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 sections related to ERA5 have been modified by adopting an alternative methodology. The revised methodology is as follows:

This study utilized re-gridded precipitation data derived from ERA5 reanalysis products provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). The original ERA5 precipitation data, available at a 0.25° × 0.25° spatial resolution, was re-gridded to a 1.0° × 1.0° resolution using the Python library xESMF. The data units were converted from meters per day (m/day) to millimeters per day (mm/day) for consistency with other datasets. The dataset is part of the FROGS (Frequent Rainfall Observations on Grids) database, which integrates various precipitation products, including satellite-based, gauge-based, and reanalysis data (Roca et al., 2019). The re-gridded dataset was selected for its spatial compatibility with the study's objectives, facilitating the evaluation of General Circulation Model (GCM) simulations in replicating observed precipitation patterns. The FROGS database provides a robust framework for intercomparison and assessment of precipitation products across different sources. FROGS database has been widely used in various studies to ensure the reliability of climate model evaluation and climate change assessment (Wood et al., 2021; Roca and Fiolleau, 2020; Petrova et al., 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:

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. Furthermore, 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. Future research should integrate greenhouse gas scenarios to improve the accuracy of climate predictions and provide a more comprehensive understanding of future climate risks. Furthermore, more bias correction methods should be used to extend the robustness of CI.

**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.