Thank you for your deep interest in our research. This study aims to determine the reliability of GCMs. Furthermore, considering the uncertainty of bias correction methods in climate research, it is focused on quantifying it after selecting an appropriate method based on its performance. On the other hand, the CI in this study allows for selecting a suitable method for each continent or grid by considering both uncertainty and performance while choosing a bias correction method. Based on these processes, highly reliable results can be sufficiently ensured in climate change applications, such as water resources and energy. We will now begin answering your questions.

Comment 1

Why did this study only use the historical period? For example, do the authors believe that bias correction performance during historical periods sufficiently accounts for uncertainty in future projections?

Answer

Thank you for your comment. Previous studies have argued that higher performance of models in the historical period can lead to higher performance in future projections. This argument implies that the historical climate data sufficiently reflects non-stationarity. However, future climate predictions are exposed to various uncertainties, which limits the discussion of the accuracy of projection performance. This study divided the historical period into training and validation periods to compare QDM, EQM, and DQM. This approach provides the following two significant advantages. First, it is possible to make an intuitive judgment on predicting the future based on the corrected precipitation data in the historical period. For example, comparing the reliable reference data and the bias-corrected GCM output values in the historical period can clarify the bias-corrected performance of QDM, EQM, and DQM. In addition, the range of uncertainties that may occur in the future can be identified in advance based on the corrected climate data.

Second, it is possible to exclude uncertainties arising from various factors, such as greenhouse gas emission scenarios and structural differences between GCMs, and to evaluate only the uncertainty due to the bias-corrected method. This is because these factors can hinder the independent evaluation of the correction method.

Therefore, this study focused on verifying and comparing the inherent performance of bias correction methods using historical period data, which is consistent with the research objective.

To clarify this, we added the following sentence to the atmosphere mapping method in Sections 2-3:

This study divided the data into a training period (1980-1996) and a validation period (1997-2014) to correct the historical period's data. This approach minimizes the influence of uncertainties associated with future projections, allowing the study to focus on evaluating the intrinsic performance differences of the QM methods.

Comment 2

What are the strengths of the Comprehensive Index? The study claims to have developed this index, but could you provide a detailed explanation of its strengths and how it differs from previous methods?

Answer

Thank you for your valuable comment. The strength of the comprehensive index (CI) proposed in this study is the consideration of uncertainty and performance, distinguishing it from previous studies that primarily focused on performance when selecting bias correction methods. There are important reasons for establishing this concept. For example, the CI enables the separation of uncertainties associated with greenhouse gas emission scenarios and the choice of bias correction methods. Greenhouse gas emission scenarios can be a significant source of uncertainty in climate research. In this context, the CI proposed in this study effectively confirms various sources of uncertainty by clearly distinguishing their ranges. Furthermore, it prevents potential uncertainties arising from existing methodologies in advance, enhancing the reliability and robustness of selecting appropriate bias correction methods. This process goes beyond choosing suitable bias correction methods; it also contributes to quantifying various techniques that require a balanced evaluation of uncertainty and performance, such as GCM selection, variant label selection, and scenario selection.

We have explained the strengths of the CI in the Discussion section as follows:

Unlike previous studies that focused on the performance of bias correction methods (Song et al., 2024a; Teutschbein and Seibert, 2012; Smitha et al., 2018), this study suggests a CI that integrates the performance and uncertainty metrics. This approach enhances the robustness of bias correction method selection and provides a more holistic evaluation framework.

The additional reference included in this study is as follows:

Smitha, P.S., Narasimhan, B., Sudheer, K.P., and Annamalai, H.: An improved bias correction method of daily rainfall data using a sliding window technique for climate change impact assessment. J. Hydrol. 556, 100–118. https://doi.org/10.1016/j.jhydrol.2017.11.010, 2018.

Comment 3

Why were TOPSIS and BMA specifically used for performance and uncertainty in the Comprehensive Index? Could other methods have been equally applicable? Please discuss whether alternatives might be feasible.

Answer

Thank you for your valuable comments. TOPSIS was used as a performance index to calculate CI in this study because it is intuitive for directly estimating the closeness between positive and negative ideal solutions to determine performance-based priorities. Furthermore, It is also a method mainly used in MME development and GCM selection in climate change research. Therefore, TOPSIS was introduced in the CI calculation process so readers can easily understand and judge it. In addition, this study used BMA because it can consider both model and prediction uncertainties of the bias correction method. This approach provides the advantage of understanding the uncertainty of the bias correction method from various aspects. The CI proposed in this study is flexible because it can apply various methodologies. For example, the performance index of CI can be widely applied from simple evaluation metrics to complex methods. In addition, the uncertainty index of the CI can be used as an alternative to BMA with various techniques such as REA, standard deviation, and variance. Moreover, the CI offers flexibility in adjusting the weights. If the user or research subject places greater importance on uncertainty, the weights can be appropriately adjusted through a reasonable approach. Therefore, the CI possesses a very wide range of applicability.

We have included the strengths and flexibility of the CI framework in the methodology section as follows:

Additionally, the methodology offers flexibility in selecting performance and uncertainty metrics. Alternative MCDA methods beyond TOPSIS can be utilized for performance indicators, or indices that effectively represent the model's performance can be employed to calculate the CI. Similarly, for uncertainty indicators, approaches such as variance, standard deviation, or other uncertainty quantification techniques can be applied to enhance the robustness of the framework further.

Comment 4

Why does DQM perform worse than other methods? Could the authors explain why this method shows lower performance compared to QDM and EQM?

Answer

Thank you for your insightful comments. In the Discussion section, we previously stated that DQM has clear limitations in addressing nonlinear climate patterns and extreme events. DQM effectively corrects precipitation based on the concept of a detrending process. However, it fails to adequately account for all quantiles of the precipitation distribution, particularly at the extremes. Our study also showed that the uncertainty of DQM was higher compared to other QM methods in regions with complex climate conditions such as Southeast Asia, East Africa, and the Alps in Europe. These findings align with previous studies, suggesting that DQM faces challenges in accurately capturing long-term climate trends and variability (Berg et al., 2022; Cannon et al., 2015).

We have supplemented the Discussion to reflect on your comments as follows:

DQM showed the highest weight variance across all continents, indicating more significant uncertainty when applied to various GCMs. This uncertainty was particularly pronounced in regions with complex climate conditions, such as Southeast Asia, East Africa, and the Alps in Europe. These results align with Berg et al. (2022), who highlighted DQM's limitations in capturing long-term climate trends and extreme events. The higher uncertainty associated with DQM suggests that, while its detrending process is effective in correcting the mean, it may struggle in regions dominated by nonlinear climate patterns, as it does not sufficiently account for all quantiles in the distribution, particularly extremes, as noted by Cannon et al. (2015).

Comment 5

Why was ERA5 chosen as the reference dataset without comparing it to other reanalysis datasets (e.g., CHIRPS, GPCP)? What criteria led the authors to determine ERA5 as the most suitable for the study's regions and precipitation characteristics?

Answer

Thank you for your valuable comments. The ERA5 dataset, developed by the European Centre for Medium-Range Weather Forecasts (ECMWF), provides high-resolution reanalysis data applicable worldwide and has been widely used to evaluate the performance of climate models. In particular, due to its high temporal and spatial resolution and comprehensive assimilation of observational data, ERA5 is highly suitable as a reference dataset for applying bias correction methods. In this context, ERA5 is frequently adopted in journals ranked Q1 or higher to emphasize the reliability of research.

https://doi.org/10.1029/2024EF004541

This study utilized the ERA5 reanalysis dataset to contextualize the 2021 heatwave within historical records from 1950 to 2022.

https://doi.org/10.1029/2021EF002625

This study employed ERA5 reanalysis data to evaluate the performance of CMIP6 models in simulating historical climate conditions.

https://doi.org/10.5194/gmd-17-191-2024

This study used ERA5 data to assess the performance of downscaled climate projections and compared them with historical reanalysis data.

https://doi.org/10.5194/gmd-17-8665-2024

This study utilized ERA5 as a reference to evaluate the performance of downscaled climate products.

To justify our use of ERA5 as a reference dataset, we added the following statement to our methodology:

ERA5 has been widely used in various studies to ensure reliability of climate model evaluation and climate change assessment (Jeong et al., 2024; Virgilio et al., 2024; Baek et al., 2024).

We have added the following references to the article:

Jeong, D.I., Yu, B., Cannon, A.J.: 2021 Heatwave Over Western North America: Structural Uncertainty and Internal Variability in GCM Projections of Humidex and Temperature Extremes. Earth's futur 12, 8, e2024EF004541. https://doi.org/10.1029/2024EF004541, 2024.

Virgilio, G.D., Tam, F.J.E., Nishant, N., Evans, J.P., Thomas, C., Riley, M.L., Beyer, K., Grose, M.R., Narsey, S., Delage, F.: Selecting CMIP6 GCMs for CORDEX Dynamical Downscaling: Model Performance, Independence, and Climate Change Signals. Earth's futur 10, 4, e2021EF002625. https://doi.org/10.1029/2021EF002625, 2024.

Baek, S.H., Ullrich, P.A., Dong, B., Lee, J.: Evaluating downscaled products with expected hydroclimatic co-variances. Geosci. Model Dev. 17, 23, 8665–8681. https://doi.org/10.5194/gmd-17-8665-2024, 2024.

Comment 6

While unifying data resolution to 1°x1° is advantageous, it might obscure detailed regional characteristics that can be captured with higher resolutions. Why was this resolution chosen, and do the authors believe that bias correction performance could differ at higher resolutions?

Answer

Thank you for your comment. Since GCM outputs are typically provided at a coarser resolution than reference data, we unified the data resolution to 1°x1° to ensure consistency across all GCM outputs and reference datasets. Using a unified resolution simplifies the comparison and evaluation of bias correction performance while avoiding inconsistencies arising from re-gridding to multiple resolutions. We acknowledge that higher-resolution data could capture more detailed regional characteristics, particularly in areas with complex topography or localized climatic patterns. However, the computational resources required to process and apply bias correction at finer resolutions for many GCMs and a long time would have been substantial.

The goal of bias correction, especially quantile mapping methods, is to minimize the differences between GCM outputs and observational data. Indeed, higher-resolution data could yield more precise correction results by capturing finer-scale climatic features. However, the objective of this study is to evaluate the performance of bias correction methods rather than to optimize their application to a specific resolution. While we acknowledge that high resolutions may enhance the accuracy of bias correction results, we believe that the resolution used does not significantly affect the comparison of bias correction methods. This is because the evaluation focuses on the inherent characteristics of the bias correction techniques rather than their interaction with specific spatial scales. In future studies, exploring how bias correction performance might vary across different resolutions when applied to regional or localized contexts would be interesting. However, for this study, we believe that the chosen resolution sufficiently serves its purpose without compromising the validity of the results.

Comment 7

Entropy theory was clearly used to determine the weights, but could the authors explain the impact and significance of low weights for certain metrics (e.g., EVS, NSE)?

Answer

Thank you for your comment. We used the entropy theory to calculate the weights for each alternative in TOPSIS. Entropy theory objectively derives weight based on the intrinsic variability of the data, minimizing the influence of subjective judgments. The metrics with more significant variability are assigned to higher weights. Of course, there are limitations in that they depend on relative variability and cannot integrate expert opinions, but entropy theory is a well-established method for ensuring the objectivity of data.

When applied to TOPSIS, entropy-based weights ensure that metrics with higher informational significance play a more critical role in determining the rankings of alternatives. EVS and NSE may have relatively low variability and thus lower weights, but they are still reflected in the decision-making process, ensuring a comprehensive evaluation of each alternative.

We added the following sentence to the methodology to highlight the advantages of entropy weights:

This study used entropy theory to calculate the weights for each criterion. Entropy weighting ensures sufficient objectivity by calculating weights based on the variability and distribution of data. This approach minimizes subjectivity, preventing biases in the weighting process.

Comment 8

EQM showed the lowest uncertainty. Is this result due to the characteristics of the methodology itself, or is it influenced by specific regional or data traits?

Answer

Thank you for your valuable comment. The low uncertainty of EQM may be attributed to a combination of the methodology's characteristics and the specific traits of the regions and data used. EQM's low uncertainty lies in the synergy between the methods and regional data characteristics. For instance, EQM adjusts biases at each precipitation distribution quantile by aligning the reference and model data's empirical cumulative distribution functions (ECDFs). Based on this, EQM provides detailed and localized corrections for model biases.

Furthermore, it applies uniformly corrections across the entire precipitation distribution, thereby reducing uncertainty. This approach may result in lower uncertainty when capturing extreme events. Since EQM relies on quantile alignment, it performs particularly well in regions with reliable precipitation distributions. This study's global-scale correction of daily precipitation from GCMs enhanced EQM's performance by leveraging the empirical cumulative distribution functions. However, it is important to note that while EQM exhibited low uncertainty in this study, its performance may vary depending on the complexity of regional climates, data quality, and temporal resolution.

We have included the reasons for EQM's high performance in the discussion section as follows:

In particular, EQM is consistent with previous studies in that it more accurately corrects observed distributions in non-stationary and highly variable climate variables, such as precipitation (Themeßl et al., 2012; Maraun, 2013; Gudmundsson et al., 2012). These positive aspects are mainly due to EQM’s ability to align the empirical ECDFs of reference and model data across all quantiles, allowing it to correct biases with high precision at both central tendencies and extremes.

Comment 9

Each QM method shows strengths and weaknesses in specific regions. Was there an attempt to develop a hybrid approach that combines these methods? For instance, could DQM’s ability to remove long-term trends and EQM’s stable performance be integrated?

Answer

Thank you for your insightful comment. It is innovative and feasible to propose an integrated methodology based on the strengths of DQM and EQM. I have previously developed a flexible double-distribution quantile mapping. This approach can use an appropriate distribution function that considers the precipitation characteristics across various regions to better correct for extreme precipitation.

Based on this experience, we can develop a hybrid methodology that integrates the strengths of DQM and EQM. DQM is effective in addressing non-stationarity by removing long-term trends from data. This concept ensures that bias-corrected data accurately reflects climate variability. It is also beneficial for long-term climate analysis by ensuring that the corrected data accurately reflects the underlying climate variability and temporal dynamics.

EQM is advantageous in ensuring accurate correction for central tendency and extreme values by aligning CDFs in all quantiles. Based on this, we can integrate the detrending function of DQM and the quantile-based correction of EQM. This hybrid approach is consistent with our goal of improving bias-correcting performance across diverse regions and datasets. We intend to explore this concept further in future research. We added the following sentence in the conclusion:

Based on the results of this study, future studies can develop hybrid methodologies that combine the strengths of each QM.

Comment 10

The authors claim that CI considers both performance and uncertainty. However, there is insufficient explanation on whether region-specific weights were applied. Was CI calculated for each grid? For example, regions with extreme precipitation distributions (e.g., the Sahara Desert) differ significantly from moderate regions (e.g., Northern Europe). If CI was calculated at a larger scale, would that be appropriate? Alternatively, is it reasonable to apply uniform weights?

Answer

Thank you for your comment. In this study, the CI was calculated for every grid cell at a resolution of 1°×1°, and the index was derived individually for each grid. This grid-wise calculation allows the CI to reflect regional climate characteristics, including regions with extreme precipitation distributions, such as the Sahara Desert, and areas with moderate conditions, such as Northern Europe. The CI proposed in this study already has the flexibility to adjust the weights according to the research objectives about using uniform weights. Furthermore, this study can be applied based on the same weighting for performance and uncertainty, and the framework of the method is designed to allow users to emphasize uncertainty or performance depending on the research objectives. For example, a study focusing on extreme climate events may assign a higher weight to uncertainty, while an application requiring model reliability may prioritize performance.

We believe that this flexibility enhances the adaptability of the CI framework and ensures its applicability to various research contexts and regional characteristics.

We added the following sentences to make the calculation process of the comprehensive index clearer:

The CI is calculated individually for every grid and can reflect climate characteristics. Framework provides flexibility in determining the weighting of uncertainty or performance depending on the study objectives.

Comment 11

Why was TOPSIS chosen over other MCDA techniques like AHP (Analytic Hierarchy Process) or VIKOR? A discussion on this decision is necessary.

Answer

Thank you for your insightful comment. We chose TOPSIS over other MCDA techniques, such as AHP or VIKOR, to minimize the influence of subjective judgment and focus only on numerical data for performance evaluation. AHP and VIKOR are valuable approaches but require expert opinions or subjective assessment, which may introduce bias into the decision-making process. On the other hand, TOPSIS provides a purely data-based framework that evaluates alternatives based on the closeness coefficient to the ideal solution and the distance from the negative ideal solution.

Furthermore, this study used entropy theory to calculate the weights of TOPSIS components to ensure objectivity further. Entropy-based weights are derived directly from the variability of the data, eliminating the need for subjective inputs and enhancing the reliability of the evaluation process.

The combination of TOPSIS and entropy theory is consistent with the goal of this study to provide an unbiased and transparent framework for evaluating the performance and uncertainty of bias correction methods across different regions and data sets.

Comment 12

Lastly, some sentences could benefit from structural refinement. To enhance the clarity and flow of the text, it would be helpful to revise the sentences throughout the manuscript.

Answer

Thank you for your comment. We have reviewed the entire article's grammar in response to your comments.

Intercomparison of bias correction methods for precipitation of multiple GCMs across six continents

Young Hoon Song[1], Eun-Sung Chung[1*]

[1] Faculty of Civil Engineering, Seoul National University of Science and Technology, 232 Gongneung-ro, Nowon-gu, Seoul 01811, Korea

[*] Correspondence to: Eun-Sung Chung eschung@seoultech.ac.kr

Abstract

This study, conducted across six continents, evaluated and compared the effectiveness of three Quantile Mapping (QM) methods: Quantile Delta Mapping (QDM), Empirical Quantile Mapping (EQM), and Detrended Quantile Mapping (DQM) for correcting daily precipitation data from 11 CMIP6 General Circulation Models (GCMs). The performance of corrected precipitation data was evaluated using ten evaluation metrics, and the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) was applied to calculate performance-based priorities. Bayesian Model Averaging (BMA) was used to quantify model-specific and ensemble prediction uncertainties. Subsequently, this study developed a comprehensive index by aggregating the performance scores from TOPSIS with the uncertainty metrics from BMA. The results showed that EQM performed the best on all continents, effectively managing performance and uncertainty. QDM outperformed other methods in specific regions and was selected more frequently than DQM when greater weight was given to uncertainty. It suggests that daily precipitation corrected by QDM is more stable than DQM. On the other hand, DQM effectively reproduces dry climate but shows the highest uncertainty in certain regions, suggesting potential limitations in capturing long-term climate trends. This study emphasizes that both performance and uncertainty should be considered when choosing a bias correction method to increase the reliability of climate predictions.

Keywords

CMIP6 GCM, Bias correction, Uncertainty, TOPSIS, Comprehensive index

1. Introduction

The Coupled Model Intercomparison Project (CMIP) General Circulation Models (GCMs) have provided critical scientific evidence to explore climate change (IPCC, 2021; IPCC, 2022). Nevertheless, GCMs exhibit significant biases compared to observational data for reasons such as incomplete model parameterization and inadequate understanding of key physical processes (Evin et al., 2024; Zhang et al., 2024; Nair et al., 2023). These deficiencies with GCM have introduced various uncertainties in climate projections, making ensuring sufficient reliability in climate change impact assessments difficult. In this context, many studies have proposed various bias correction methods to reduce the discrepancies between observational data and GCM simulations, thereby providing more stable results than raw GCM-based assessments (Cannon et al., 2015; Themeßl et al., 2012; Piani et al., 2010). Despite these advancements, the suggested bias correction methods differ in their physical approaches, resulting in discrepancies in the climate variables adjusted for historical periods. Furthermore, the distribution of precipitation across continents and specific locations causes variations in the correction outcomes depending on the method used, which makes it challenging to reflect extreme climate events in future projections and adds another layer of confusion to climate change research (Song et al., 2022b; Maraeun, 2013; Ehret et al., 2012; Enayati et al., 2021). Thus, exploring multiple aspects to make reasonable selections when applying bias correction methods specific to each continent and region is necessary.

