CAUCHY: Jurnal Matematika Murni dan Aplikasi
Volume 9(2) (2024), Pages 297-309
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Bias Correction of Lake Toba Rainfall Data Using Quantile Delta Mapping

Syukri Arif Rafhida, Sri Nurdiati*, Retno Budiarti, Mohamad Khoirun Najib

Departemen Matematika, IPB University, Bogor, Indonesia

Email: nurdiati@apps.ipb.ac.id

ABSTRACT

Lake Toba, located in North Sumatra, is the largest tectonic and volcanic lake in Indonesia. Lake Toba has an equatorial climate characterized by abundant rainfall throughout the year. High rainfall, coupled with annual decreases due to climate change, results in a vulnerability to the unpredictable extreme weather, causing harm to the surrounding communities. Consequently, a rainfall prediction model is needed to anticipate the impacts of such extreme rainfall. One of the rainfall prediction models used is ERA5-Land. However, this prediction model has biases that can be avoided. This study applies quantile delta mapping (QDM) to correct biases in ERA5-Land rainfall predictions for Lake Toba using BMKG observations. Two methods—monthly and full-period distribution—were used. Results show improvement in model accuracy, particularly in Silaen, Laguboti, and Doloksanggul stations during dry seasons. However, the first method improves the model distribution more in Silaen and Laguboti, while the second method improves the model distribution more in Doloksanggul. These findings suggest that QDM can improve the accuracy of rainfall prediction models, aiding in more reliable weather-related planning and mitigation efforts in the Lake Toba region.

Keywords: Lake Toba, quantile delta mapping, rainfall

INTRODUCTION

Lake Toba, located in North Sumatra Province, is the largest tectonic and volcanic lake in Indonesia [1]. It spans seven regencies: Karo, Simalungun, Dairi, Toba Samosir, Samosir, North Tapanuli, and Humbang Hasundutan, with geographical coordinates between 98° 31' 2"–98° 9' 14" East Longitude and 2° 19' 15"–2° 54' 2" North Latitude. The lake was formed 1.3 million years ago from the ancient eruption of Mount Toba, which was the largest volcanic eruption in the world [2]. It has a surface area of 1,124 km^2, with a length of about 50 km, a width of about 27 km, and an average depth of 228 meters [3].

High rainfall in Lake Toba as a result of equatorial climate is a crucial aspect of the local community's life. However, studies [4], [5] reported that rainfall in Lake Toba has decreased due to climate change. This decrease in rainfall should be noted to study the potential for future extreme weather events. Rainfall data in an area is continuously recorded to gain more accurate insights. The validity of the data increases proportionally with the amount of data collected [6]. However, a common problem encountered is incomplete data at some rainfall recording stations, which are also

Submitted: September 12, 2024 Reviewed: September 24, 2024 Accepted: October 25, 2024
DOI: http://dx.doi.org/10.18860/ca.v9i2.29124

unevenly distributed [7], [8]. This data loss can be caused by the negligence of rainfall recording personnel due to manual data entry or damage to rainfall recording instruments due to lack of maintenance. This condition makes rainfall forecasting a crucial aspect for residents around Lake Toba to take necessary preventive actions in facing potential extreme weather and its impacts.

Many models have been developed to predict rainfall, one of which is the ECMWF Reanalysis 5th Generation in land variables (ERA5-Land) for rainfall, developed by the European Centre for Medium-Range Weather Forecasts (ECMWF) [9]. ERA5-Land has become one of the frequently used sources of weather and climate forecasting data in Indonesia [10], [11]. However, like other forecasting models, ERA5-Land often shows systematic errors when compared to observational data in the field [12], [13]. This highlights the need for bias correction to make the model data more reflective of actual conditions.

Various bias correction methods have been developed to downscale climate variables and reduce model bias. The study [14] mentions that, based on their schemes, commonly applied methods can be divided into three categories. The first category considers the spatial consistency of climate variables, such as the analog approach. The second category adjusts the mean and standard deviation of climate variables in model simulations based on observations, such as linear scaling and delta change methods. The third category corrects the probability distribution of model simulations by applying correction factors to the Cumulative Distribution Function (CDF) of the modeled and observed variables, for example, quantile mapping (QM) and distribution mapping (DM) based on empirical and theoretical CDFs. QM is effective in eliminating model bias, both for the mean and standard deviation, as well as for extreme events [15]. Many bias correction methods, including QM and DM as well as the second category of bias correction methods, assume that the statistical relationship between past model simulations and observations remains unchanged in the future, even as the distribution is expected to change over time. These conventional bias correction methods can deliberately distort the climate change signal, necessitating methods that efficiently preserve changes; thus, quantile delta mapping (QDM) was developed.

