We sincerely appreciate your interest in our paper. As you have pointed out, this study aims to significantly contribute to determining the reliability of GCMs in climate predictions and impact assessments. Previous bias correction studies have primarily focused on selecting appropriate methods based on performance. However, this approach has limitations due to the uncertainties arising from various factors. Our study suggested a Comprehensive Index that integrates uncertainty quantification and performance evaluation in selecting bias correction methods. This approach provides an opportunity to further enhance the reliability of climate models. Specifically, the results of this study are expected to be effectively utilized in future processes for selecting GCMs and evaluating various methods applied to models, facilitating the selection of the most suitable approaches. Once again, we sincerely appreciate your interest in and valuable feedback. We will now begin addressing your questions.

**Comment 1**

Why was only the historical period used? If there is a specific reason, it must be clearly stated.

**Answer**

This study used the three quantile mapping methods by dividing the historical period into training and validation periods. This approach has two main reasons. First, this study divided the historical period into two periods to analyze the performance and uncertainty during the validation period. This concept can intuitively evaluate how precipitation corrected by bias correction methods can be projected into the future. Specifically, the historical period allows for direct comparison with reliable observational data such as ERA5, providing a robust basis to accurately assess the differences between bias-corrected GCM outputs and observed data. This approach facilitates a comparison of the bias correction performance of QDM, EQM, and DQM. Furthermore, using historical data helps remove uncertainties associated with future simulations, enabling the study to focus solely on the performance differences of the correction methods. This was a critical reason to enhance the reliability of the research.

Second, future climate projections are subject to additional uncertainties arising from factors such as GHG emission scenarios and structural differences among GCMs. These factors can hinder the evaluation of the bias correction methods themselves. By using only the historical period, this study concentrated on verifying and comparing the inherent performance of the correction methods, aligning with the study's objectives.

We have added the following sentences to Section 2.3 Quantile Mapping to clarify this.

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**

Developing a comprehensive index is a significant strength of this study. However, there is insufficient explanation of the differences from previous methods to highlight its originality. Please address this. Additionally, the comprehensive index appears to be overly limited. Such a restrictive methodology may reduce the efficiency of the approach, which warrants further discussion.

**Answer**

Thank you for your valuable comment. We have revised the manuscript to clarify further the originality of the comprehensive index proposed in this study and highlight its differences from previous methods. Unlike earlier studies, which primarily focused on the performance of bias correction methods, this study introduces a CI that considers both performance and uncertainty simultaneously. This consideration is a significant advancement, allowing for a more comprehensive and balanced evaluation of bias correction methods. By integrating uncertainty into the evaluation process, the CI addresses a critical limitation of traditional performance-centered approaches, enhancing the reliability and robustness of method selection for climate change research. Additionally, in response to concerns that CI may be seen as overly restrictive, the methodology details that incorporating uncertainty along with performance allows the approach to be leveraged by a variety of factors. The added discussion can explain why this comprehensive evaluation framework is essential for making well-informed decisions regarding climate projections.

The discussion section explains the strengths of CI:

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.

In response to your concerns, we have revised the methodology section to address the flexibility of the CI framework. Specifically, the following sentence has been added to highlight the adaptability of performance and uncertainty metrics:

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 3**

DQM performs worse overall compared to other methods. The reasons for this must be explained. However, as QM methods are not newly proposed in this study but adopted from existing ones, it is unnecessary to deeply analyze the cause (this could be a topic for a future study). A brief explanation in the context of specific climate phenomena should suffice.

**Answer**

Thank you for pointing out the need to explain why DQM performs worse than other methods. Based on the discussion section, DQM's performance can be attributed to its limitations in handling nonlinear climate patterns and extreme events. DQM effectively corrects the meaning through its detrending process. It does not sufficiently account for all quantiles in the precipitation distribution, particularly extremes. This limitation increases uncertainty, especially in regions with complex climate conditions like Southeast Asia, East Africa, and the Alps in Europe. These findings are consistent with previous studies, highlighting DQM's difficulty in accurately capturing long-term climate trends and variability (Berg et al., 2022; Cannon et al., 2015). In response to your comment, the following paragraph is included in the discussion section of the manuscript:

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 4**

The authors used entropy as a criterion weight for TOPSIS. Is there a reason for this choice? Since MCDA is widely used in other fields, incorporating more complex weights might be more appropriate.