Many studies have developed appropriate bias correction methods based on various theories, which have reduced the difference between GCM simulations and observed precipitation (Abdelmoaty and Papalexiou, 2023; Shanmugam et al., 2024; Rahimi et al., 2021). The Quantile Mapping (QM) series has been widely adopted among bias correction methods due to its conceptual simplicity, ease of application, and adaptability to various methodologies. However, although standard QM methods have high performance in correcting stationary precipitation, they are less efficient in non-stationary data, such as extreme precipitation events (Song et al., 2022b). To address these limitations, a recent study proposed an improved QM approach to reflect future non-stationary precipitation across all quantiles of historical precipitation (Rajulapati and Papalexiou, 2023; Cannon et al., 2015; Cannon, 2018; Song et al., 2022b). In recent years, climate studies using GCMs have adopted several improved QM methods that offer higher performance than previous methods to correct historical precipitation and project it accurately into the future. For example, Song et al. (2022b) performed bias correction on daily historical precipitation over South Korea using distribution transformation methods they developed and found that the best QM method varied depending on the station. Additionally, previous studies have reported that QM performance varied by grid and station (Ishizaki et al., 2022; Chua et al., 2022). From this perspective, these improved QMs may only guarantee uniform results across some grids and regions. Therefore, to analyze positive changes in future climate impact assessments, selecting appropriate bias correction methods based on a robust framework is essential.

Multi-criteria decision analysis (MCDA) is efficient for prioritization because it can aggregate diverse information from various alternatives. MCDA has been extensively used across different fields to select suitable alternatives, with numerous studies confirming its stability in priority selection (Chae et al., 2022; Chung and Kim, 2014; Song et al., 2024a). Moreover, MCDA has been employed in future climate change studies to provide reasonable solutions to emerging problems, including the selection of bias correction methods for specific regions and countries (Homsi et al., 2019; Saranya and Vinish, 2021). However, MCDA's effectiveness is sensitive to the source and quality of alternatives, making accurate ranking challenging when information is lacking or overly focused on specific criteria (Song and Chung, 2016). Small-scale regional and observation-based studies have conducted GCM performance evaluations, but global and continental-scale evaluations are rare due to the substantial time and cost required.

GCM simulation includes uncertainties from various sources, such as model structure, initial condition, boundary condition, and parameters (Pathak et al., 2023; Cox and Stephenson, 2007; Yip et al., 2011; Woldemeskel et al., 2014). The selection of bias correction methods contributes significantly to uncertainty in climate change research using GCMs. Jobst et al. (2018) argued that GHG emission scenarios, bias correction methods, and GCMs are primary sources of uncertainty in climate change assessments across various fields. The extensive uncertainties in GCMs complicate the efficient establishment of adaptation and mitigation policies. This issue has increased awareness of the uncertainties inherent in historical simulations. Consequently, many studies have focused on estimating uncertainties using diverse methods to quantify these uncertainties (Giorgi and Mearns, 2002; Song et al., 2022a; Song et al., 2023). Although it is impossible to drastically reduce the uncertainty of GCM outputs due to the unpredictable nature of climate phenomena, uncertainties in GCM simulations can be reduced using ensemble principles, such as multi-model ensemble development using a rational approach (Song et al., 2024). However, accurately identifying biases in simulation precipitation remains challenging due to the lack of comprehensive equations reflecting Earth's physical processes. In this context, climate change studies have aimed to quantify the uncertainty of historical climate variables in GCMs, offering insights into the variability of GCM simulations (Pathak et al., 2023). Bias-corrected precipitation of GCMs using QM has shown high performance in the historical period, which is expected to result in better future predictions. However, the physical concepts of various QMs may lead to more significant uncertainty in the future (Lafferty et al., 2023). Therefore, efforts should be made to consider and reduce uncertainty in the GCM selection process. It will ensure the reliability of predictions by selecting an appropriate bias-correcting method.

This study aims to compare the performance of three bias correction methods using daily historical precipitation data (1980-2014) from CMIP6 GCMs across six continents (South America: SA; North America: NA; Africa: AF; Europe: EU; Asia: AS; and Oceania: OA). Ten evaluation metrics were used to assess the performance of daily precipitation corrected by the three QM methods for each continent. Subsequently, the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) of MCDA was applied to select an appropriate bias correction method for each continent. Additionally, the uncertainty in daily precipitation for historical periods was quantified using Bayesian Model Averaging (BMA). By integrating performance scores from TOPSIS and uncertainty metrics from BMA, this study developed a Comprehensive Index (CI), which was then used to select the best bias correction method for each continent. This comprehensive approach ensures a balanced consideration of both performance and uncertainty, enhancing understanding of the bias correction process based on the distribution of daily precipitation across continents.

2. Datasets and methods

2.1 General Circulation Model

This study used 11 CMIP6 GCM to perform bias correction for daily precipitation in the historical period. This study used daily precipitation to correct bias because the natural variability relative to projected anthropogenically forced trends is much larger for precipitation than for temperature (Deser et al., 2012). Table 1 presents basic information, including model names, resolution, and variant labels. The model resolution of 11 CMIP6 GCMs was equally re-gridded to 1°×1° using linear interpolation. Furthermore, this study's ensemble member of CMIP6 GCMs was the first member of realizations (r1).

Table 1. Information of CMIP6 GCMs in this study

Models	**Resolution**	**Climate variables**	**Variant label**
ACCESS-CM2	1.2° × 1.8°	Daily precipitation	r1i1p1f1
ACCESS-ESM1-5	1.2° × 1.8°		
BCC-CSM2-MR	1.1° × 1.1°		
CanESM5	2.8° × 2.8°		
CESM2-WACCM	0.9° × 1.3°		
CMCC-CM2-SR5	~ 0.9°		
CMCC-ESM2	0.9° × 1.25°		
EC-Earth3-Veg-LR	1.0° × 1.0°		
GFDL-ESM4	1.4° × 1.4°		
INM-CM4-8	~ 0.9°		
IPSL-CM6A-LR	1.1° × 1.1°		

2.2 Reference data

This study utilized ERA5 reanalysis data from the European Center for Medium-Range Weather Forecasts (ECMWF) as reference data. The model physics of ERA5 reanalysis data improved as it employed an Integrated Forecasting System based on CY41r2 (Hersbach et al., 2020). ERA5 has been widely used in various studies to ensure the reliability of climate model evaluation and climate change assessment (Jeong et al., 2024; Virgilio et al., 2024; Baek et al., 2024). The model resolution selected in this study was 1.0° × 1.0°, which was provided by the institution for research availability. The accuracy of assessing GCM simulation is crucial for replicating the spatial and temporal variability of observed data (Hamed et al., 2023). In this context, the ERA5 product has been commonly used to reproduce observed precipitation, for the evaluation of GCMs' performances.

2.3 Quantile mapping

This study employed three (Quantile delta mapping, QDM; Detrended quantile mapping, DQM; Empirical quantile mapping, EQM) QM methods to correct the simulation of CMIP6 GCMs, and these methods are commonly used in climate change research based on the climate models (Switanek et al., 2017). This study divided the data into a training period (1980-1996) and a validation period (1997-2014) to correct the historical period's data. This approach minimizes the influence of uncertainties associated with future projections, allowing the study to focus on evaluating the intrinsic performance differences of the QM methods. The frequency-adaptation technique, as described by Themeßl et al. (2012), was applied to address potential biases and improve the accuracy of the corrections. The corrected precipitation using the QM used a cumulative distribution function, as shown in Equation 1, to reduce the difference from the reference data.

$$\hat{x}_{m,p}(t) = F_{o,h}^{-1}\{F_{m,h}[x_{m,p}(t)]\} \quad (1)$$

where, $\hat{x}_{m,p}(t)$ presents the bias-corrected results. $F_{o,h}$ represents the cumulative distribution function (CDF) of the observed data, and $F_{m,h}$ presents the CDF of the model data. The subscripts o and m denote observed and model data, respectively, and the subscript h denotes the historical period.

QDM, developed by Cannon et al. (2015), preserves the relative changes ratio of modeled precipitation quantiles. In this context, QDM consists of bias correction terms derived from observed data and relative change terms obtained from the model. The computation process of QDM is carried out as described in Equation (2) to (4).

$$\hat{x}_{m,p}(t) = \hat{x}_{o:m,h:p}(t) \cdot \Delta_m(t) \quad (2)$$

$$\hat{x}_{o:m,h:p}(t) = F_{o,h}^{-1}[F_{m,p}^{(t)}\{x_{m,p}(t)\}] \quad (3)$$

$$\Delta_m(t) = \frac{x_{m,p}(t)}{F_{m,h}^{-1}[F_{m,p}^{(t)}\{x_{m,p}(t)\}]} \quad (4)$$

where, $\hat{x}_{o:m,h:p}(t)$ presents the bias corrected daily precipitation for the historical period, and $\Delta_m(t)$ the relative change in the model simulation between the reference period and the target period. In addition, the target period is calculated by multiplying the relative change ($\Delta_m(t)$) at time (t) multiplied by the bias-corrected precipitation in the reference period. $\Delta_m(t)$ is defined as $\widehat{x_{m,p}}(t)$ divided by $F_{o,h}^{-1}[F_{m,p}^{(t)}\{x_{m,p}(t)\}]$. $\Delta_m(t)$ preserving the relative change between the reference and target periods. DQM, while more limited compared to QDM, integrates additional information regarding the projection of future precipitation. Furthermore, climate change signals estimated from DQM tend to be consistent with signals from baseline climate models. The computational process of DQM is performed as shown in Equation (5).

$$\hat{x}_{m,p} = F_{o,h}^{-1}\left\{F_{m,h}\left[\frac{\bar{X}_{m,h}X_{m,h}(t)}{\bar{X}_{m,p}(t)}\right]\right\}\frac{\bar{X}_{m,p}(t)}{\bar{X}_{m,h}} \quad (5)$$

where, $\bar{X}_{m,h}$ and $\bar{X}_{m,p}$ represent the long-term modeled averages for the historical reference period and the target period, respectively.

EQM is a method that corrects the quantiles of the empirical cumulative distribution function from a GCM simulation based on a reference precipitation distribution using a corrected transfer function (Dequé, 2007). The calculation process of EQM can be represented as follows in Equation (6).

$$\hat{x}_{m,p}(t) = F_{o,h}^{-1}(F_{m,h}(x_{m,p}(t))) \quad (6)$$

All these QMs can be applied to historical data correction in this approach. The bias correction is performed based on the relative changes between a reference period and a target period in the past, ensuring that the relative changes between these periods are preserved in the corrected data (Ansari et al., 2023; Tanimu et al., 2024; Cannon et al., 2015).

2.4 Evaluation metrics

This study used ten evaluation metrics to assess the output performance of three quantile mapping methods against the reference data for the validation period (1997-2014). Seven evaluation metrics used in this study are as follows: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2), Percent bias (Pbias), Nash-Sutcliffe Efficiency (NSE), Kling-Gupta efficiency (KGE), Median Absolute Error (MdAE), Mean Squared Logarithmic Error (MSLE), Explained Variance Score (EVS), and Jenson-Shannon divergence (JS-D). The equations of seven evaluation metrics are presented in Table 2.

Table 2. Information of the seven-evaluation metrics used in this study

Metrics	Equations	Factors	References
RMSE	$= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{sim} - X_i^{ref}\right)^2}$	X_i^{ref} reference data X_i^{sim} Bias corrected GCM	
MAE	$= \sum_{i=1}^{n}\left\lvert X_i^{sim} - X_i^{ref}\right\rvert$		
R^2	$= 1 - \frac{\sum_{i=1}^{n}(X_i^{sim} - X_i^{ref})^2}{(X_i^{ref} - \bar{X}_i^{ref})^2}$		Galton, 1886
Pbias	$= \frac{\sum_{i=1}^{n}(X_i^{ref} - X_i^{sim})}{\sum_{i=1}^{n} X_i^{ref}} \times 100$		

NSE	$= 1 - \frac{\sum_{i=1}^{n}(X_i^{sim} - X_i^{ref})^2}{\sum_{i=1}^{n}(X_i^{ref} - \bar{X}_i^{ref})^2}$		Nash and Sutcliffe, 1970
MdAE	$= median(\lvert X_i^{sim} - X_i^{ref} \rvert)$		
MSLE	$= \frac{1}{n}\sum_{i=1}^{n}(\log(1 + X_i^{sim}) - \log(1 + X_i^{ref}))^2$		
EVS	$= 1 - \frac{Var(X^{sim} - X^{ref})}{Var(X^{ref})}$		
KGE	$= 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2}$	r Pearson product-moment correlation α Variability error β: Bias term	Gupta et al. 2009
JS-D	$= \frac{1}{2} D_{KL}\left(P \parallel \frac{P+Q}{2}\right) + \frac{1}{2} D_{KL}\left(Q \parallel \frac{P+Q}{2}\right)$	$P(x)$: Probability density distribution of reference data $Q(x)$: Probability density distribution of GCM D_{KL}: KL-D	Lin, 1991

Ten evaluation metrics selected in this study assess GCM performance from various perspectives, including error (RMSE, MAE, MdAE, and MSLE), deviation (Pbias), accuracy (R^2, NSE), variability (EVS), correlation and overall performance (KGE), and distributional differences (JSD). These metrics complement each other by offering a comprehensive evaluation framework. For instance, while NSE evaluates the overall fit of the simulated data to observations, KGE provides a holistic view by integrating correlation, variability, and bias into a single efficiency score, and JS-D captures the difference between the distributions of the reference data and the bias-corrected GCM output.

2.5 Generalized extreme value

This study used generalized extreme value (GEV) to compare the extreme precipitation calculated by the bias-corrected GCM at each grid of six continents over the historical period. The historical precipitation was compared with the distribution of reference data and bias-corrected GCM above the 95th quantile of the Probability Density Function (PDF) of the GEV distribution (Hosking et al. 1985). In addition, this study compared the distribution differences between the reference data based on the GEV distribution and the corrected GCM using JSD. GEV distribution is commonly used to confirm extreme values in climate variables. The PDF of the GEV distribution is shown in Equation 7, and the parameters of the GEV distribution were estimated using L-moment (Hosking, 1990).

$$g(x) = \frac{1}{\alpha}\left[1 - k\frac{x-\epsilon}{\alpha}\right]^{\frac{1}{k}-1} exp\left\{-\left[1 - k\frac{x-\epsilon}{\alpha}\right]^{\frac{1}{k}}\right\} \quad (7)$$

where, k, α, and ε represents a shape, scale, and location of the GEV distribution, respectively.

2.6 Bayesian model averaging (BMA)

The BMA is a statistical technique that combines multiple models to provide predictions that account for model uncertainty (Hoeting et al., 1999). BMA is used to integrate predictions from GCMs to improve the robustness and reliability of the resulting assemblies. The posterior probability of each model is calculated based on Bayes' theorem as shown in Equation 8.

$$P(M_k \mid D) = \frac{P(D|M_k)P(M_k)}{\sum_{j=1}^{K} P(D|M_j)P(M_j)} \quad (8)$$

where, $P(M_k)$ is the prior probability of model M_k, and $P(D \mid M_k)$ s the likelihood of the data D given model M_k, $P(M_k \mid D)$ is the posterior probability of model M_k. In addition, the BMA prediction $\hat{Q}_{BMA}$ is the weighted average of the predictions from each model as shown in Equation 9.

$$\hat{Q}_{BMA} = \sum_{k=1}^{K} P(M_k \mid D)\hat{Q}_k \quad (9)$$

where, $\hat{Q}_k$ is the prediction from model M_k. In this study, BMA was used to quantify the model uncertainty and ensemble prediction uncertainty for daily precipitation corrected by three QM methods (QDM, EQM, and DQM) applied to 11 CMIP6 GCMs, as shown in Equations 10 and 11.

$$\alpha_w^2 = \frac{1}{K}\sum_{k=1}^{K}(w_k - \overline{w})^2 \quad (10)$$

where, K is the number of models, $w_k = P(M_k \mid D)$ is the weight of model M_k, $\overline{w}$ is the mean of the weights, given by $\overline{w} = \frac{1}{K}\sum_{k=1}^{K} w_k$. A higher variance in model weights indicates more significant prediction differences, implying greater model uncertainty.

$$\sigma BMA = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(\hat{Q}_k - \hat{Q}BMA)^2} \quad (11)$$

σBMA is standard deviation of the BMA ensemble predictions, $\hat{Q}_k$ is the prediction from each model M_k, $\hat{Q}BMA$ is the weighted average prediction from BMA. This standard deviation represents the variability among the ensemble predictions and serves as an indicator of uncertainty. A lower standard deviation implies higher consistency among predictions, indicating lower uncertainty, while a higher standard deviation suggests greater variability and higher uncertainty.

2.7 TOPSIS

This study used TOPSIS to calculate a rational priority among three QM methods based on the outcomes derived from evaluation metrics. Furthermore, the closeness coefficient calculated using TOPSIS was used as the performance metric for the CI. Proposed by Hwang and Yoon (1981), TOPSIS is a multi-criteria decision-making technique frequently used in water resources and climate change research to select alternatives (Song et al., 2024). As described in Equation 12 and 13, the proximity of the three QM methods is calculated based on the Positive Ideal Solution (PIS) and the Negative Ideal Solution (NIS).

$$D_i^+ = \sqrt{\sum_{j=1}^{n} w_j (f_j^+ - f_{i,j})^2} \quad (12)$$

$$D_i^- = \sqrt{\sum_{j=1}^{n} w_j (f_j^- - f_{i,j})^2} \quad (13)$$

where, D_i^+ is the Euclidean distance of each criterion from the PIS, summing the whole criteria for an alternative f_j^+, j presents the normalized value for the alternative f_j^+. w_j presents weight assigned to the criterion j. D_i^- is the distance between the alternative f_j^- and the NIS. The relative closeness is calculated as shown in Equation 14. The optimal value is closer to 1 and represents a reasonable alternative.

$$C_i = \frac{D_i^-}{(D_i^- + D_i^+)} \quad (14)$$

This study used entropy theory to calculate the weights for each criterion. Entropy weighting ensures sufficient objectivity by calculating weights based on the variability and distribution of data. This approach minimizes subjectivity, preventing biases in the weighting process.