Quantile delta mapping (QDM), developed based on QM, retains QM's advantages while accounting for CDF changes over different time periods. The main advantage of QDM is its ability to preserve the trend in climate model output, which is a critical aspect in climate analysis [16]. In the QDM method, there are two types of distribution searches that can be used. The first is distribution search based on months [10], [11], and the second is distribution search without considering months [17]. Many studies have adopted the QDM method as a bias correction tool in their research. For example, Gumus et al. [18] used QDM along with other methods such as QM and detrended quantile mapping to improve the accuracy of temperature and precipitation data from the Coupled Model Intercomparison Project Phase 6 (CMIP6) data. Meanwhile, Reboita et al. [19] used QDM to correct bias in CMIP6 data for precipitation in South America. However, the use of QDM for climate models in Indonesia, particularly Lake Toba, is still limited. Therefore, this research aims to perform statistical bias correction using the quantile delta mapping (QDM) approach on rainfall prediction data for the Lake Toba region through monthly and full-period distribution search methods and to evaluate the prediction model before and after correction.

METHODS

Data

This study focuses on bias correction of rainfall predictions using data from the ERA5-Land model and BMKG (Meteorological, Climatological, and Geophysical Agency) observational data for the Lake Toba region, North Sumatra, Indonesia. The data used in this study are monthly total rainfall data from January 1973 to December 2017. The observational data were obtained from the Meteorology, Climatology, and Geophysics Agency (BMKG) at seven weather station locations. As shown in Figure 1, the locations of the seven rainfall stations around Lake Toba, listed from 1–7, are Pangururan, Parapat, Silaen, Laguboti, Siborong-borong, Doloksanggul, and Merek. The data were provided in Microsoft Excel 97-2003 format. Some cells contained null data, necessitating pre-processing using the k-nearest neighbors (knn) imputation method.

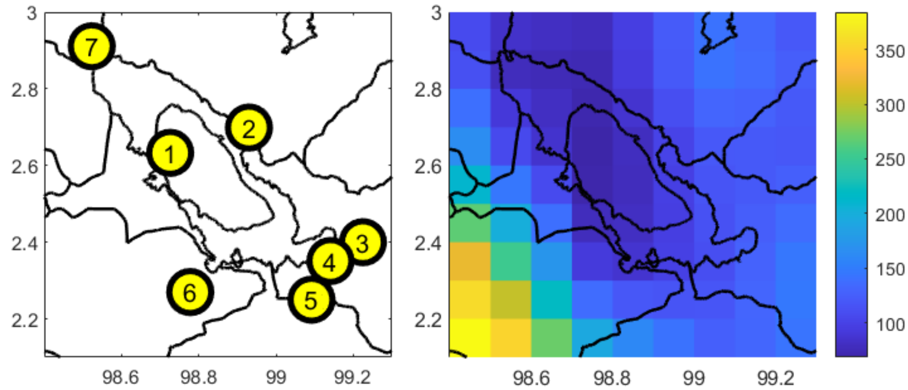

Figure 1. Locations of Weather Stations Around Lake Toba (Left), ERA5-Land Data for January 1973 (Right)

The model data used in this study is ERA5-Land data from the ECMWF (European Centre for Medium-Range Weather Forecasts), presented in a grid format as shown in Figure 1. The data has a spatial resolution of 0.1° × 0.1° and was downloaded for coordinates around Lake Toba, specifically 2.1°–3° North Latitude, 98.4°–99.1° East Longitude, in netCDF format. The ERA5-Land model data files are released monthly and include four variables: longitude, latitude, time, and tp (total precipitation). The time variable indicates the prediction time for monthly rainfall, while the tp variable represents the average monthly precipitation on a specific grid, which is converted into total monthly rainfall by multiplying by the number of days in the month.