**Answer**

Thank you for your comment regarding using entropy to determine the criterion weights in TOPSIS. The reason for selecting entropy is to assign objective weights based on the data. Entropy weights reflect the variability and distribution of each criterion within the dataset, minimizing subjectivity in the weighting process. This ensures that the calculated weights are determined solely by the data, which is crucial for studies comparing bias correction methods across diverse and complex regions.

We have added the following sentence to the methodology section in response to your comment.

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.

While we acknowledge that more intricate weighting schemes, such as expert judgment or hybrid approaches, could integrate additional perspectives, our objective was to establish a straightforward, generalized framework to apply without requiring domain-specific expertise. The simplicity and scalability of entropy weighting make it a suitable choice for determining weights.

**Comment 5**

Model performance can vary significantly depending on the climate zone, but in most cases, the performance is reported to be very high (e.g., EVS: 0.98 (DQM)). Such results indicate exceptional performance. This should be mentioned in the Discussion section.

**Answer**

Thank you for your comment. The three QM methods were undoubtedly closer to ERA5 compared to the raw GCMs. Additionally, their performance is fine for applications in various fields; they demonstrate high performance. In response to your comments, we have added the following paragraph to the Discussion section.

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.

**Comment 6**

Why was only ERA5 selected for comparison? It may not pose a significant issue without a valid reason, but considering multiple reference datasets is essential for evaluating performance.

**Answer**

Thank you for your comment. ERA5, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), provides high-resolution reanalysis data applicable on a global scale and has been widely used for evaluating the performance of climate models. ERA5’s consistent methodology, high temporal and spatial resolution, and comprehensive assimilation of observational data make it highly suitable for assessing the performance of bias correction methods. Furthermore, ERA5 is frequently adopted in high-impact journals ranked in Q1 or above to emphasize the reliability of analyses.

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

This study used ERA5 reanalysis data to contextualize the 2021 heatwave within historical records, covering the period from 1950 to 2022.

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

In the study, ERA5 reanalysis data were employed 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 evaluate the performance of downscaled climate projections by comparing them against historical reanalysis data.

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

In the study, ERA5 data were employed as a reference to assess the performance of downscaled climate products. This application of ERA5 data underscores its credibility and widespread acceptance in high-impact climate research, reinforcing our decision to use ERA5 as the reference dataset in our study.

We added the following statement to the methodology to justify the use of ERA5 as reference data:

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 7**

Why does DQM yield higher estimates for extreme precipitation? Could this be because DQM emphasizes variability in the data due to detrending?

**Answer**

Thank you for your comment. DQM is a bias correction method designed to correct discrepancies between GCMs and reference data and address long-term trends and variability in precipitation. Unlike other QM methods, DQM considers a detrending process that removes long-term trends from modeled data before applying bias correction. This method helps capture bias in extreme climate signals but can result in some discrepancies when comparing bias-corrected precipitation to observed values. DQM preserves the expected climate signals derived from GCM outputs, ensuring consistency with modeled trends. Nevertheless, DQM can overestimate extreme precipitation events compared to observations, as shown in the results of this study. This is because detrending in the correction process can amplify variability, leading to further overestimation in the tail of the precipitation distribution. Furthermore, this method has higher uncertainties in regions with nonlinear or complex climate patterns due to its sensitivity to variability and dependence on the detrending process. This study applied DQM to correct daily precipitation data for six continents using 11 CMIP6 GCMs. DQM effectively captured relative variability and produced results consistent with the predicted climate signal, but overestimated extreme precipitation values ​​compared to observations, which were more frequent in countries and continents where extreme precipitation frequently occurred.

The following three factors may contribute to the overestimation of extreme precipitation corrected using DQM:

1. The detrending process emphasizes the relative variability of data, making DQM particularly sensitive to extreme values. If the model's extreme values are adjusted more significantly than the observed data, the distribution width in the extreme range may be amplified due to the detrending process.

2. DQM adjusts the model data by matching its cumulative distribution function to that of the reference data. However, if the model's distribution is broader than the observed data for extreme values, this mismatch may lead to overestimation. The emphasis on relative values through detrending can further expand the extremes during adjustment.

3. DQM restores the long-term mean to the corrected data as part of the process. This restoration may overestimate the relative contribution of extremes, mainly when the differences in mean and distribution between the observed and modeled values are significant. The more pronounced these differences, the greater the potential overestimation of extremes.

We describe this issue of DQM-corrected extreme precipitation in the Discussion section as follows:

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.