2.8 Comprehensive index (CI)

This study proposed a CI to select the best QM method by combining performance scores and model uncertainty indicators. The CI integrates the performance scores (closeness coefficient) derived from the TOPSIS method with the uncertainty quantified using BMA. This approach allows for a balanced evaluation that considers both the effectiveness of the QM methods and the associated uncertainties. Uncertainty was quantified in two ways. Model-specific weight variance was calculated using the variance of the model weights assigned by BMA, representing the uncertainty in selecting the appropriate QM. The standard deviation of BMA ensemble prediction was calculated to capture the spread and, thus, the uncertainty of the ensemble forecasts. Both the indicators were normalized using a min-max scaler to ensure comparability. The CI is calculated individually for every grid and can reflect climate characteristics. Framework provides flexibility in determining the weighting of uncertainty or performance depending on the study objectives. Additionally, the methodology offers flexibility in selecting performance and uncertainty metrics. Alternative MCDA methods beyond TOPSIS can be utilized for performance indicators, or indices that effectively represent the model's performance can be employed to calculate the CI. Similarly, for uncertainty indicators, approaches such as variance, standard deviation, or other uncertainty quantification techniques can be applied to enhance the robustness of the framework further. Finally, the calculation process of the CI is performed as shown in Equations 15 and 16.

$$UI = \frac{V_w + \sigma_e}{2} \quad (15)$$

$$CI = \alpha \times C_i - \beta \times UI \quad (16)$$

where, UI represents the uncertainty indicator. V_w and σ_e represent the normalized weight variance and the normalized ensemble standard deviation, respectively, calculated using BMA. C_i represents the closeness coefficient calculated from TOPSIS. α represents the weight given to the performance score, β represents the weight given to the uncertainty indicator. Furthermore, by adjusting the weights α and β, the study evaluated the QM methods under different scenarios. Equal weight (α= 0.5, β=0.5) balances performance and uncertainty equally, and the emphasized performance weight (α= 0.7, β=0.3) prioritize performance over uncertainty. The emphasized uncertainty weight (α= 0.3, β=0.7) prioritize uncertainty over performance. The results from the CI provide a holistic evaluation of the QM methods, considering both their effectiveness in bias correction and the reliability of their predictions.

3. Result

3.1 Assessment of bias correction reproducibility across continents

3.1.1 Comparison of bias correction effects

This study applied three QM methods to correct daily precipitation data from 11 CMIP6 GCMs across six continents. Figure 1 presents the results of comparing daily precipitation data before and after bias correction using the Taylor diagram. In general, the precipitation corrected by DQM showed a larger difference from the reference data than other methods. In contrast, EQM performed better than DQM, and many models showed results close to the reference data. The precipitation corrected by QDM also showed good performance in most continents but slightly lower than EQM. Nevertheless, QDM showed clearly better results than DQM.

Regarding correlation coefficients, precipitation corrected by DQM showed relatively high values between 0.8 and 0.9 but lower than EQM and QDM. The precipitation corrected by EQM showed high agreement with the reference data, recording correlation coefficients above 0.9 in most continents. QDM generally showed similar correlation coefficients to EQM but slightly lower values than EQM in North America and Asia.

For RMSE, precipitation corrected by DQM was higher than EQM and QDM, indicating that the corrected precipitation differed more from the reference data. On the other hand, EQM had the lowest RMSE and showed superior performance compared to other methods. QDM had slightly higher RMSE than EQM but still outperformed DQM.

In terms of standard deviation, precipitation corrected by DQM was higher or lower than the reference data in most continents. On the other hand, precipitation corrected by EQM was similar to the reference data and almost identical to the reference data in Africa and Asia. QDM was similar to the reference data in some continents but showed slight differences from EQM. These results imply that the precipitation corrected by the three methods outperforms the raw simulation, which confirms that the GCM's daily precipitation is reliably corrected in the historical period.

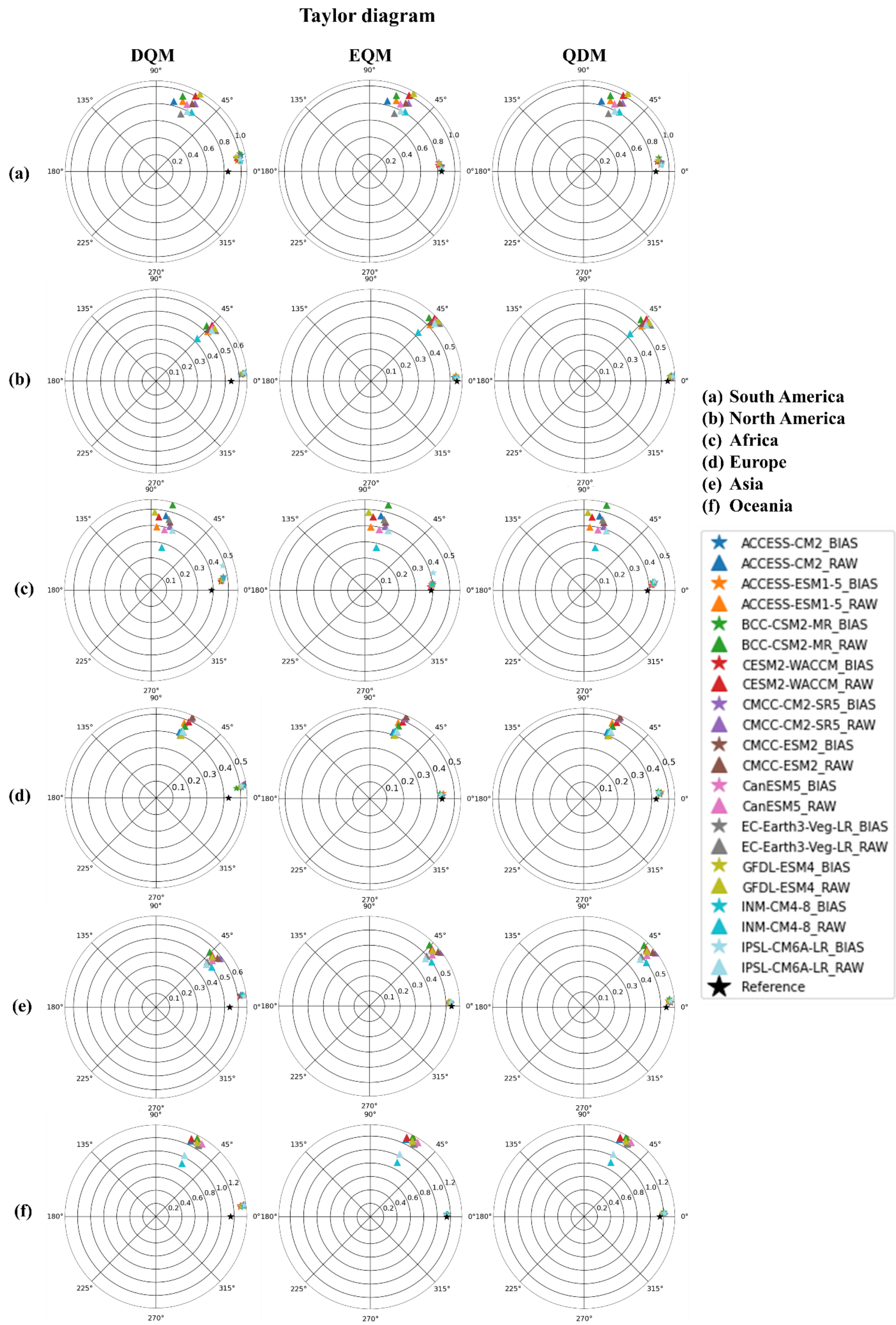

Figure 1. Comparison of raw and corrected daily precipitation on six continents using Taylor diagrams

3.1.2 Spatial distribution of bias correction performance

This study evaluated the performance of daily precipitation across six continents using ten evaluation metrics for 11 CMIP6 GCMs. Figures 2 and S1 present the spatial patterns of these evaluation metrics, calculated for daily precipitation from the bias corrected GCMs in South America. Overall, the precipitation corrected by EQM demonstrated lower JSD values, as well as higher EVS and KGE values, compared to other methods. The precipitation corrected by EQM showed higher EVS in certain regions but slightly lower performance in MdAE and Pbias across some grids. DQM exhibited performance similar to EQM and QDM in most evaluation indices but was relatively lower in most evaluation metrics. The precipitation corrected by the three methods was underestimated compared to the reference data in northern South America, while it was overestimated in eastern South America. In addition, precipitation corrected by the DQM method tended to be overestimated more than the other methods, while the EQM method showed the opposite result. Furthermore, the daily precipitation corrected by EQM showed the lowest overall error and high performance in both NSE and R^2. QDM and DQM also performed well but exhibited slightly larger errors in some regions than EQM.

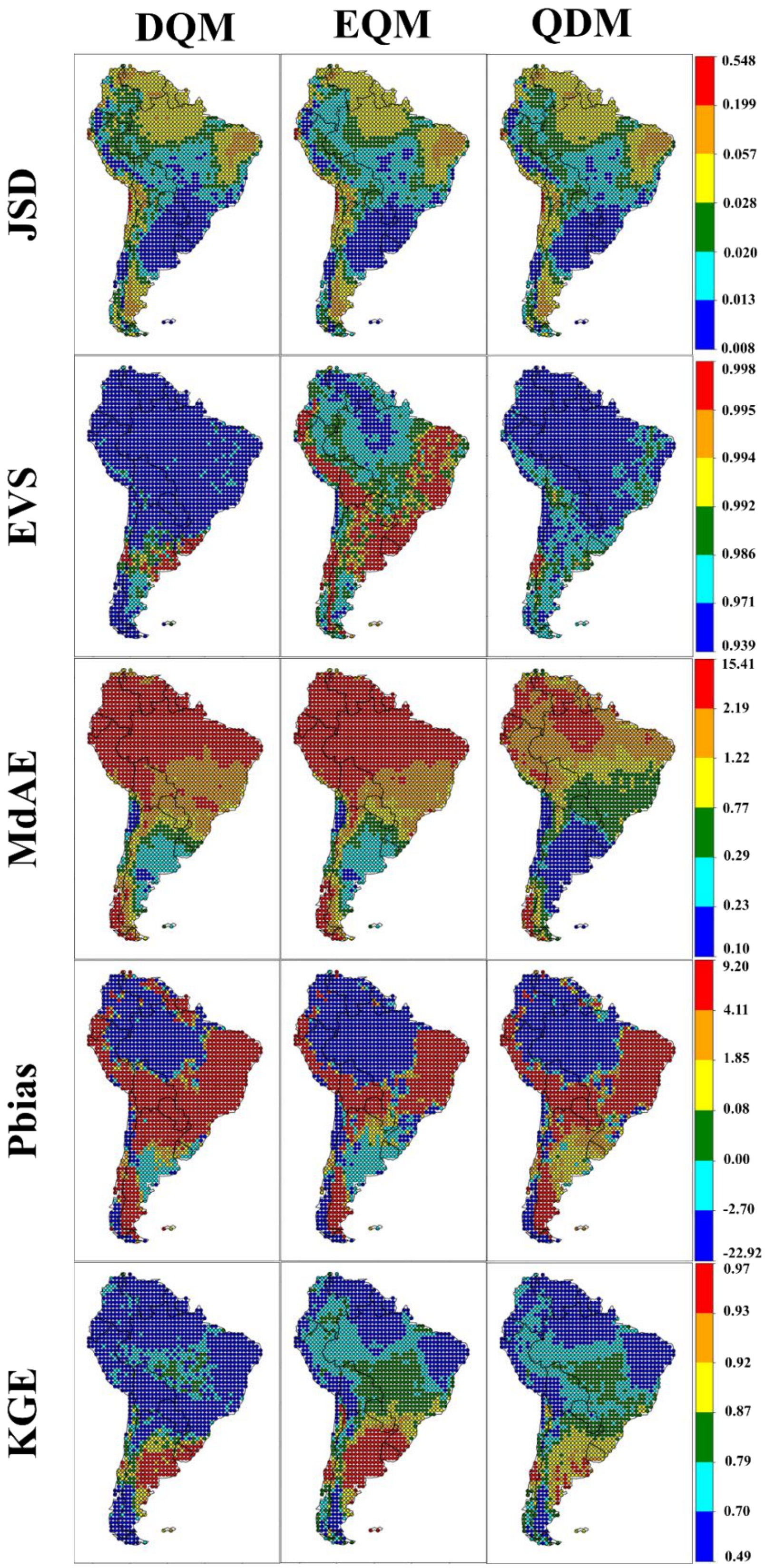

Figure 2. Performance comparison of DQM, EQM, and QDM using evaluation metrics (JSD, EVS, MdAE, Pbias, and KGE) for daily precipitation in South America.

Figures 3 and S2 present the spatial patterns of these evaluation metrics, calculated for daily precipitation from the bias corrected GCMs in South America. Regarding error metrics (MAE, MSLE, RMSE, and MdAE), precipitation corrected using DQM showed relatively lower performance across North America, with substantial errors in the southern region. In contrast, precipitation corrected using EQM demonstrated superior performance across the continent compared to other methods. QDM exhibited similar error performance to EQM but slightly higher errors in the southern region.

For correlation metrics (NSE and R^2), DQM-corrected precipitation had lower performance than other methods, although some grid cells in the central and eastern regions showed high performance, with values exceeding 0.995. The precipitation corrected using EQM showed the highest performance, especially in the central and eastern regions, where most grid points showed correlation coefficients above 0.995. QDM, while achieving correlation metrics above 0.978 for most grid points, had slightly lower performance than the other methods.

Regarding Pbias, all three methods tended to overestimate precipitation relative to the reference data across most grid points in North America, while corrected precipitation in Greenland was underestimated. For JSD, EVS, and KGE metrics, EQM-corrected precipitation showed the highest performance, with DQM and QDM performing lower than EQM.

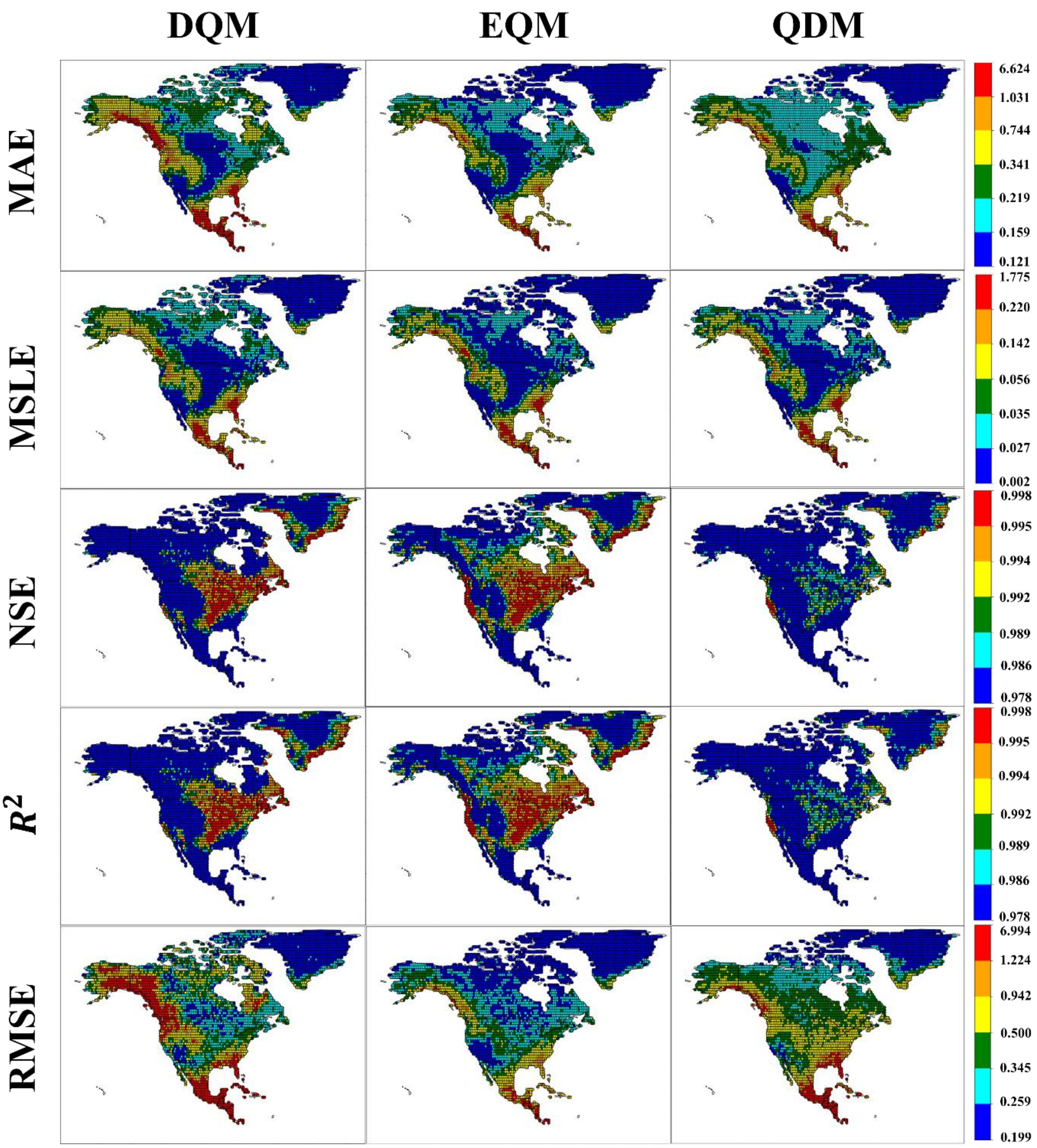

Figure 3. Performance comparison of DQM, EQM, and QDM using evaluation metrics (MAE, MSLE, NSE, R^2, and RMSE) for daily precipitation in North America.

In this study, the daily precipitation in Africa was corrected using three QM methods, and the performance is shown in Figures 4 and S3. Overall, the JSD of precipitation corrected by the three methods showed similar spatial patterns, but the precipitation of DQM showed lower performance than the other methods in the southern region. In terms of EVS, the precipitation of DQM showed higher variability than the other methods. The precipitation of QDM showed lower variability in southern Africa than DQM, but overall, it showed higher variability than EQM. The precipitation of EQM showed lower variability in southern and central Africa but still showed high variability in the northern region. Analyzing the error performance, the precipitation corrected by QDM showed the best performance compared to the other methods. In particular, QDM showed the highest performance in North Africa (MAE: 0.03, and MSLE: 0.004), and EQM's error performance was lower than QDM's in most indicators but better than DQM's. Finally, EQM performed the highest in correlation metrics (NSE and R^2), and QDM performed better than DQM.