Data Extraction

The observational data are only available at the specific coordinates of the seven weather stations, whereas the ERA5-Land model data are provided in a grid format. To standardize both data formats into time series data of total monthly rainfall at the seven weather station locations, an interpolation process of the model data to the coordinates of the weather stations is necessary. In this study, spline interpolation was used to obtain the model rainfall data. The observational data are stored as $x_{o,h}$ for the years 1973-2003 and $x_{o,p}$ for the years 2004-2017, while the model data are stored as $x_{m,h}$ for the years 1973-2003 and $x_{m,p}$ for the years 2004-2017.

Distribution Identification

The identification of the distribution was performed on the data $x_{o,h}$, $x_{m,h}$ for the years 1973–2003, and $x_{m,p}$ for a time window of t years, in accordance with the QDM method. The selected value of t is 20 years. This approach has been previously applied by [20] in estimating rainfall distribution. The distributions used are listed in Table 1. The parameters of each distribution were estimated using the maximum likelihood estimator (MLE) with the help of MATLAB software. The distribution that best fits the data was determined based on the Kolmogorov-Smirnov one-sample error statistic.

Table 1. The distribution equation used to identify the distribution of observed and model rainfall data.

Distribution Name	Probability Density Function
Extreme Value	$f(x \mid \mu, \sigma) = \frac{1}{\sigma} e^{-\left(\frac{x-\mu}{\sigma}\right)} e^{-e^{-\left(\frac{x-\mu}{\sigma}\right)}}$
Generalized Extreme Value	$f(x\|\mu, \sigma, \xi) = \frac{1}{\sigma} e^{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}} \left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi-1}$
Logistic	$f(x\|\mu, s) = \frac{1}{s} \frac{e^{-(x-\mu)/s}}{(1+e^{-(x-\mu)/s})^2}$
Normal	$f(x\|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Exponential	$f(x \mid \mu) = \frac{1}{\mu} e^{\left(-\frac{x}{\mu}\right)}$
Gamma	$f(x\|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}}$
Inverse Gaussian	$f(x \mid \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} e^{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}}$
Log-Logistic	$f(x\|\alpha, \beta) = \frac{\beta}{x}\left(1 + \left(\frac{x}{\alpha}\right)^{\beta+1}\right)^{-1}$
Log-Normal	$f(x\|\mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(ln(x)-\mu)^2}{2\sigma^2}}$
Weibull	$f(x\|\lambda, k) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}$

Bias Correction for Prediction Models

Bias correction in statistical models is developed by establishing a relationship between observed and predicted rainfall data to form a specific correction factor $[y = f(x)]$. This function connects the values of the Cumulative Distribution Function (CDF) of observed rainfall and the CDF of predicted rainfall. The relationship is expressed as: $\mathrm{CDF}_{\mathrm{observasi}}(y) = \mathrm{CDF}_{\mathrm{prediksi}}(x)$ [21]. The CDF values can be calculated by integrating the Probability Density Function (PDF).

Quantile Delta Mapping

Quantile Delta Mapping (QDM) in this study introduced with detailed formulas by Cannon et al. [17]. By comparing QDM with other QM methods, [16] demonstrated the superiority of QDM in preserving trends in climate model outputs. The QDM method output is corrected rainfall data ($\hat{x}_{m,p}$) obtained by using following equations:

$$\hat{x}_{m,p}(t) = \hat{x}_{o:m,h:p}(t) \times \Delta_m(t)$$
$$\hat{x}_{o:m,h:p}(t) = F_{o,h}^{-1}\left(F_{m,p}^{(t)}\left(x_{m,p}(t)\right)\right) \tag{1}$$

$$\Delta_m(t) = \frac{x_{m,p}(t)}{F_{m,h}^{-1}\left(F_{m,p}^{(t)}\left(x_{m,p}(t)\right)\right)}$$

where model data, observation data, historical data, and prediction data are denoted with subscripts m, o, p and p respectively. Bias-corrected historical periodic data represented by $\hat{x}_{o:m,h:p}(t)$, and $\Delta_m(t)$ represents the relative change from historical data to the prediction period. $F_{m,h}$ and $F_{o,h}$ respectively represent the CDFs of the ERA5-Land model data and BMKG observational data during the historical period. $F_{m,p}^{(t)}$ denotes the time-dependent CDF of the model data within a specific time window [18].