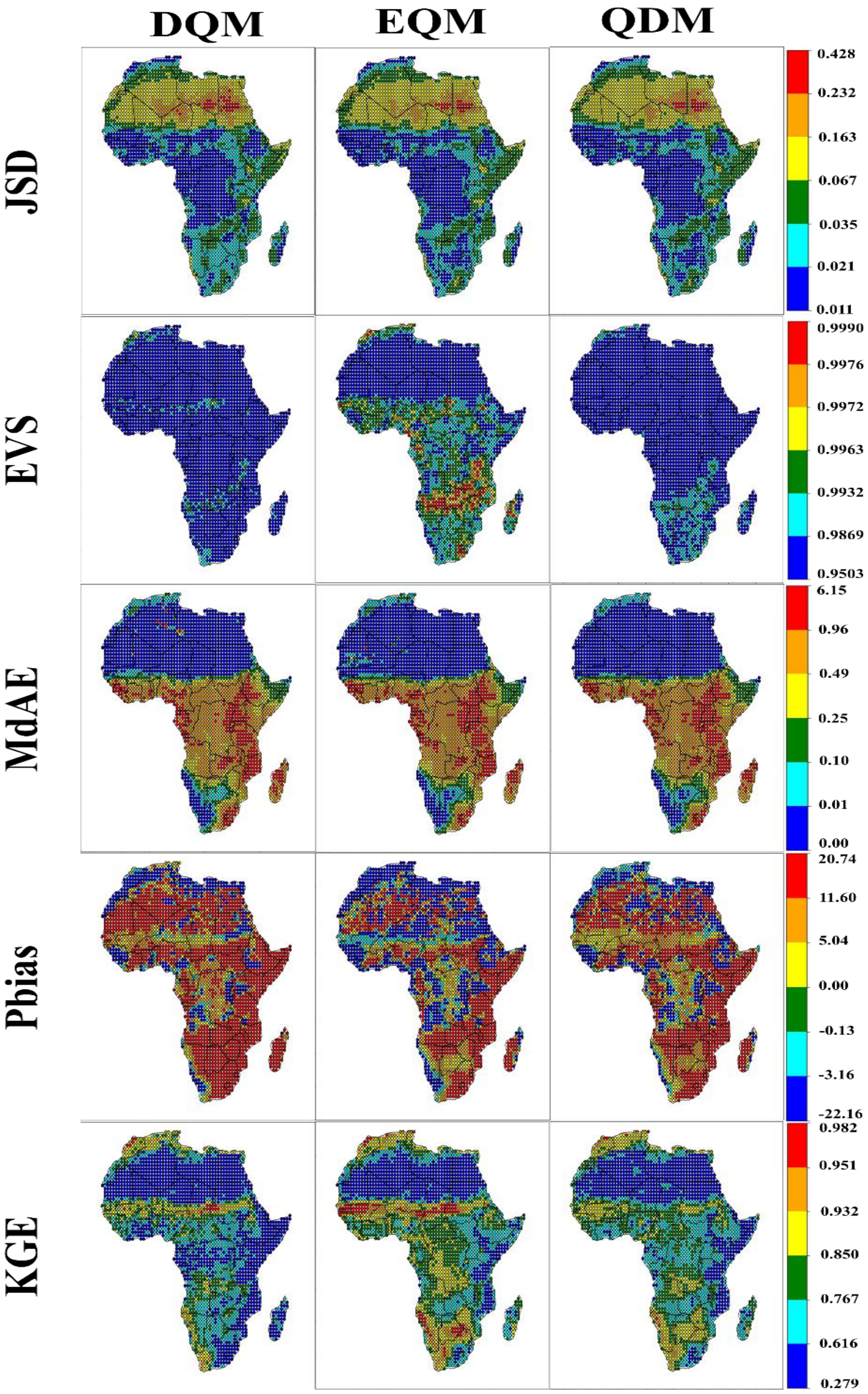

Figure 4. Performances of DQM, EQM, and QDM using evaluation metrics (JSD, EVS, MdAE, Pbias, and KGE) for daily precipitation in Africa.

Figures 5 and S4 show the spatial results of the grid-based evaluation metrics for the European region. In terms of error metrics, EQM-corrected precipitation performed the best across Europe compared to other methods. In contrast, QDM-corrected precipitation performed similarly to DQM in MAE and MSLE but significantly outperformed DQM in RMSE. Regarding NSE and R, EVS, and KGE metrics, EQM-corrected precipitation performed overwhelmingly better than other methods. QDM precipitation performed better than DQM, while DQM performed the worst. Regarding Pbias, EQM-corrected precipitation was underestimated compared to the reference data in most parts of Europe. In contrast, QDM-corrected precipitation was more similar to the reference data compared to other methods, and DQM precipitation was overestimated compared to the reference data except in central Europe.

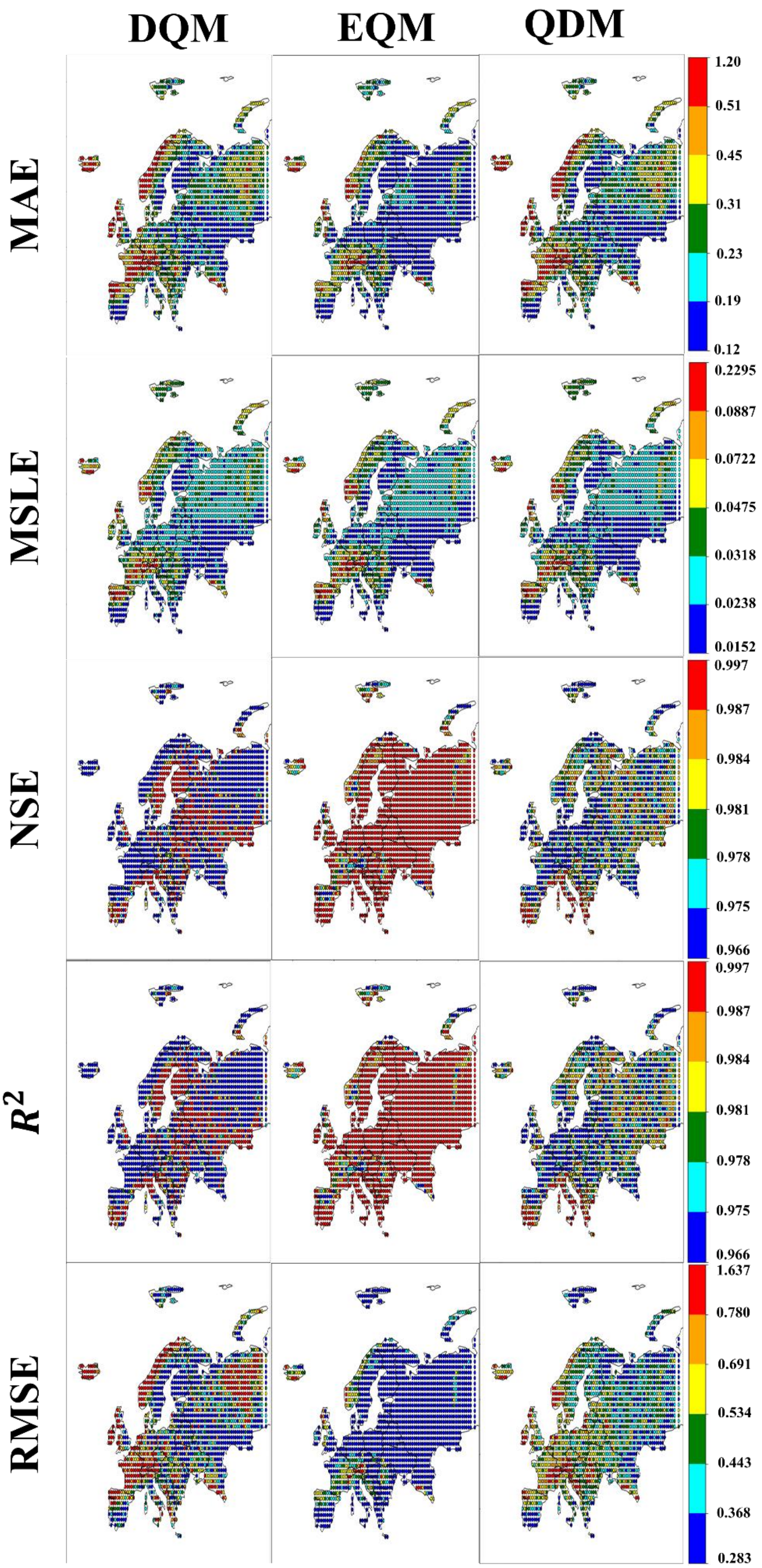
DQM
EQM
QDM
MAE
MSLE
NSE
R^2
RMSE
1.20
0.51
0.45
0.31
0.23
0.19
0.12
0.2295
0.0887
0.0722
0.0475
0.0318
0.0238
0.0152
0.997
0.987
0.984
0.981
0.978
0.975
0.966
0.997
0.987
0.984
0.981
0.978
0.975
0.966
1.637
0.780
0.691
0.534
0.443
0.368
0.283

Figure 5. Performances of DQM, EQM, and QDM using evaluation metrics (MAE, MSLE, NSE, R^2, and RMSE) for daily precipitation in Europe.

Figures 6 and S5 show the results of spatially quantifying the corrected precipitation in Asia using various evaluation metrics. Regarding error metrics, EQM-corrected precipitation stands out with its superior performance, particularly in RMSE, which was consistently below 1.35 in most areas except for certain parts of Central Asia. In contrast, DQM-corrected precipitation showed the poorest performance in error metrics. QDM-corrected precipitation demonstrated a performance similar to EQM but slightly lower in East Asia and North Asia. In NSE and R, the precipitation corrected by EQM performed better than other methods, especially in Southwest and East Asia. In contrast, the precipitation corrected by DQM performed lower than other methods. Regarding EVS, the precipitation corrected by EQM showed the lowest variability, while QDM showed higher variability than EQM but lower variability than DQM. In the case of Pbias, precipitation corrected by DQM was overestimated compared to the reference data throughout Asia. The precipitation corrected by EQM was underestimated in most regions except Central Asia. Precipitation in QDM showed a similar spatial pattern to that in EQM, but the range of Pbias was more diverse.

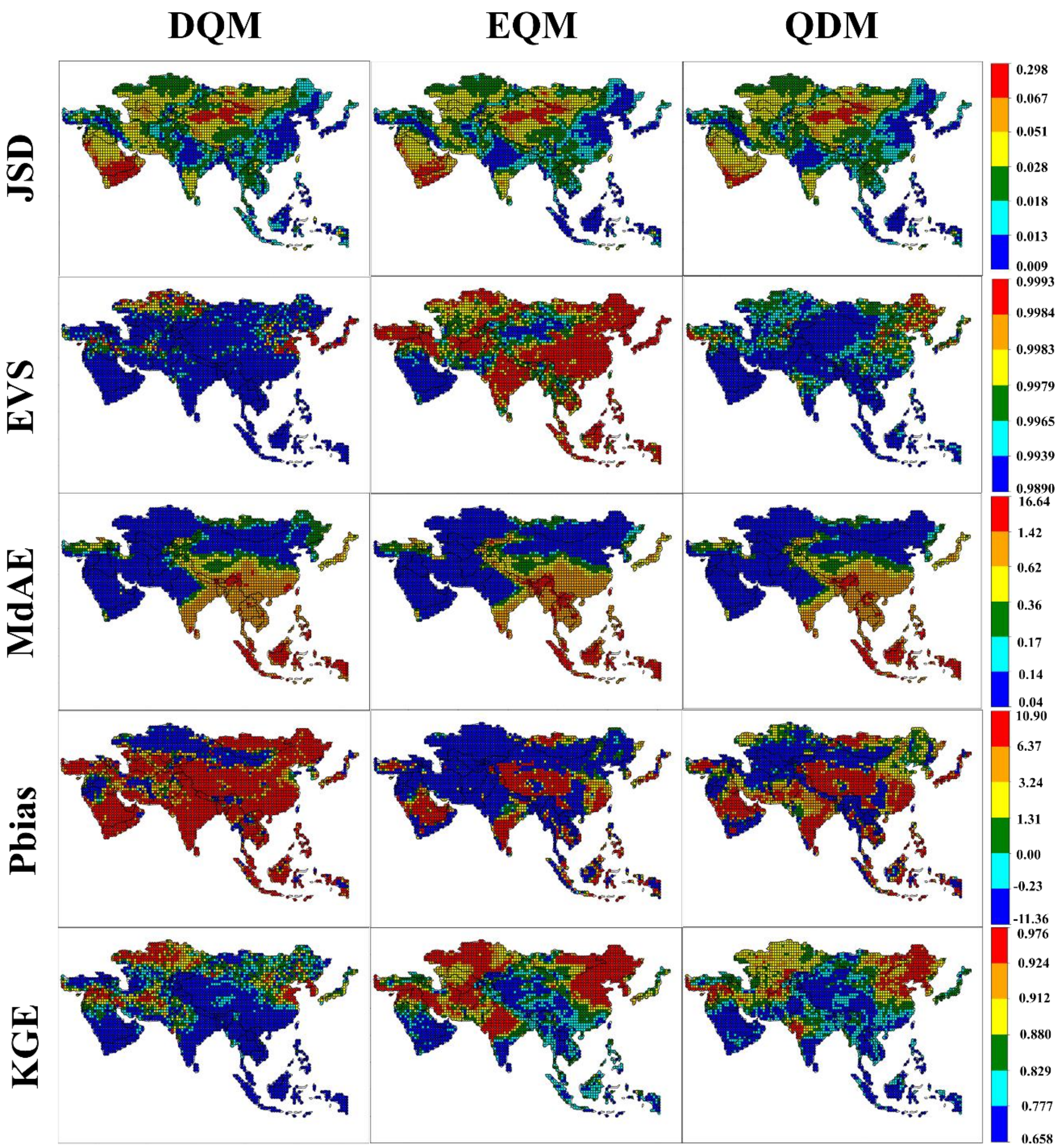

Figure 6. Performances of DQM, EQM, and QDM using evaluation metrics (JSD, EVS, MdAE, Pbias, and KGE) for daily precipitation in Asia.

Figures 7 and S6 show the results of spatially quantifying the corrected daily precipitation in Oceania using various evaluation metrics. In terms of error metrics, the precipitation estimated by the three QM methods performed similarly in MAE, MdAE, and MSLE. However, the precipitation corrected by EQM performed better in RMSE than the other methods. In the case of JSD, all three methods performed well.

Regarding EVS, the precipitation corrected by EQM showed lower variability than the other methods, and DQM showed higher performance than QDM. In Pbias, the precipitation adjusted by QDM was overestimated compared to the reference data in Oceania, while the precipitation corrected by DQM and EQM was underestimated compared to the reference data in central and southern Oceania. Finally, in KGE, precipitation corrected by EQM showed the highest performance, while DQM showed the lowest.

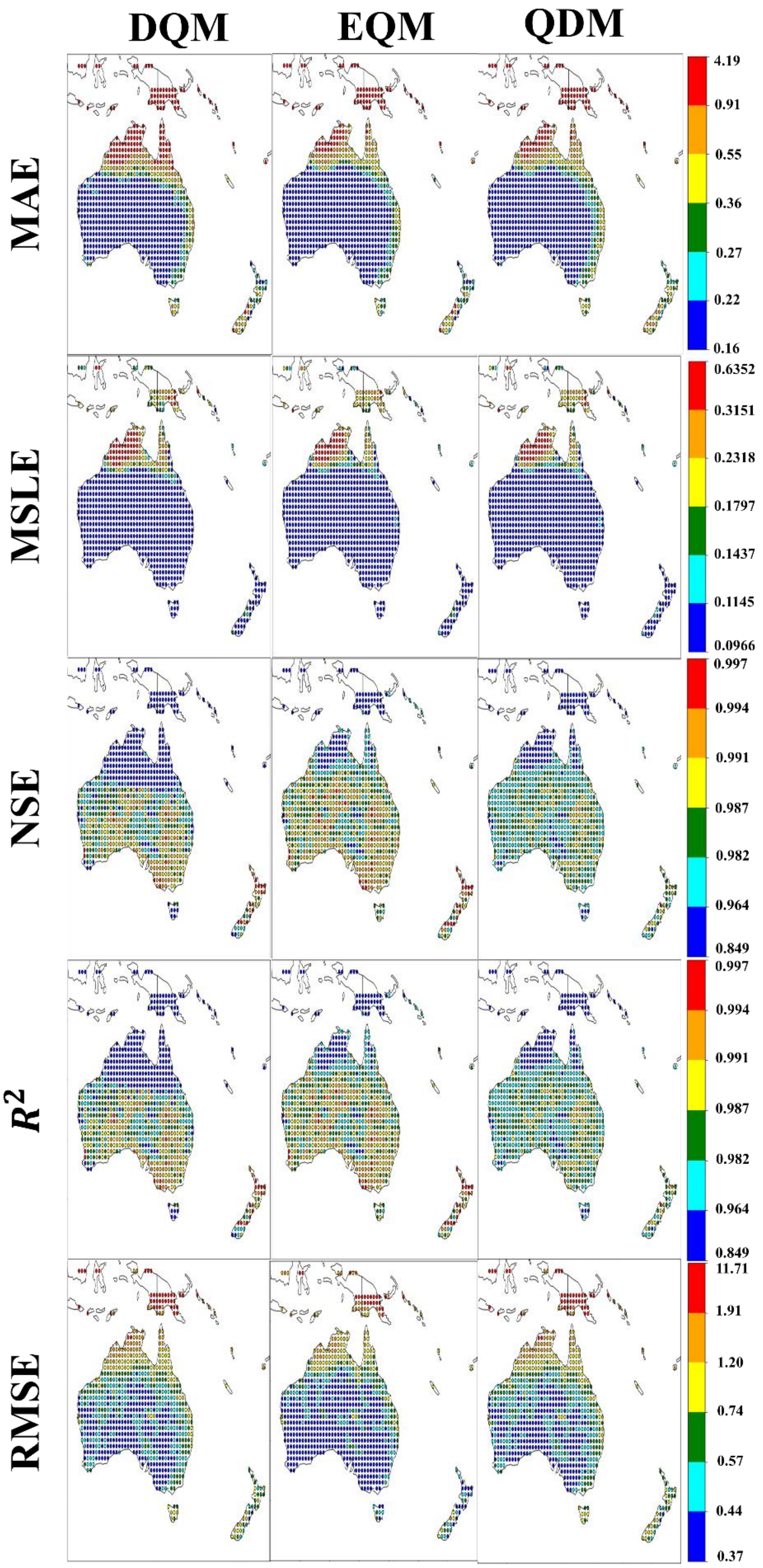

Figure 7. Performances of DQM, EQM, and QDM using evaluation metrics (MAE, MSLE, NSE, R^2, and RMSE) for daily precipitation in Asia.

Figure 8 visualizes the results of evaluating the bias-corrected precipitation data using 11 CMIP6 GCMs on six continents using ten evaluation metrics as boxplots. Overall, the precipitation corrected by EQM outperforms the other methods on most continents. In particular, EQM performs the best on the error metrics. QDM performs slightly lower than EQM but still maintains a high level of performance on all continents. On the other hand, DQM has more significant errors and relatively poor performance compared to the other methods on most metrics.

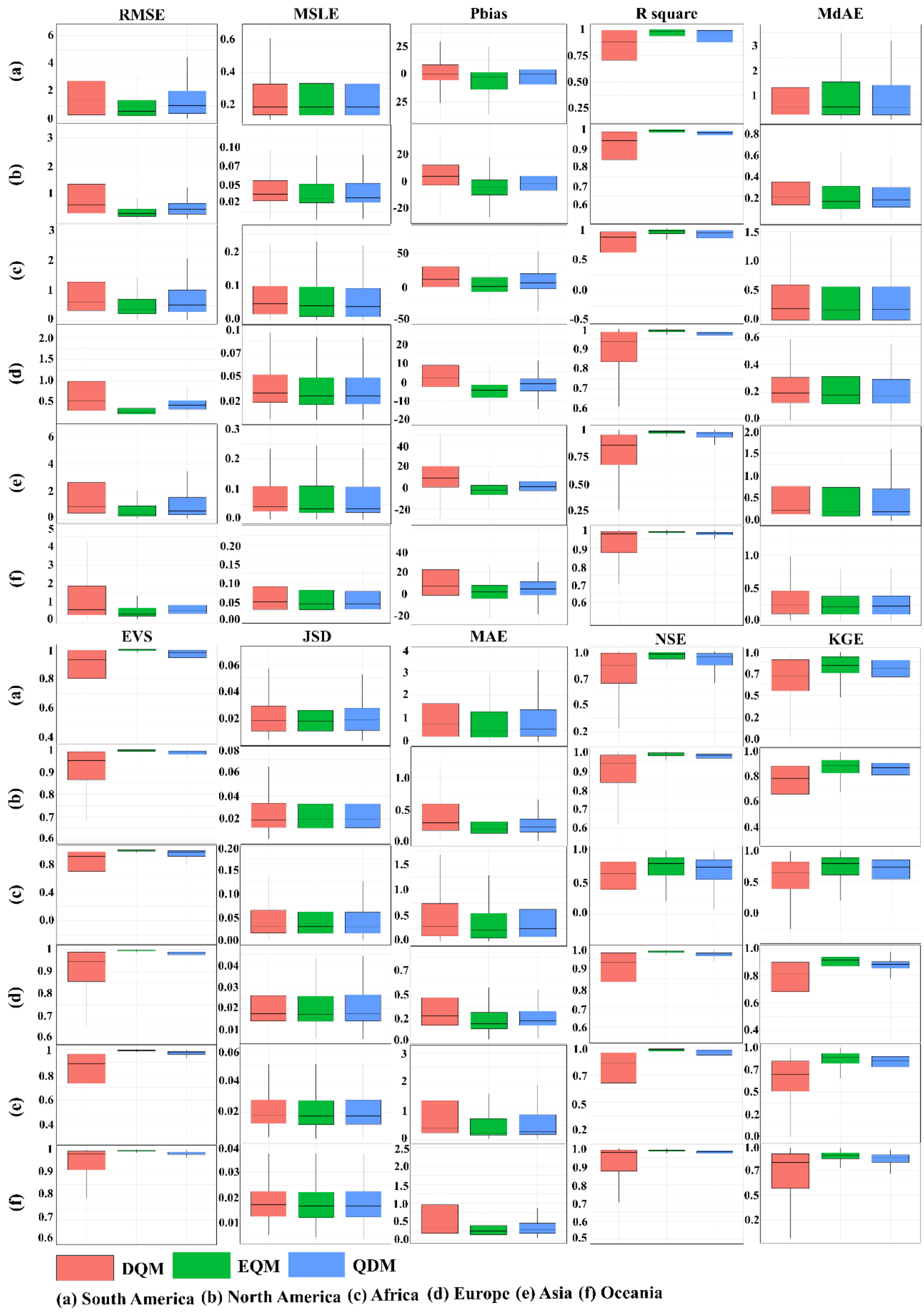

Figure 8. Performances of DQM, EQM, and QDM of historical period precipitation using boxplots based on ten evaluation metrics

3.1.3 Comparison of reproducibility for extreme daily precipitation

This study compared the daily extreme precipitation corrected by three methods using the GEV distribution. Figure 9 compares the distribution differences of the daily precipitation adjusted by the biased bias correction methods based on the GEV distribution using the JSD. In general, the JSD values for precipitation from DQM, EQM, and QDM are very low for most continents, indicating that the GEV distributions are almost identical among the three methods. Although there are some outliers, the overall distribution differences are not significant, suggesting little difference among the three methods when correcting for historical precipitation. However, in Europe, unlike other continents, the differences between the first and third quartiles of the JSD are relatively significant, indicating that the distributions can vary significantly from grid to grid depending on the QM method.

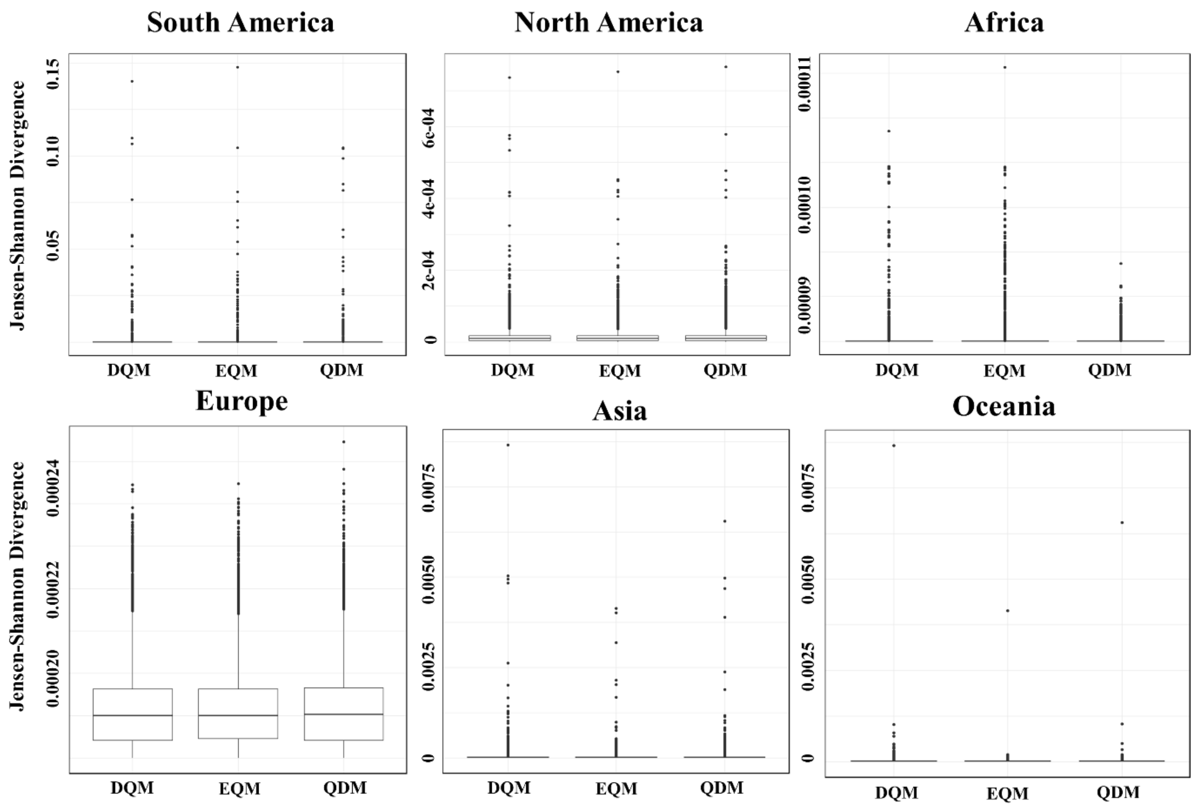

Figure 9. Comparison of distribution differences for GEV distribution using JSD across six continents.

Figure 10 shows the probability density functions for extreme precipitation above the 95th percentile of the GEV distribution. Overall, DQM shows the highest probability density for extreme precipitation across all continents and has the widest distribution, indicating that DQM corrects more extreme precipitation. On the other hand, EQM shows the lowest probability density and conservatively corrects for extreme precipitation. QDM shows probability densities between EQM and DQM across most continents but closer to EQM.

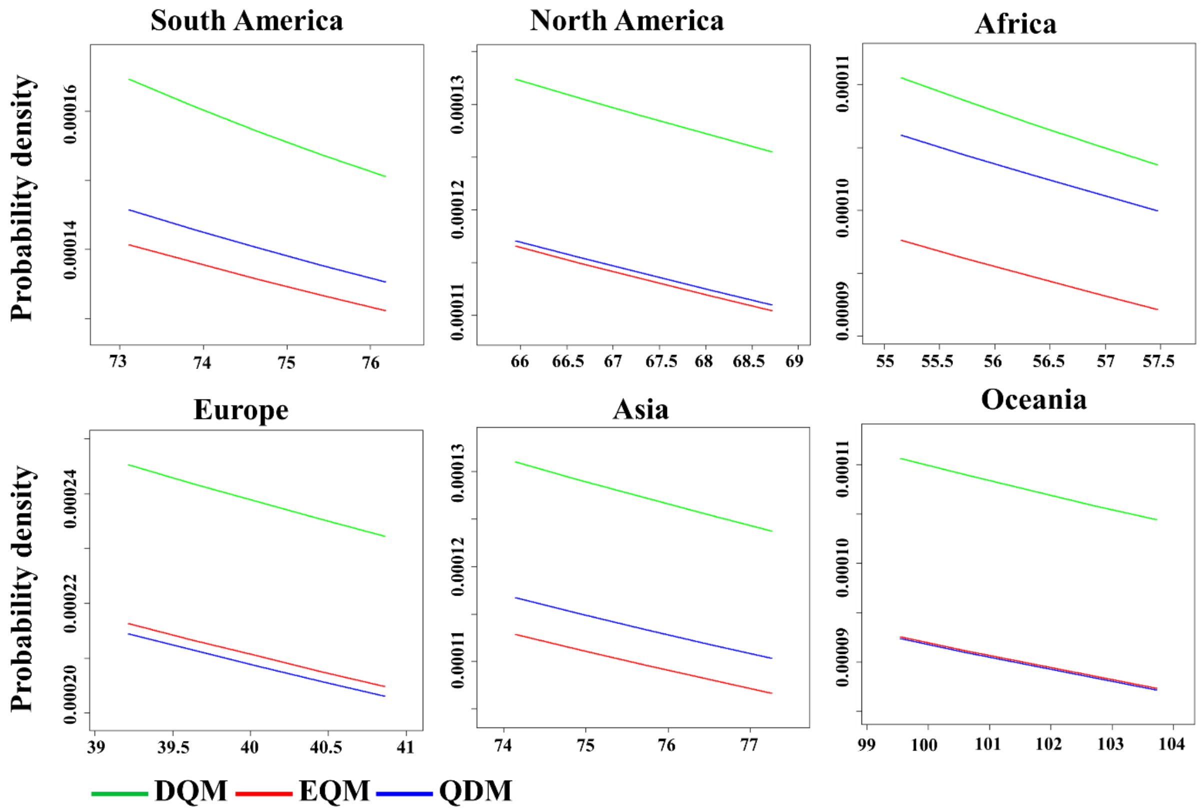

Figure 10. Comparison of probability densities for extreme precipitation values above the 95th percentile using GEV.

3.2 Prioritization of bias correction methods based on performance

3.2.1 Results of weight for evaluation metrics

In this study, the weights were calculated by applying entropy theory to the evaluation metrics used in the TOPSIS analysis, and the results are presented in Table 3. JSD had the highest weight in South America because the estimated JSD from 11 CMIP6 GCMs was an important metric for evaluating model performance differences. These results indicate that the differences between distributions are significant. On the other hand, EVS and NSE in South America had very low weights, suggesting that the variability and efficiency of precipitation were considered less important than other indicators. For North America, the RMSE, MSLE, and MAE metrics were of significant importance, as evidenced by their high weights. These error metrics revealed substantial regional differences. In contrast, EVS carried a negligible weight, suggesting it was less important in explaining variability in North America. For Africa, MdAE and JSD metrics were of considerable importance, as indicated by their high weights. These metrics were key evaluation factors in Africa. Conversely, EVS carried a low weight, suggesting it was considered relatively less important. RMSE had the highest weight in Europe, and KGE also had a relatively high weight, indicating that these metrics were considered important evaluation criteria in Europe. In Asia, MAE and MSLE had high weights, suggesting that these metrics were important evaluation metrics. On the other hand, EVS and NSE were considered less important due to their low variability. JSD, KGE, RMSE, and MAE were assigned high weights in Oceania, indicating that these metrics are essential factors. On the other hand, R^2 and NSE were assigned low weights.

Table 3. Entropy-based weights for evaluation metrics across different continents

	RMSE	MAE	R^2	NSE	KGE	Pbias	MdAE	MSLE	EVS	JSD
South America	0.1439	0.1536	0.0001	0.0001	0.0005	0.0238	0.1754	0.1934	0.0004	0.3088
North America	0.2289	0.1908	0.0001	0.0001	0.0007	0.0118	0.2152	0.2117	0.0001	0.1411
Africa	0.1319	0.1686	0.0002	0.0002	0.0002	0.0855	0.2436	0.1911	0.0002	0.1786
Europe	0.2821	0.1762	0.0022	0.0022	0.0063	0.0378	0.1754	0.1666	0.0021	0.1490
Asia	0.2073	0.1954	0.00003	0.00003	0.0001	0.0305	0.2300	0.2024	0.00003	0.1342
Oceania	0.2384	0.2204	0.0013	0.0013	0.0068	0.0214	0.2338	0.2093	0.0012	0.0660

3.2.2 Selection of the best bias correction method based on TOPSIS

Figures 11 and S7 present the best bias correction method selected for each continent using the TOPSIS approach. In Figure 11(a), the spatial distribution of the most effective bias correction method across the grid points of each continent is shown. In contrast, Figure 11(b) shows the number of grid points selected for each QM method. In South America, EQM was chosen as the best method in most grid points, with EQM being selected in over 1,500 grid points. In contrast, QDM was selected in fewer than 700 grid cells, making it the least chosen method in South America. Across all continents except South America, EQM was selected as the best model in the majority of grid cells, with the number of selected grid points (North America: 7,583; Africa: 2,879; Europe: 2,719; Asia: 8,793; and Oceania: 1,659). On the other hand, DQM was the least chosen method across all continents. For QDM, although it was the second most selected method across all continents except South America, the difference in the number of grid points between QDM and EQM is significant.

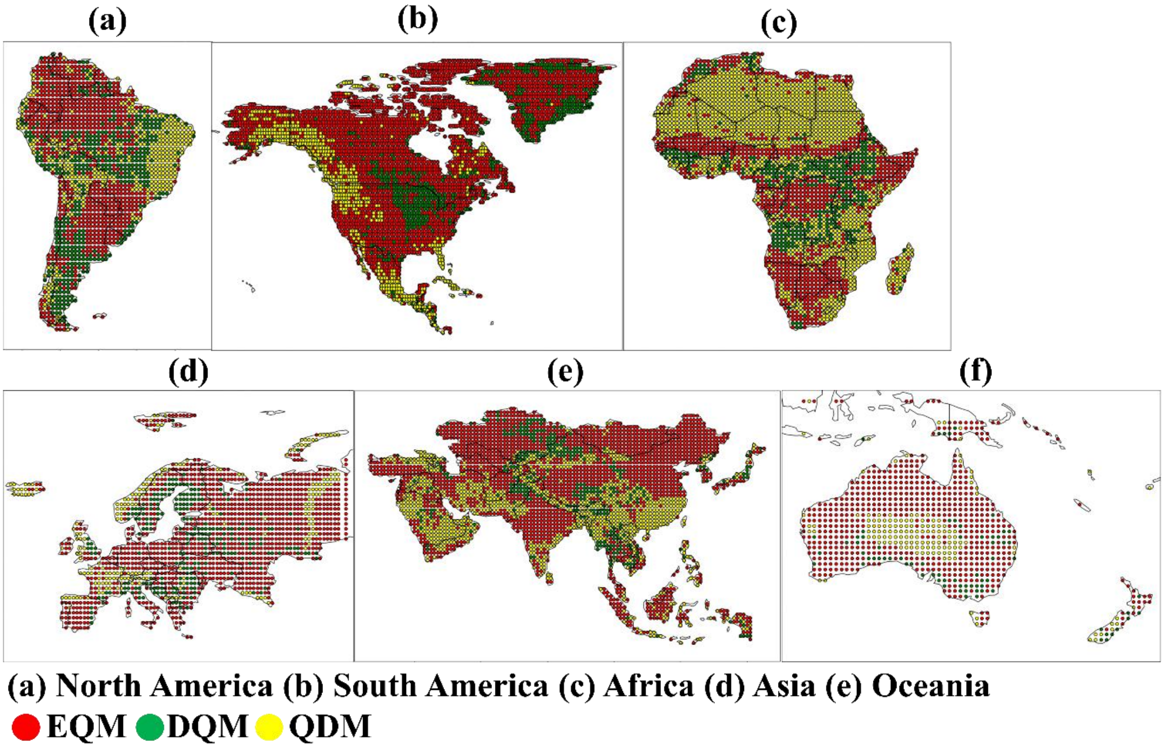

Figure 11 Spatial distribution for selected best bias correction methods across continents using TOPSIS

3.3 Uncertainty quantification of bias corrected daily precipitation

3.3.1 Uncertainty by model

This study quantifies the daily precipitation uncertainty of 11 CMIP6 GCMs, corrected using three different BMA methods. Figure 12 shows the distribution of GCM weight variances calculated by BMA across six continents. In South America, the highest weight variance was observed mainly in DQM. EQM showed high weight variance in the northern region but lower variance than DQM in most other regions. QDM exhibited the lowest weight variance, with values less than 0.00113 in most regions. In North America, EQM had the lowest weight variance, with values between 0.00055 and 0.00024 in most regions. QDM showed the lowest model uncertainty across North America, with more regions where weight variances were closer to 0 than the other methods. On the other hand, DQM exhibited high weight variance overall, with exceptionally high model uncertainty in the northeast and southern regions. In

Africa, EQM's weight variance was estimated to be low overall, resulting in low model uncertainty in most regions. For QDM, weight variance was low in some regions but higher than 0.00113 in others. DQM showed high weight variance in most regions except for the northern area, indicating high model uncertainty across the continent. EQM's weight variance was the lowest in Europe compared to the other methods, with weight variances close to 0 across the continent. QDM also showed low weight variance overall, though higher than EQM. DQM exhibited high weight variance in most regions except for Central Europe. In Asia, EQM showed low weight variance in most regions except Southeast Asia. QDM's weight variance was similar to EQM's, though some regions had higher model uncertainty. DQM showed high weight variance in most regions except for some Southwest and North Asian areas. For Oceania, the weight variances of EQM and DQM were mainly similar, but DQM showed a higher weight variance overall.

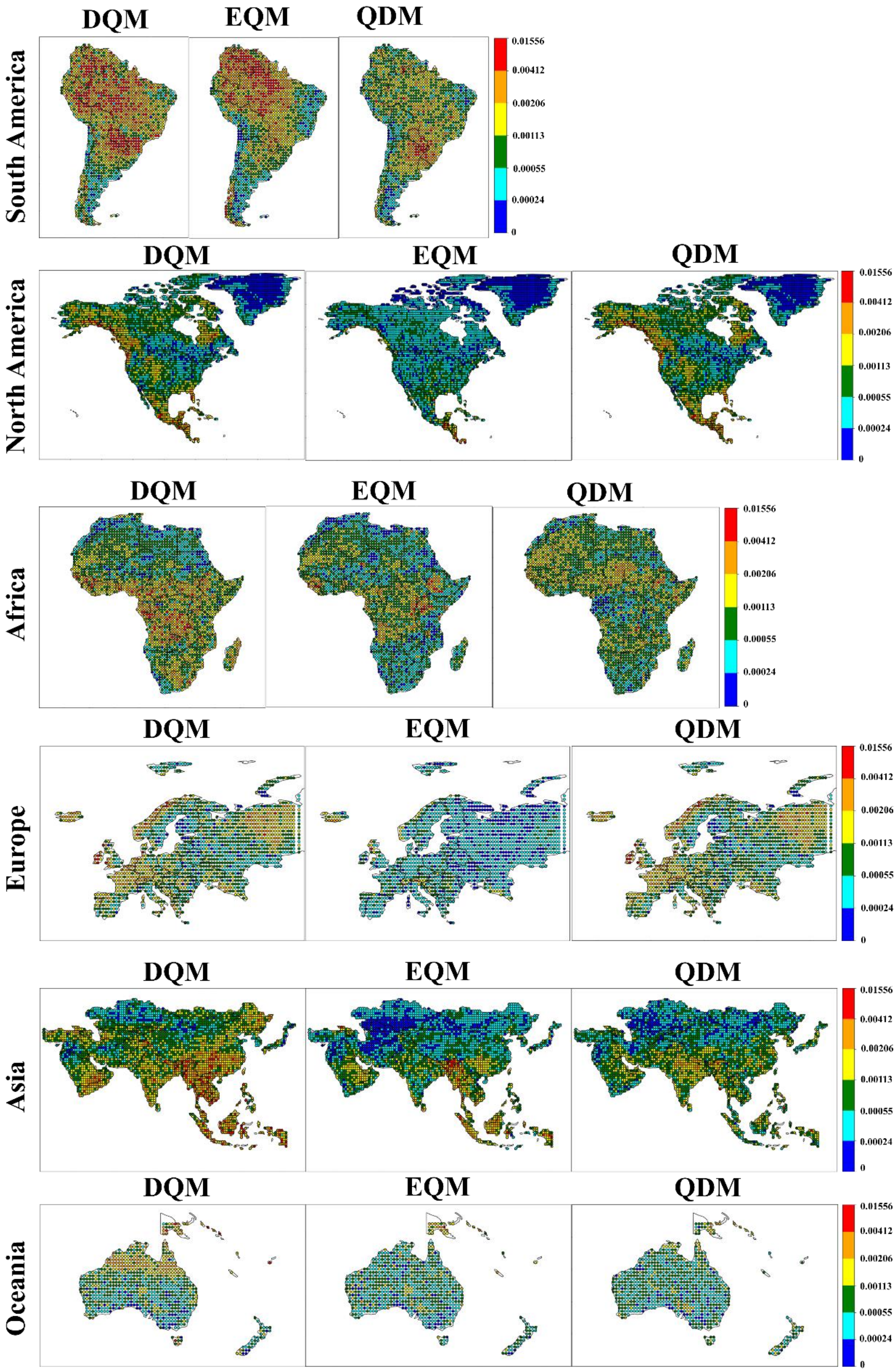

Figure 12. Spatial distribution of weight variance across continents for bias corrected CMIP6 GCMs using BMA

Figure 13 shows the distribution of GCM weight variances calculated using BMA across six continents, presented as boxplots. Overall, EQM has the smallest weight variance, and QDM has the second smallest weight variance on all continents except South America. In contrast, in South America, QDM has the smallest weight variance, and EQM has the second smallest. DQM consistently has the largest weight variance across all continents, indicating the highest model uncertainty.