At this step, the $x_{m,p}$ data obtained during the extraction phase is bias-corrected using the QDM method. This study uses two types of QDM methods for bias correction. In a 20-year time window, the distribution of $x_{m,p}$ is identified for each month, resulting in 12 distributions for 12 months, and also identified comprehensively without separating by month.

Evaluation of the Corrected Model

The model evaluation is used to compare the PDF of the prediction model before and after correction. At this step, the model is evaluated based on the difference in MAE between the corrected and uncorrected models, as well as on the Kolmogorov-Smirnov two-sample test between the observational data and model data. The MAE difference will be positive if the MAE of the corrected model is smaller than that of the uncorrected model, and vice versa. The K-S test will be considered successful if the p-value of the data is greater than 0.01.

RESULTS AND DISCUSSION

Data Overview

The area around Lake Toba has an equatorial climate characterized by two rainfall peaks. This pattern is influenced by the movement of the convergence zone towards the north and south, following the apparent movement of the sun [22]. The convergence zone is where two air masses from different locations meet and then move upward. This convergence zone is commonly known as the Inter-Tropical Convergence Zone (ITCZ) and is located in low latitudes, including Lake Toba. A general overview of rainfall in the Lake Toba area is visualized in Figure 2.

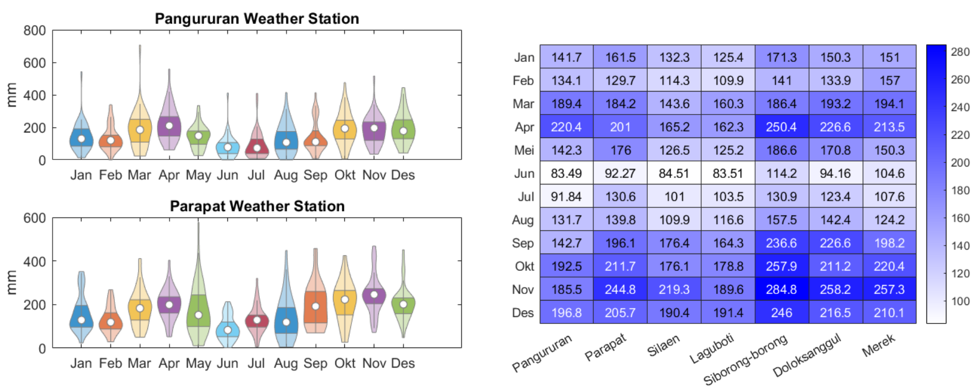

Figure 2. Rainfall at stations around Lake Toba in 1972-2017 with violin chart (left), and using heatmap (bottom)

Figure 2 shows that Pangururan and Parapat stations have two identical rainfall peaks in March–April and October–November, as seen from the average violin chart. More generally, based on the heatmap in Figure 2, the highest average rainfall in the Lake Toba area occurs during the second peak in October–November, with the highest average at the Siborong-borong Weather Station in November. On the other hand, the lowest average rainfall occurs during the dry season of the equatorial climate, in June–July, as indicated by the white color on the heatmap. A comparison with model data is shown in Figure 3.

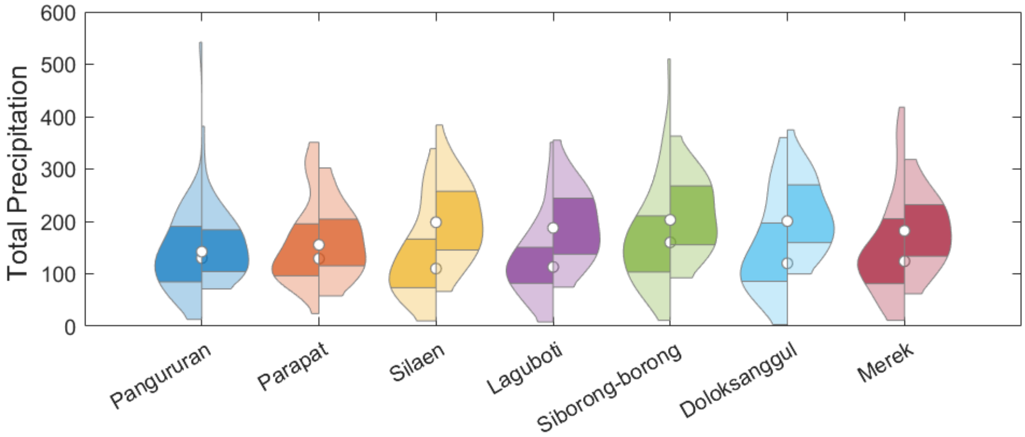

Figure 3. Comparison of rainfall data observation (left) and ERA5-Land model (right) at seven stations in January

Based on Figure 3, there is a difference between observed rainfall data and model data. The ERA5-Land model rainfall data tends to be overestimated at all seven weather stations. The most overestimated data is at the Silaen Weather Station, which has the highest median difference.