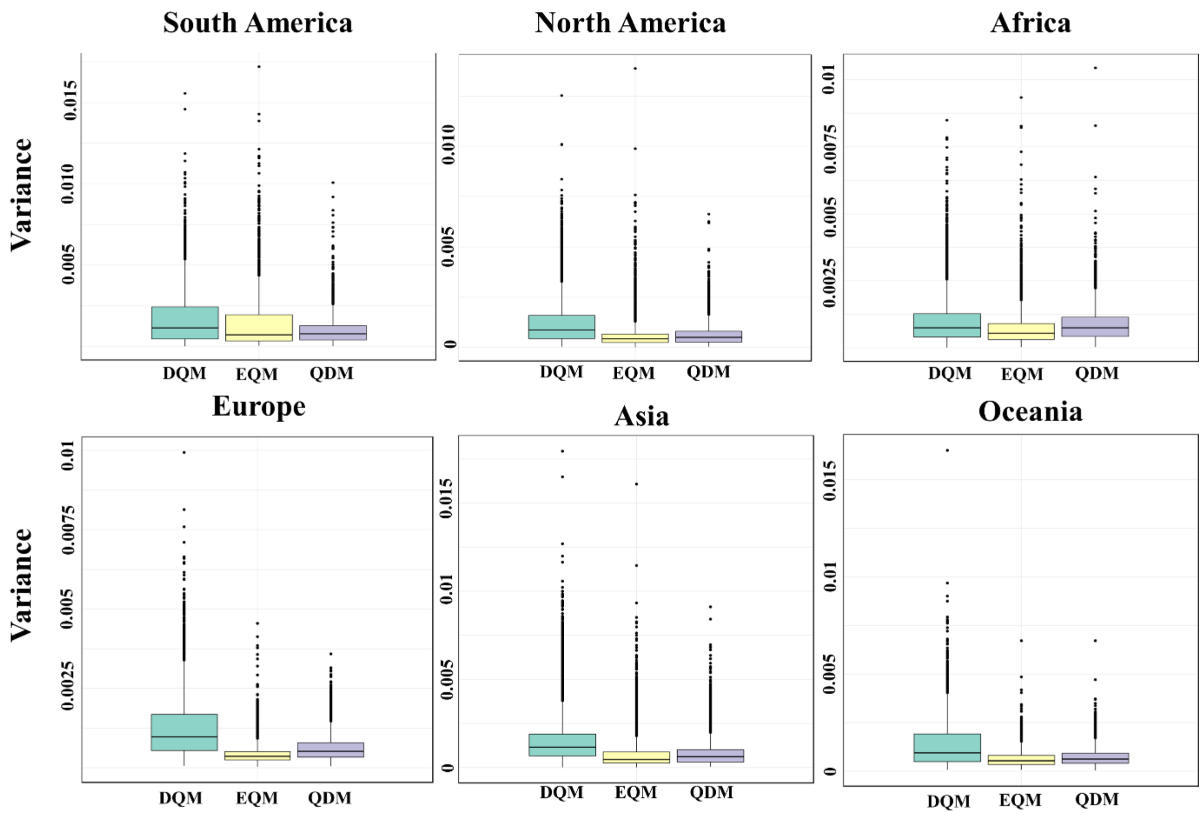

Figure 13. Weight variance for bias correction methods across six continents using box plots.

3.3.2 Uncertainty by ensemble prediction

This study developed a daily precipitation ensemble for the historical period based on 11 CMIP6 GCMs using BMA. Figure 14 shows the standard deviation of daily precipitation for the historical period by continent for the ensemble developed using BMA with 11 CMIP6 GCMs. Overall, the ensemble predicted using EQM provided stable precipitation projection with low standard deviations across most continents. The QDM ensemble showed similar results to EQM for most continents except Oceania, but the standard deviations were slightly higher. On the other hand, the ensemble using DQM exhibited higher standard deviations than the other methods for all continents and had the largest prediction uncertainty. In Oceania, the ensembles predicted by the three methods showed similar results. However, the prediction uncertainty was estimated to be lower in the order of EQM, DQM, and QDM due to slight differences.

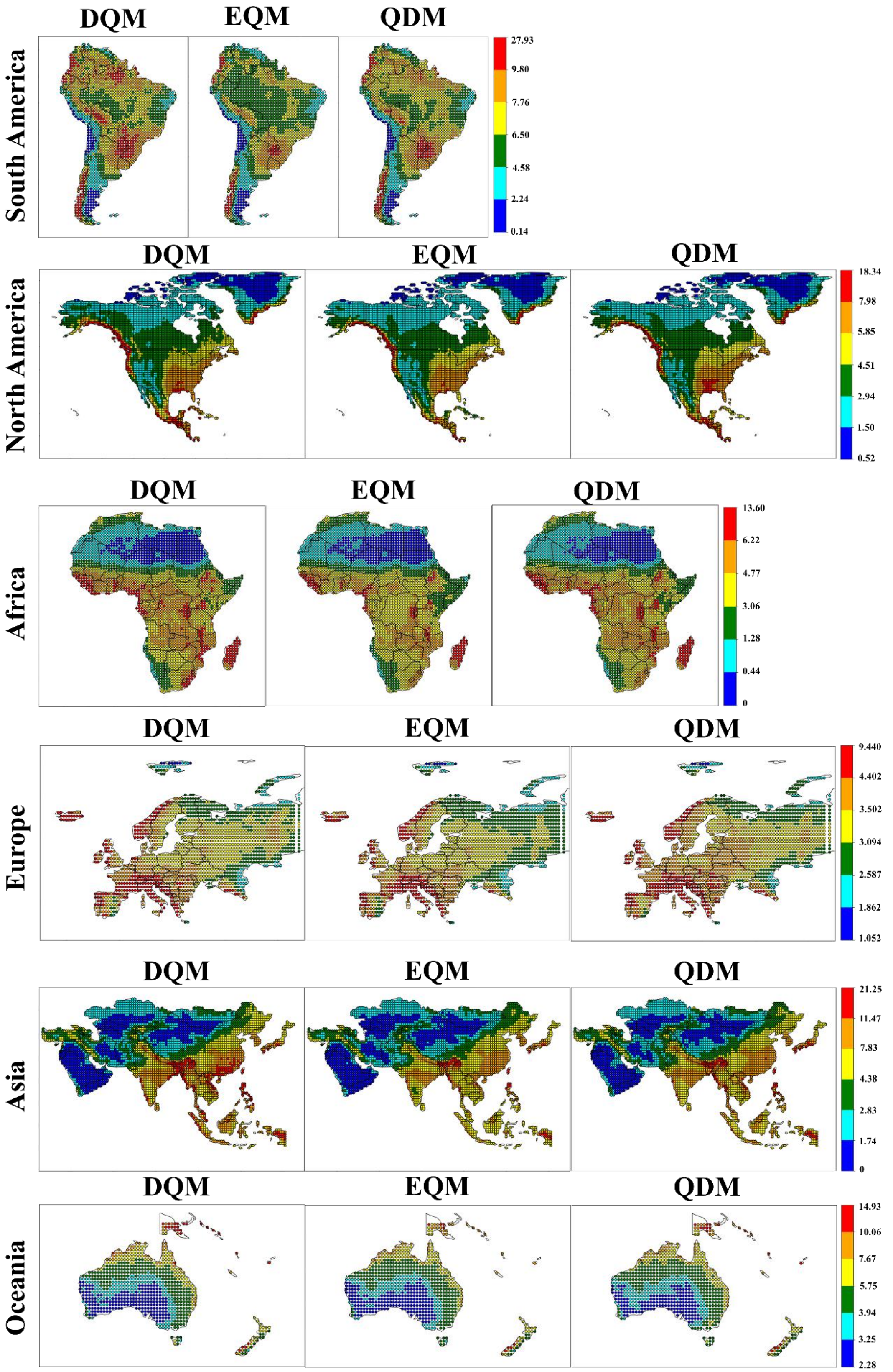

Figure 14. Spatial distribution of standard deviation for daily precipitation across continents for bias corrected CMIP6 GCMs using BMA

Figure 15 shows the standard deviation of daily precipitation for the ensemble forecasted by BMA using three methods, DQM, EQM, and QDM, in a boxplot for each continent. Overall, the EQM ensemble showed the lowest standard deviation across all continents, providing the most stable daily precipitation forecasts. The QDM ensemble showed slightly higher standard deviations than EQM for most continents, but there was no significant difference between the two methods. In contrast, the DQM ensemble showed the highest standard deviation and the largest prediction uncertainty.

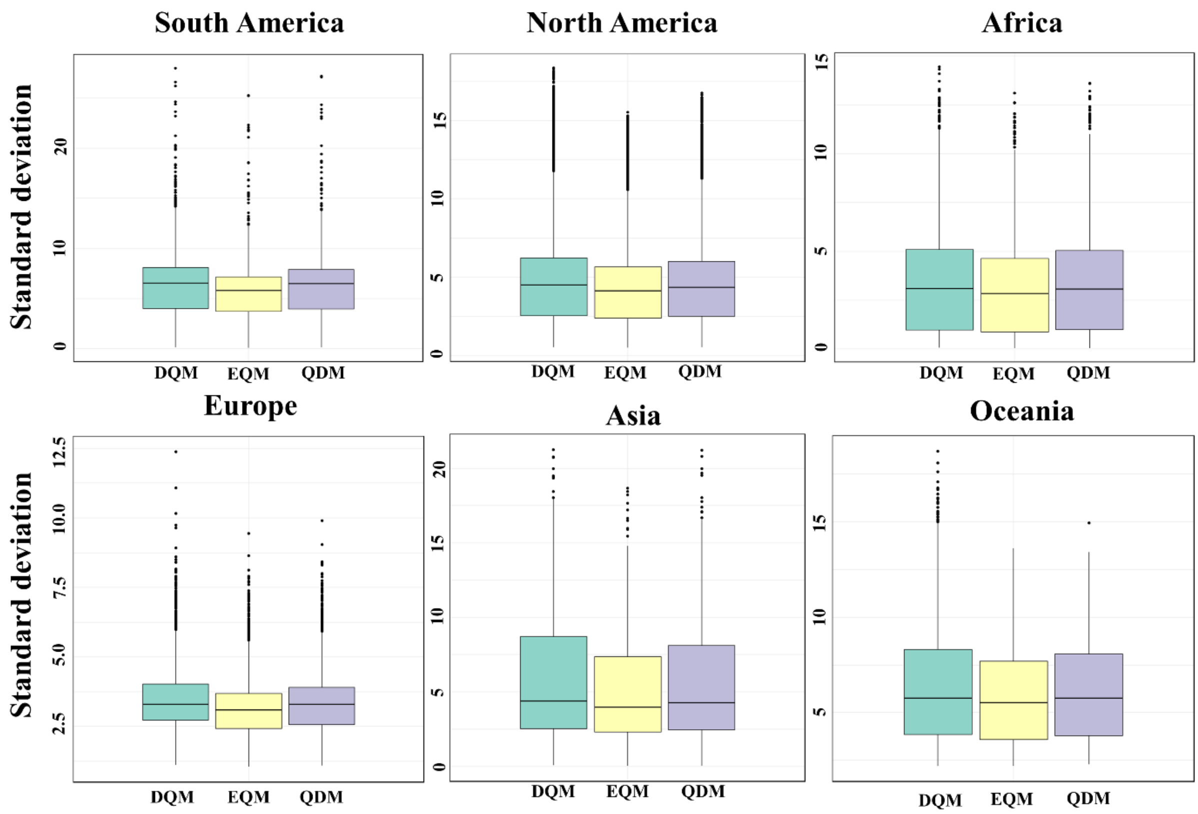

Figure 15. Spatial distribution of standard deviation for daily precipitation across continents for bias corrected CMIP6 GCMs using BMA

3.4 Evaluation of bias correction methods using CI

3.4.1 Results of CI by each weighting case

This study compared three QM methods by generating a CI based on three cases of weighting values that considered both model performance and uncertainty. Figures 16, S8, and S9 show the comprehensive indices calculated by applying equal weights and weights emphasizing performance and uncertainty, respectively.

EQM showed the highest CI across all continents when equal weights were applied. However, the index was lower in southern Europe and southeastern North America, but it calculated high values in most other regions. QDM showed high index values in some regions, although they were lower than those of EQM. For example, the CI results were high in the northern and western parts of North America and the central part of Europe. On the other hand, DQM was generally unsuitable in most regions but showed a relatively high index in Oceania.

When weights that emphasized performance were applied, DQM showed a high index in the central part of South America but low performance in most continents. Nevertheless, DQM showed a better index than QDM in some parts of Oceania. EQM showed the best index across most continents. While QDM was less suitable than EQM, it was still evaluated as a useful method in some continents.

Even when applying weights that increased the emphasis on uncertainty, similar results were obtained with the other weighting values. In particular, EQM was evaluated as the most suitable model across all continents, while DQM showed the opposite results.

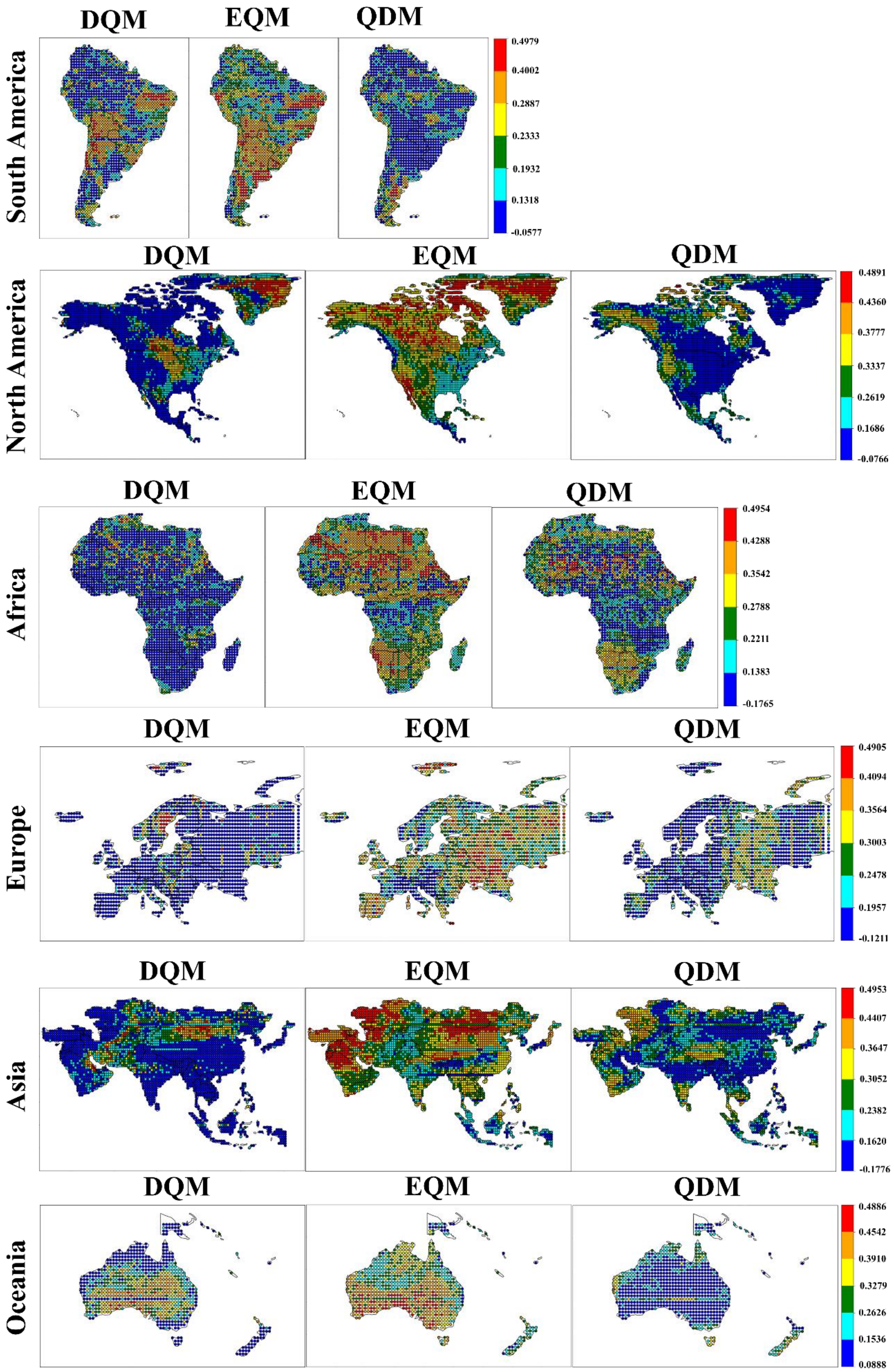
DQM
EQM
QDM
South America
0.4979
0.4002
0.2887
0.2333
0.1932
0.1318
-0.0577
North America
0.4891
0.4360
0.3777
0.3337
0.2619
0.1686
-0.0766
Africa
0.4954
0.4288
0.3542
0.2788
0.2211
0.1383
-0.1765
Europe
0.4905
0.4094
0.3564
0.3003
0.2478
0.1957
-0.1211
Asia
0.4953
0.4407
0.3647
0.3052
0.2382
0.1620
-0.1776
Oceania
0.4886
0.4542
0.3910
0.3279
0.2626
0.1536
0.0888

Figure 16. Spatial distribution of comprehensive indices for bias correction methods with equal weights (α: 0.5, β: 0.5) across continents

Figure 17 presents a comparison of the comprehensive indices for three QM methods with different weights for each continent using box plots. Overall, all methods showed higher indices than the other weighting values in the values that emphasized more weight on performance. In all weighted values, DQM showed the lowest indices in all continents except for South America and Oceania, where it was slightly higher or similar to QDM. EQM showed the best composite indices in all continents, outperforming performance and uncertainty. QDM showed high comprehensive indices in most continents, and the gap with EQM narrowed significantly in the weighting values that emphasized performance more. Nevertheless, QDM overall had lower comprehensive indices than EQM.