Distribution Identification

The next step is to identify the distribution of observed and modeled rainfall using 10 distributions. This test shows the sample data error magnitude compared to the ten distributions, which have been optimized using the Maximum Likelihood Estimator (MLE). The distribution with the smallest one-sample K-S test score is selected. The results of the distribution identification for the seven station data for $x_{o,h}$ in January 1973–2003 are presented in Table 2.

Table 2. Identification of Rainfall Distribution in January 1973–2003 with K-S Test

Distribution	Pan	Par	Sil	Lag	Sbb	Dol	Mer
EV	0.2913	0.2596	0.1914	0.1759	0.2236	0.2052	0.1881
GEV	**0.1184**	**0.0986**	0.0791	**0.0658**	0.0742	**0.0646**	**0.0584**
LOG	0.1296	0.1786	0.1102	0.0845	0.1032	0.0962	0.0937
NOR	0.2404	0.2141	0.1601	0.1298	0.1747	0.1486	0.1384
EXP	0.2488	0.4044	0.1851	0.2163	0.2380	0.2338	0.2162
GAM	0.1512	0.1734	0.0708	0.0877	0.0945	0.0930	0.0849
ING	0.1611	0.1519	0.1436	0.2085	0.1670	0.2901	0.1961
LL	0.1227	0.1213	**0.0674**	0.0806	**0.0674**	0.0711	0.0827
LN	0.1316	0.1444	0.0962	0.1389	0.1082	0.1484	0.1343
WB	0.1731	0.1797	0.0888	0.0665	0.1156	0.0885	0.0752

The K-S test values shown in Table 2 represent the maximum difference between the empirical CDF and the estimated CDF obtained using MLE. The bold values represent the smallest K-S test values at each station. The distribution with the smallest K-S test value is selected as the estimated distribution for that station and period. The best distribution for $x_{o,h}$ in January for the Pangururan, Parapat, Laguboti, Doloksanggul, and Merek stations follows the generalized extreme value distribution, while the Silaen and Siborong-borong stations follow the log-logistic distribution. This process is repeated for each $x_{o,h}$, $x_{m,h}$ per month in both the monthly and full-period QDM types during the observation period and the time window for the full-period QDM type.

Bias Correction

In this study, a bias correction process for rainfall was conducted to develop a new prediction model. The method used for bias correction is the QDM method as performed by [17]. The $x_{m,h}$ data was corrected based on the $x_{o,h}$ data at each station within a specific time period. This study uses two types of time periods. The first type of bias correction is performed monthly during the historical and prediction periods, resulting in twelve different distributions for each month [10], [11]. The second type of bias correction is conducted for the entire year, resulting in only one distribution [17]. The comparison between the CDF of observed data, pre-correction model data, and post-correction model data of the first type of bias correction, or monthly correction, at one of the rainfall stations are presented in Figure 4 for the Merek Weather Station during the prediction period from 2004 to 2017. While the comparison between three CDF of the second type or full-period of bias correction are presented in Figure 5.