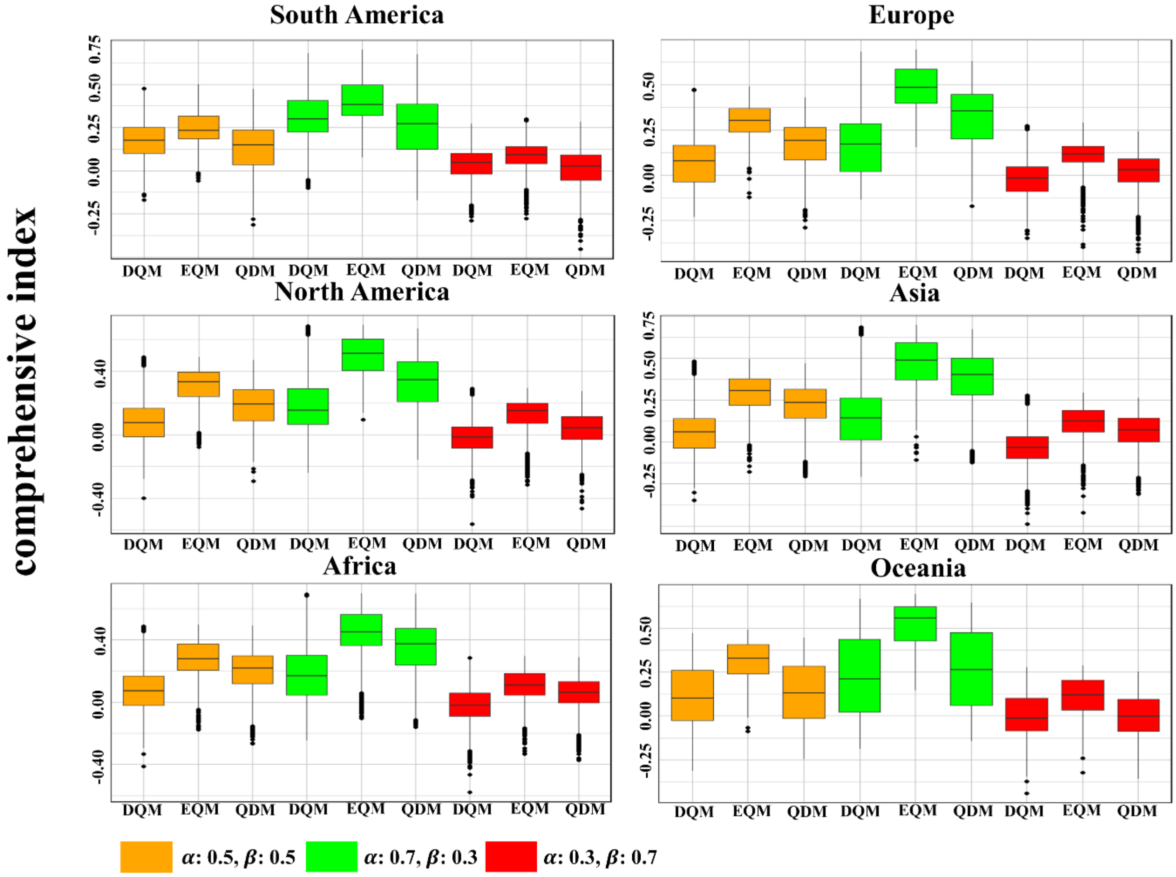

Figure 17. CI for three bias correction methods across continents with varying weights on performance and uncertainty

3.4.2 Selection of best bias correction method

Based on the CI, this study selected the best bias correction method for each continent. Figure 18 shows how the best bias correction method was selected for each continent by applying various weighting values of the CI. Overall, EQM was selected as the best correction method for most continents in all weighting values and was selected more than other methods in North America, Europe, Asia, and Oceania. DQM was selected the least in most continents except for South America and Oceania, and the number of selected grids tended to decrease as the weighting for uncertainty increased. QDM was selected as the proper bias correction method in western North America, southern and eastern Africa, and northern Europe. In addition, QDM was selected the most in Southeast Asia in all weighting values.

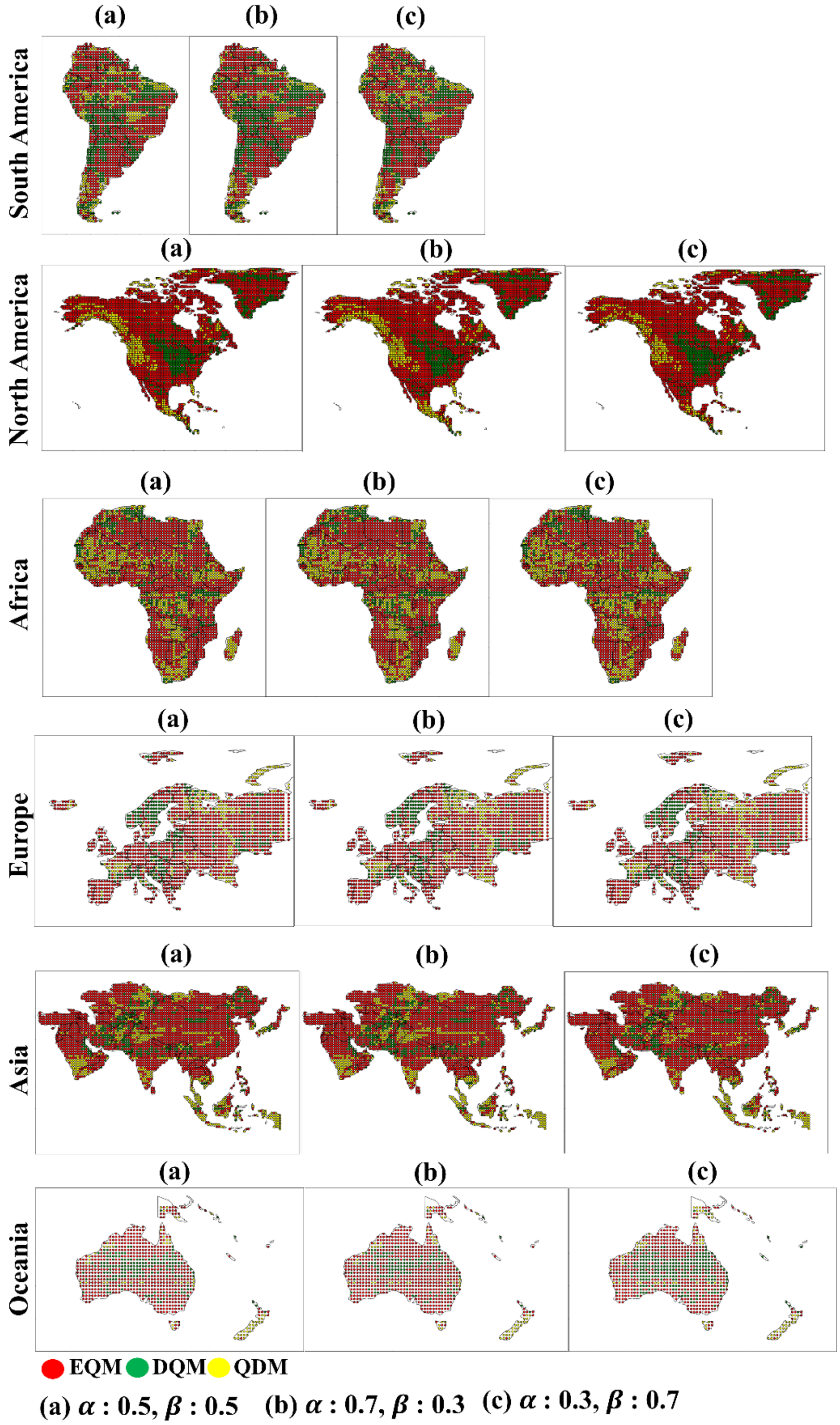

Figure 18. Selection of best bias correction methods across continents based on CI depending on weighting values.

Figure 19 shows the number of selected grids for the best bias correction method across continents based on three weighting values. Overall, EQM was the most frequently selected method across all weighting values, demonstrating superior performance across all continents compared to the other methods. Interestingly, as the weight for uncertainty increased, the number of grids where EQM was selected also increased, while the number decreased as the weight for performance increased. In contrast, QDM was chosen as the second-best method on most continents, except for South America and Oceania. The number of selected grids for QDM slightly increased as the performance weight increased. DQM was the least selected method across most continents, indicating that it was the least suitable overall.

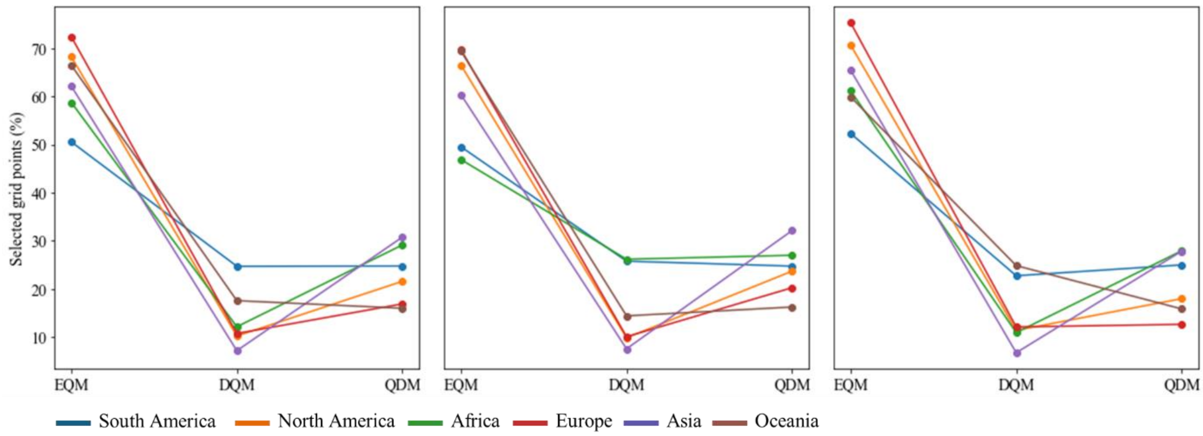

Figure 19. Ratios of selected grids for best bias correction methods across continents based on different weighting values

4. Discussion

Bias correction methods are widely used in correcting GCM outputs, and previous studies have compared the performance of various methods (Homsi et al., 2019; Saranya and Vinish, 2021). Among these, Quantile Mapping (QM) has consistently shown superior performance compared to other methods, making it a widely used approach for bias correction. In particular, QDM, EQM, and DQM, which are the focus of this study, are frequently employed in research exploring and applying climate change projections based on GCM outputs (Cannon et al., 2015; Switanek et al., 2016; Song et al., 2022a). Analyzing the strengths and limitations of these three methods will provide valuable insights for climate researchers, enabling them to choose the most suitable bias correction method for specific regions. In this context, this study further evaluates the performance of QDM, EQM, and DQM, especially for daily precipitation, and investigates how these methods perform across different regions. Unlike previous studies that focused on the performance of bias correction methods (Song et al., 2024a; Teutschbein and Seibert, 2012; Smitha et al., 2018), this study suggests a CI that integrates the performance and uncertainty metrics. This approach enhances the robustness of bias correction method selection and provides a more holistic evaluation framework. This section discusses the strengths and weaknesses of each method from various perspectives to provide a more balanced assessment.

4.1 Evaluation of bias correction methods performance

The daily precipitation corrected by the three QM methods outperformed the raw GCM data (see Figure 1). All three methods showed strong overall performance, as indicated by the Taylor diagram, producing consistently good results across different regions. This highlights the need to use multiple performance metrics to fully understand the strengths and weaknesses of the three QM methods, as relying on a single analysis or macroscopic perspective can overlook important details. From this perspective, many studies have emphasized the application of a multifaceted analysis in selecting bias correction methods (Homsi et al., 2019; Cannon et al., 2015; Berg et al., 2022; Song et al., 2023). The spatial distribution of correction performance, as discussed in Section 3.1.2, varies significantly by continent. Figures 2 to 7 reveal that the evaluated metrics differ across continents, underscoring the importance of region-specific correction methods. This finding aligns with Song et al. (2023), highlighting the importance of selecting appropriate correction methods based on the precipitation distribution at observation sites. Moreover, studies such as Homsi et al. (2019) and Saranya and Vinish (2021) also emphasize the variability in bias correction performance depending on the regional climate and data characteristics, reinforcing the need for tailored approaches. Of course, the three QM methods showed high performance across most continents, effectively correcting the biases in daily precipitation from GCMs. However, the corrected daily precipitation varies subtly among the three methods, with these differences becoming more pronounced in extreme events or specific evaluation metrics. For example, the three QM methods tend to perform less effectively in regions with high precipitation, but their performance also varies by grid (e.g., southern India in Asia: RMSE; central Oceania: Pbias and EVS; central Europe: Pbias, MdAE, and KGE). While EQM performs well across most continents, DQM and QDM show superior results in specific regions. Similar results were made by Cannon et al. (2015), which highlighted differences in the performance of bias correction methods, particularly in handling extreme precipitation events. QDM's error-related metrics (South America: RMSE, MAE, and MSLE) are nearly identical to EQM's, yet QDM outperforms EQM regarding MdAE on more grids. These findings suggest that a more nuanced and detailed analysis of precipitation corrected by GCMs is necessary, aligning with the conclusions of Gudmundsson et al. (2012), which emphasize that the effectiveness of bias correction methods can vary significantly depending on local climate characteristics, highlighting the importance of selecting appropriate methods for each region. These results suggest a more detailed precipitation analysis from corrected GCMs is needed.

This study compared the three QM methods for daily precipitation events above the 95th percentile (extreme precipitation) using the GEV distribution, as shown in Figure 10. The results indicate that DQM tends to correct more extreme precipitation events than QDM, aligning with previous findings that DQM captures a broader range of extremes. The unique characteristics of DQM caused these results. DQM overestimated the corrected extreme precipitation due to the relative variability in the data introduced through detrending, and the subsequent reintroduction of the long-term mean during the correction step widened the range of extreme precipitation, leading to overestimation compared to the reference data in areas with high variability. At the same time, QDM and EQM take a more conservative approach (as noted in previous studies such as Cannon et al., 2015). These findings suggest that EQM and QDM may be more suitable in regions vulnerable to floods and extreme weather events that require a more balanced and cautious approach. However, when comparing the differences in GEV distributions, there was no significant difference between methods in regions like Oceania and Europe (see Figure 9). These results imply that EQM can better handle extreme values or outliers in the data by directly comparing and correcting past and future distributions. In particular, EQM is consistent with previous studies in that it more accurately corrects observed distributions in non-stationary and highly variable climate variables, such as precipitation (Themeßl et al., 2012; Maraun, 2013; Gudmundsson et al., 2012). These positive aspects are mainly due to EQM's ability to align the empirical ECDFs of reference and model data across all quantiles, allowing it to correct biases with high precision at both central tendencies and extremes. Although there are significant advantages in observing the results of the correction method in detail from various perspectives, presenting these results without integrating them into a reasonable framework can increase confusion and uncertainty in climate change research

(Wu et al., 2022). Therefore, it is essential to introduce a structured framework such as MCDA to provide a single integrated result.

4.2 Uncertainties of model and ensemble prediction in bias correction methods

In climate modeling, quantifying uncertainty is essential to assess the reliability of bias-corrected precipitation data. This study applied BMA to quantify the uncertainty of three QM methods on a continental basis, addressing both model-specific and ensemble prediction uncertainties. Similar to the findings by Cannon et al. (2015), this analysis demonstrates how different bias correction methods yield varying uncertainty levels based on the underlying climate models. Notably, EQM showed the lowest weight variance across most continents, which means that the inter-model uncertainty for 11 GCMs corrected by EQM is lower than that of the other QM methods. The low uncertainty associated with EQM aligns with previous studies like Themeßl et al. (2012), which found that EQM consistently reduced discrepancies between modeled and observed data across regions. EQM's ability to manage extreme precipitation and anomalous values based on observed distributions contributes to its reliability, a feature also emphasized by Gudmundsson et al. (2012). On the other hand, DQM showed the highest weight variance across all continents, indicating more significant uncertainty when applied to various GCMs. This uncertainty was particularly pronounced in regions with complex climate conditions, such as Southeast Asia, East Africa, and the Alps in Europe. These results align with Berg et al. (2022), who highlighted DQM's limitations in capturing long-term climate trends and extreme events. The higher uncertainty associated with DQM suggests that, while its detrending process is effective in correcting the mean, it may struggle in regions dominated by nonlinear climate patterns, as it does not sufficiently account for all quantiles in the distribution, particularly extremes, as noted by Cannon et al. (2015). QDM, though showing lower weight variance than DQM, still demonstrated higher uncertainty than EQM in regions with diverse climate characteristics. These results are consistent with the study of Tong et al. (2021), suggesting that QDM performs better under moderate precipitation scenarios. However, the uncertainty may increase under highly variable or extreme weather conditions. Furthermore, this study extended the uncertainty analysis to ensemble predictions, calculating the standard deviation of daily precipitation for each continent using BMA. The EQM-based ensemble consistently exhibited low standard deviations across all continents, indicating that EQM offers the most stable and reliable precipitation predictions. This finding echoes the conclusions drawn by Teng et al. (2015), where EQM provided more accurate and less uncertain projections. In contrast, DQM presented the most significant prediction uncertainty, reinforcing the need for caution when applying DQM in studies that require high-confidence data. These results emphasize the importance of weighing performance and uncertainty when choosing a suitable bias correction method. EQM's consistent performance in reducing uncertainty across model-specific and ensemble forecasts highlights its robustness as a preferred choice for climate research. However, the substantial uncertainty associated with DQM suggests that its use should be limited to regions where its detrending process can be beneficial. Overall, these findings stress the critical role of uncertainty quantification in climate change impact assessments and underscore the need for selecting bias correction methods based on a comprehensive evaluation of both performance and uncertainty.

4.3 Integrated assessment of bias correction methods

This study selected the optimal QM method for each continent based on the CI, which considers uncertainty and performance. The critical point is that uncertainty is decisive when selecting a bias correction method. As shown in Figure 19, the optimal correction method varies depending on the continent, and the selected method also changes depending on the weight. These results suggest that uncertainty still exists, as Berg et al. (2022) pointed out, and that uncertainty must be considered when selecting the optimal method. In other words, even if the QM method has high performance, it is difficult to make a reasonable selection if the uncertainty contained in the method is significant. Overall, EQM showed the highest CI value in all continents, which means that it provides the most balanced results in terms of performance and uncertainty. These results are consistent with previous studies (Lafon et al., 2013; Teutschbein and Seibert, 2012; Teng et al., 2015) that showed high precipitation correction accuracy and excellent performance, especially under complex climate conditions. QDM was evaluated highly in some regions but performed worse than EQM overall. Berg et al. (2022) also pointed out that QDM is superior in general climate conditions but may perform worse in extreme climate situations, suggesting that this may increase the uncertainty of QDM in extreme climates. DQM was evaluated as an unsuitable method in most regions due to low CI values, which is consistent with the limitations of DQM mentioned in Cannon et al. (2015) and Berg et al. (2022). It was confirmed that DQM performs relatively well in dry climates but may perform worse in various climate conditions. In addition, some differences were observed with the results based on

TOPSIS. For example, DQM was selected more than QDM in South America, but when the uncertainty weight was applied, QDM was selected more. Conversely, in Oceania, QDM was selected more than DQM, but when the uncertainty weight was increased to 0.7, DQM was selected more. These results are consistent with those of Lafferty and Sriver (2023), showing that when significant uncertainty exists, uncertainty can be greater despite high bias correction performance. In conclusion, EQM is the most balanced method regarding performance and uncertainty and will likely be preferred in future climate modeling studies. However, there may be more suitable QM methods depending on the region, and a comprehensive evaluation with various weights is needed. Therefore, when establishing climate change response strategies or policy decisions, it is essential to take a multifaceted approach that considers uncertainty together rather than relying on a single indicator or performance alone. It will enable more reliable predictions and better decision-making.