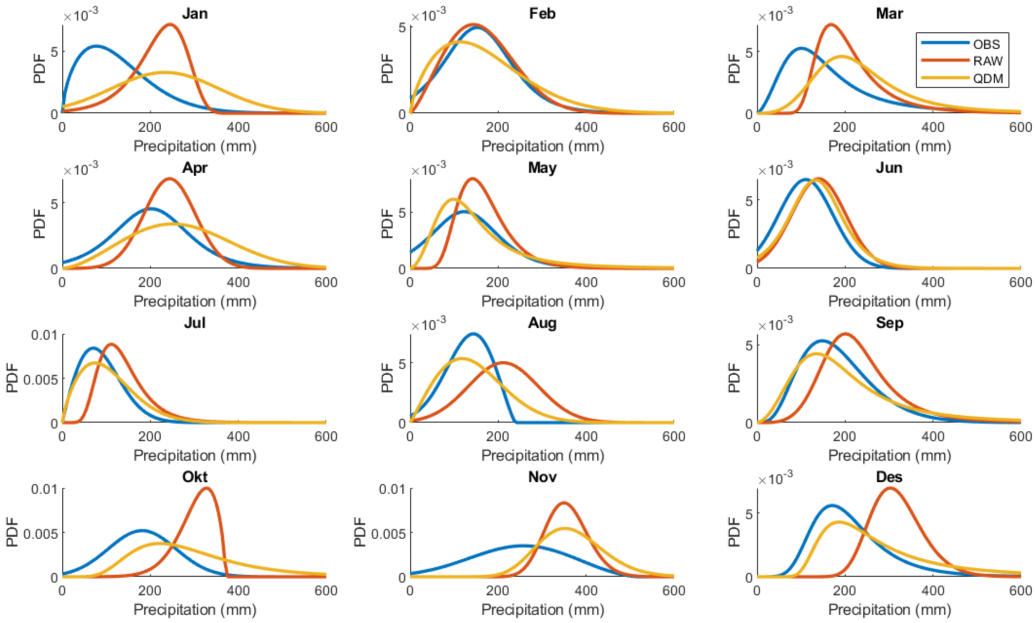

Figure 4. PDF of $\boldsymbol{x_{o,p}}$, $\boldsymbol{x_{m,p}}$, and $\boldsymbol{\hat{x}_{m,p}}$ Data from 2004–2017 at Merek Weather Station using QDM with Monthly Correction

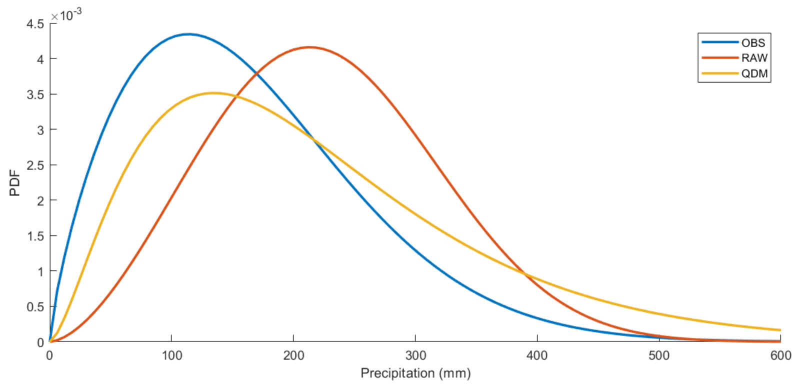

Figure 5. PDF of $\boldsymbol{x_{o,p}}$, $\boldsymbol{x_{m,p}}$, and $\boldsymbol{\hat{x}_{m,p}}$ Data from 2004–2017 at Merek Weather Station using QDM with Full-Period Correction

Based on the bias correction results using monthly QDM at the Merek Weather Station shown in Figure 4, there are several months where the corrected model PDF performs better than the PDF of the uncorrected model. Although the improvement is not very significant, the corrected model's PDF using QDM in January, March, May, June, July, August, September, October, November, and December appears to be closer to the observed PDF, as indicated by the average distribution. On the other hand, the bias correction results using full-period QDM at the Merek Weather Station shown in Figure 5 demonstrate that the average PDF of the corrected model has closely approached the observed PDF. These results indicate that while QDM does not significantly make the corrected model data identical to the observed data, it does show that the distribution of the corrected model ($\hat{x}_{m,p}$) is closer to the observed distribution ($x_{o,p}$) than the uncorrected model distribution ($x_{m,p}$). This process is applied to all rainfall stations.

Corrected Model Evaluation

After the bias correction process is completed and the new corrected rainfall model is obtained, the next step is to evaluate each created model. This process is done to measure the accuracy and precision of the corrected model on previously uncalculated data, specifically the $x_{m,p}$ data. Accuracy reflects how close the corrected model is to the actual rainfall values, and precision shows how small the differences are between the corrected model's data patterns and the observed rainfall data.