5. Conclusion

This study corrected and compared historical daily precipitation from 11 CMIP6 GCMs using three QM methods. Eleven statistical metrics were used to evaluate the precipitation performance corrected by three QM methods, and TOPSIS was applied to select performance-based priorities. BMA was applied to quantify model-specific and ensemble prediction uncertainties. Additionally, suitable QM methods were selected and compared using a CI that integrates TOPSIS performance scores with BMA uncertainty metrics. The conclusions of this study are as follows:

1. EQM showed the highest overall index across all continents, indicating that it provides the most balanced approach in terms of performance and uncertainty.
2. DQM effectively reproduced the dry climate in North Africa and parts of Central and Southwest Asia but showed the highest uncertainty across all continents. These results suggest that DQM may lose some long-term trend information, making it less reliable in regions prone to extreme weather events.
3. QDM performed better in certain regions, such as Southeast Asia, and was selected more often than DQM when uncertainty was given greater weight. QDM may be a promising alternative in areas where uncertainty plays a significant role.

4. Selecting an appropriate QM is required for high performance, and significant uncertainty can complicate rational decision-making. Therefore, a multifaceted approach considering performance and uncertainty is essential in climate modeling.

In conclusion, EQM has emerged as the preferred method due to its balanced performance, but this study emphasizes the importance of regional assessment and careful consideration of uncertainty when selecting a QM method. Future research should integrate greenhouse gas scenarios to improve the accuracy of climate predictions and provide a more comprehensive understanding of future climate risks. Based on the results of this study, future studies can develop hybrid methodologies that combine the strengths of each QM.

Code availability

Codes for benchmarking the xclim of python package are available from https://doi.org/10.5281/zenodo.10685050 (Bourgault et al., 2024). Furthermore, the CI proposed in this study, along with the TOPSIS and BMA used within it, is available at https://doi.org/ 10.5281/zenodo.14351816 (Song, 2024b).

Data availability

The data used in this study are publicly available from multiple sources. CMIP6 General Circulation Models (GCMs) outputs were obtained from the Earth System Grid Federation (ESGF) data portal at https://esgf-node.llnl.gov/search/cmip6/. Users can select data types such as climate variables, time series, and experiment ID, which can be downloaded as NC files. Furthermore, CMIP6 GCMs output can also be accessed in Eyring et al. (2016) The ERA5 reanalysis dataset used in this study is available through the Copernicus Data Store (CDS) provided by ECMWF (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview). ERA5 is available at https://doi.org/10.24381/cds.bd0915c6 (Hersbach et al., 2023). The daily precipitation datasets from CMIP6 GCM and ERA5 used in this study are available at https://doi.org/10.6084/m9.figshare.27999167.v5 (Song, 2024c).

Author contributions

Young Hoon Song: Conceptualization, Methodology, Data curation, Funding acquisition, Visualization, Writing – original draft, Writing – review & editing. Eun Sung Chung: Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Validation, Writing-review & editing

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study was supported by National Research Foundation of Korea (NRF) (RS-2023-00246767_2; 2021R1A2C200569914)

Reference

1. Abdelmoaty, H.M., and Papalexiou, S.M.: Changes of Extreme Precipitation in CMIP6 Projections: Should We Use Stationary or Nonstationary Models? J. Clim. 36(9), 2999-3014, https://doi.org/10.1175/JCLI-D-22-0467.1, 2023.
2. Ansari, R., Casanueva, A., Liaqat, M.U., and Grossi, G.: Evaluation of bias correction methods for a multivariate drought index: case study of the Upper Jhelum Basin. GMD 16(7), 2055-2076, https://doi.org/10.5194/gmd-16-2055-2023, 2023.
3. Berg P., Bosshard, T., Yang, W., and Zimmermann, K.: MIdASv0.2.1 – MultI-scale bias AdjuStment. GMD 15, 6165-6180, https://doi.org/10.5194/gmd-15-6165-2022, 2022
4. Bourgault, P., Huard, D., Smith, T.J., Logan, T., Aoun, A., Lavoie, J., Dupuis, É., Rondeau-Genesse, G., Alegre, R., Barnes, C., Beaupré Laperrière, A., Biner, S., Caron, D., Ehbrecht, C., Fyke, J., Keel, T., Labonté, M.P., Lierhammer, L., Low, J.F., Quinn, J., Roy, P., Squire, D., Stephens, Ag., Tanguy, M., Whelan, C., Braun, M., Castro, D.: xclim: xarray-based climate data analytics (0.48.1). Zenodo [Code], https://doi.org/10.5281/zenodo.10685050, 2024.
5. Cannon, A. J., Sobie, S. R., and Murdock, T.Q.: Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? J. Clim. 28(17), 6938-6959, https://doi.org/10.1175/JCLI-D-14-00754.1, 2015.
6. Cannon, A.J.: Multivariate quantile mapping bias correction: an N-dimensional

probability density function transform for climate model simulations of multiple variables. Clim. Dyn. 50, 31–49. https://doi.org/10.1007/s00382-017-3580-6, 2018.

7. Chae, S. T., Chung, E. S., and Jiang, J.: Robust siting of permeable pavement in highly urbanized watersheds considering climate change using a combination of fuzzy-TOPSIS and the VIKOR method. Water Resour. Manag. 36(3), 951–969, https://doi.org/10.1007/s11269-022-03062-y, 2022.
8. Chua, Z.W., Kuleshov, Y., Watkins, A.B., Choy, S., and Sun, C.: A Comparison of Various Correction and Blending Techniques for Creating an Improved Satellite-Gauge Rainfall Dataset over Australia. Remote Sens, 14(2), 261, https://doi.org/10.3390/rs14020261, 2022.
9. Chung, E. S., and Kim, Y.J.: Development of fuzzy multi-criteria approach to prioritize locations of treated wastewater use considering climate change scenarios. JEM 146, 505–516, https://doi.org/10.1016/j.jenvman.2014.08.013, 2014.
10. Cox, P., and Stephenson, D.: A changing climate for prediction. Science 317(5835), 207–208, https://www.science.org/doi/10.1126/science.1145956, 2007.
11. Deser, C., Phillips, A., Bourdette, V., and Teng, H.: Uncertainty in climate change projections: the role of internal variability. Clim. Dyn. 38, 527–546, https://doi.org/10.1007/s00382-010-0977-x, 2012
12. Déqué, M.: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. Glob. Planet. Change. 57(1-2), 16-26, https://doi.org/10.1016/j.gloplacha.2006.11.030, 2007.
13. Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions "Should we apply bias correction to global and regional climate model data?". HESS 16(9), 3391-3404, https://doi.org/10.5194/hess-16-3391-2012, 2012.
14. Enayati, M., Bozorg-Haddad, O., Bazrafshan, J., Hejabi, S., and Chu, X.: Bias correction capabilities of quantile mapping methods for rainfall and temperature variables. Water and Climate change 12(2), 401-419, https://doi.org/10.2166/wcc.2020.261, 2021.
15. Evin, G., Ribes, A., and Corre, L.: Assessing CMIP6 uncertainties at global warming levels. Clim Dyn. https://doi.org/10.1007/s00382-024-07323-x, 2024.
16. Eyring, V., Bony, S., Meehl, G., Senior, C., Stevens, B., Stouffer, R., and Taylor, K.:

Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geoscientific Model Development, 9(5), 1937–1958. 2016. https://doi.org/10.5194/gmd-9-1937-2016

17. Galton, F.: Regression Towards Mediocrity in Hereditary Stature. The Journal of the Anthropological Institute of Great Britain and Ireland 15, 246-263, https://doi.org/10.2307/2841583, 1886.
18. Giorgi, F., and Mearns, L.O.: Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "reliability ensemble averaging" (REA) method, J. Clim. 15, 1141–1158, https://doi.org/10.1175/1520-0442(2002)015<1141:COAURA>2.0.CO;2, 2000.
19. Gupta, H.V., Kling, H., Yilmaz, K.K., and Martinez, G.F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. 377(1–2), 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009
20. Gudmundsson, L., Bremnes, J.B., Haugen, J.E., and Engen-Skaugen, T.: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods. HESS 16(9), 3383–3390, https://doi.org/10.5194/hess-16-3383-2012, 2012.
21. Hamed, M.M., Nashwan, M.S., Shahid, S., Wang, X.J., Ismail, T.B., Dewan, A., and Asaduzzaman, M.d: Future Köppen-Geiger climate zones over Southeast Asia using CMIP6 Multimodel Ensemble. Atmos. Res. 283(1), 106560, https://doi.org/10.1016/j.atmosres.2022.106560, 2023.
22. Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Q. J. Roy. Meteor. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.
23. Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C.,

Dee, D., and Thépaut, J.-N.: ERA5 hourly data on pressure levels from 1940 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://doi.org/10.24381/cds.bd0915c6, 2023.

24. Homsi, R., Shiru, M. S., Shahid, S., Ismail, T., Harun, S. B., Al-Ansari, N., and Yaseen, Z.M: Precipitation projection using a CMIP5 GCM ensemble model: a regional investigation of Syria. Eng. Appl. Comput. Fluid Mech. 14(1), 90–106, https://doi.org/10.1080/19942060.2019.1683076, 2019.
25. Hoeting J.A., Madigan D., Raftery A.E., and Volinsky C.T.: BayesIan model averaging: A tutorial (with discussion). Stat. Sci. 214, 382-417, https://doi.org/10.1214/ss/1009212519, 1999.
26. Hosking, J.R.M., Wallis, J.R., and Wood, E.F.: Estimation of the generalized extreme value distribution by the method of probability weighted monents. Technometrics 27, 251–261, https://doi.org/10.1080/00401706.1985.10488049, 1985.
27. Hosking, J.R.M.: L-moments: Analysis and estimation of distributions using linear combinations of order statistics. J. R. Stat. 52**,** 105–124, https://doi.org/10.1111/j.2517-6161.1990.tb01775.x, 1990.
28. Hwang, C. L., and Yoon, K.: Multiple attribute decision making: Methods and applications. Springer-Verlag. https://doi.org/10.1007/978-3-642-48318-9. 1981.
29. IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, https://doi.org/10.1017/9781009157896, 2021.
30. IPCC: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Pörtner, H.-O., Roberts, D. C., Tignor, M., Poloczanska, E. S., Mintenbeck, K., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V., Okem, A., and Rama, B., Cambridge University Press, https://doi.org/10.1017/9781009325844, 2022.

31. Ishizaki, N.N., Shiogama, H., Hanasaki, N., Takahashi, K., and Nakaegawa, T.: Evaluation of the spatial characteristics of climate scenarios based on statistical and dynamical downscaling for impact assessments in Japan. Int. J. Climatol. 43(2), 1179-1192, https://doi.org/10.1002/joc.7903, 2022.
32. Jobst, A.M., Kingston, D.G., Cullen, N.J., and Schmid, J.: Intercomparison of different uncertainty sources in hydrological climate change projections for an alpine catchment (upper Clutha River, New Zealand). HESS 22, 3125-3142, https://doi.org/10.5194/hess-22-3125-2018, 2018.
33. Lafferty, D.C., and Sriver, R.L.: Downscaling and bias-correction contribute considerable uncertainty to local climate projections in CMIP6. npj Clim Atmos Sci 6, 158, https://doi.org/10.1038/s41612-023-00486-0, 2023.
34. Lafon, T., Dadson, S., Buys, G., and Prudhomme, C.: Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. Int. J. Climatol. 33, 1367-1381, http://dx.doi.org/10.1002/joc.3518, 2013.
35. Lin, J.: Divergence measures based on the Shannon entropy. IEEE Transactions on Information Theory 37(1), 145–151, https://doi.org/10.1109/18.61115, 1991.
36. Maraun, D.: Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. J. Clim. 26(6), 2137-2143, https://doi.org/10.1175/JCLI-D-12-00821.1, 2013.
37. Nair, M.M.A., Rajesh, N., Sahai, A.K., and Lakshmi Kumar, T.V.: Quantification of uncertainties in projections of extreme daily precipitation simulated by CMIP6 GCMs over homogeneous regions of India. Int. J. Climatol. 43(15), 7365-7380, https://doi.org/10.1002/joc.8269, 2023.
38. Nash, J.E., and Sutcliffe, J.V.: River flow forecasting through conceptual models part I—A discussion of principles. J. Hydrol. 10**,** 282–290, https://doi.org/10.1016/0022-1694(70)90255-6Return to ref 1970 in article, 1970.
39. Pathak, R., Dasari, H.P., Ashok, K., and Hoteit, I., Effects of multi-observations uncertainty and models similarity on climate change projections. npj clim. atmos. sci. 6, 144, https://doi.org/10.1038/s41612-023-00473-5, 2023.
40. Piani, C., Weedon, G. P., Best, M., Gomes, S. M., Viterbo, P., Hagemann, S., and Haerter, J.O.: Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. J. Hydrol. 395(3-4), 199-215,

https://doi.org/10.1016/j.jhydrol.2010.10.024, 2010.

41. Rahimi, R., Tavakol-Davani, H., and Nasseri, M.: An Uncertainty-Based Regional Comparative Analysis on the Performance of Different Bias Correction Methods in Statistical Downscaling of Precipitation. Water Resour. Manag. 35, 2503–2518, https://doi.org/10.1007/s11269-021-02844-0, 2021.
42. Rajulapati, C.R., and Papalexiou, S.M.: Precipitation Bias Correction: A Novel Semi-parametric Quantile Mapping Method. Earth Space Sci. 10(4), e2023EA002823, https://doi.org/10.1029/2023EA002823, 2023.
43. Saranya, M.S., and Vinish, V.N.: Evaluation and selection of CORDEX-SA datasets and bias correction methods for a hydrological impact study in a humid tropical river basin, Kerala. Water Climate Change 12(8), 3688-3713, https://doi.org/10.2166/wcc.2021.139, 2021.
44. Shanmugam, M., Lim, S., Hosan, M.L. Shrestha, S., Babel, M.S., and Virdis, S.G.P.: Lapse rate-adjusted bias correction for CMIP6 GCM precipitation data: An application to the Monsoon Asia Region. Environ Monit Assess. 196, 49, https://doi.org/10.1007/s10661-023-12187-5, 2024.
45. Smitha, P.S., Narasimhan, B., Sudheer K.P., and Annamalai, H.: An improved bias correction method of daily rainfall data using a sliding window technique for climate change impact assessment. J. Hydrol. 556, 100-118. https://doi.org/10.1016/j.jhydrol.2017.11.010, 2018
46. Song, J. Y., and Chung, E.S.: Robustness, uncertainty, and sensitivity analyses of TOPSIS method to climate change vulnerability: Case of flood damage. Water Resour. Manag., 30(13), 4751–4771, https://doi.org/10.1007/s11269-016-1451-2, 2016.
47. Song, Y.H., Shahid, S., and Chung, E.S.: Differences in multi-model ensembles of CMIP5 and CMIP6 projections for future droughts in South Korea. Int. J. Climatol. 42(5), 2688-2716, https://doi.org/10.1002/joc.7386, 2022a.
48. Song, Y.H., Chung, E.S., and Shahid, S.: The New Bias Correction Method for Daily Extremes Precipitation over South Korea using CMIP6 GCMs. Water Resour. Manag. 36, 5977–5997, https://doi.org/10.1007/s11269-022-03338-3, 2022b.
49. Song, Y.H., Chung, E.S., and Shahid, S.: Uncertainties in evapotranspiration projections associated with estimation methods and CMIP6 GCMs for South Korea.

Sci. Total Environ. 825, 153953, https://doi.org/10.1016/j.scitotenv.2022.153953, 2023.

50. Song, Y.H., Chung, E.S., and Shahid, S.: Global Future Climate Signal by Latitudes Using CMIP6 GCMs. Earths Future 12(3), e2022EF003183, https://doi.org/10.1029/2022EF003183, 2024a.

51. Song, Y.H.: Comprehensive Index and Performance-Related Code, Zenodo [Code], https://zenodo.org/records/14351816. 2024b

52. Song, Y.H.: Historical Daily Precipitation Data of CMIP6 GCMs and ERA5, Figshare [Dataset], https://doi.org/10.6084/m9.figshare.27999167.v5. 2024c

53. Switanek, M.B., Troch, P.A., Castro, C.L., Leuprecht, A., Chang, H.I., Mukherjee, R., and Demaria E.M.C.: Scaled distribution mapping: a bias correction method that preserves raw climate model projected changes. HESS 21(6), 2649-2666, https://doi.org/10.5194/hess-21-2649-2017, 2017.

54. Tanimu, B., Bello, AA.D., Abdullahi, S.A. Ajibike, M.A., Yaseen, Z.M., Kamruzzaman, M., Muhammad, M.K.I., and Shahid, S.: Comparison of conventional and machine learning methods for bias correcting CMIP6 rainfall and temperature in Nigeria. Theor. Appl. Climatol. 155, 4423–4452, https://doi.org/10.1007/s00704-024-04888-9, 2024.

55. Teutschbein, C., and Seibert, J.: Bias correction of regional climate model simulations for 575 hydrological climate-change impact studies: Review and evaluation of different 576 methods. J. Hydrol. 16, 12-29, http://dx.doi.org/10.1016/j.jhydrol.2012.05.052, 2012.

56. Teng, J., Potter, N. J., Chiew, F. H. S., Zhang, L., Wang, B., Vaze, J., and Evans, J.P.: 2015. How does bias correction of regional climate model precipitation affect modelled runoff? HESS 19, 711–728, https://doi.org/10.5194/hess-19-711-2015, 2015.

57. Themeßl, M.J., Gobiet, A., and Heinrich, G.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. Int. J. Climatol. 31(10), 1530-1544, https://doi.org/10.1002/joc.2168, 2012.

58. Tong, Y., Gao, X., Han, Z., Xu, Y., and Giorgi, F.: Bias correction of temperature and precipitation over China for RCM simulations using the QM and QDM methods. Clim. Dyn. 57, 1425-1443, https://doi.org/10.1007/s00382-020-05447-4, 2021.

59. Yip, S., Ferro, C.A.T., Stephenson, D.B., and Hawkins, E.: A simple, coherent framework for partitioning uncertainty in climate predictions. J. Clim. 24(17), 4634–4643, https://doi.org/10.1175/2011JCLI4085.1, 2011.
60. Woldemeskel, F. M., Sharma, A. Sivakumar, B., and Mehrotra, R.: A framework to quantify GCM uncertainties for use in impact assessment studies. J. Clim. 519, 1453–1465, https://doi.org/10.1016/j.jhydrol.2014.09.025, 2014.
61. Wu, Y., Miao, C., Fan, X., Gou, J., Zhang, Q., and Zheng, H.: Quantifying the uncertainty sources of future climate projections and narrowing uncertainties with Bias Correction Techniques. Earths Future, 10(11), e2022EF002963, 2022.
62. Zhang, S., Zhou, Z., Peng, P., and Xu, C.: A New Framework for Estimating and Decomposing the Uncertainty of Climate Projections. J. Clim. 37(2), 365-384, https://doi.org/10.1175/JCLI-D-23-0064.1, 2024.