MAE Test

Model evaluation is conducted based on the Mean Absolute Error (MAE) between the corrected data and the observed data. MAE measures how accurately the model predicts values. MAE is calculated by taking the absolute difference between the model predictions and the observed values and then averaging all these absolute values. The lower the MAE value, the better the model is at predicting values accurately. The results of the MAE test are shown in Figure 6.

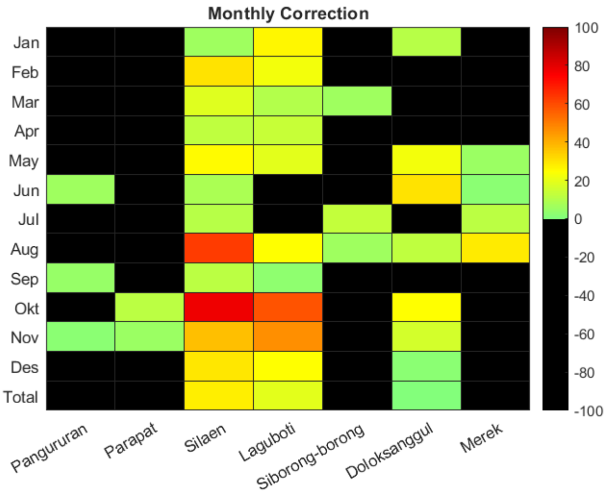

Figure 6. MAE Difference Between Rainfall Models Before and After Bias Correction in the Prediction Period (2004–2017) Using QDM with Monthly Correction

The MAE test results are shown in Figure 6, where the green-red color gradient indicates that the corrected model has a lower MAE than the uncorrected model, resulting in a positive MAE difference, while the black color indicates the opposite. From the total MAE difference, it can be seen that QDM improves the accuracy of the ERA5-Land model by reducing the model's MAE at the Silaen, Laguboti, and Doloksanggul Stations, while the QDM model shows less improvement at other stations. This first type of QDM improves 41 out of 84 months across the seven stations. The method also predominantly improves model data in May, June, August, October, and November. The same is done for QDM with full-period bias correction, shown in Figure 7.

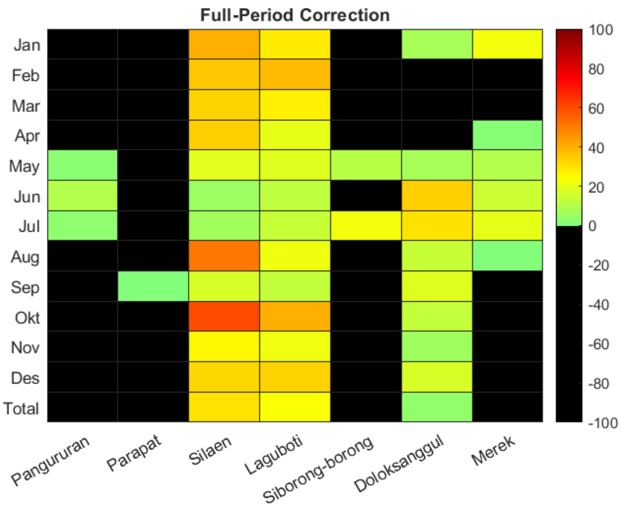

Figure 7. MAE Difference Between Rainfall Models Before and After Bias Correction in the Prediction Period (2004–2017) Using QDM with Full-Period Correction

Based on Figure 7, QDM with full-period correction improves the MAE of the ERA5-Land model at the Silaen, Laguboti, and Doloksanggul Stations, as seen from the total MAE difference. This second type of QDM improves 45 out of 84 months across the seven weather stations. This type of QDM more effectively reduces the MAE difference in May, June, July, August, and September at most stations. However, this QDM method is less effective in reducing the MAE at the Parapat Weather Station.

Two-Sample Kolmogorov-Smirnov Test

In addition to MAE, this study uses the two-sample Kolmogorov-Smirnov (K-S) test to compare two distributions, namely the observed data with the model data before and after correction [18]. The null hypothesis of this test assumes that both data samples come from the same distribution, while the alternative hypothesis states that the two data samples come from different distributions. The diagnostic test is considered successful if the null hypothesis is not rejected, meaning that no modeled quantile lies far outside the 99% confidence interval. Figure 8 shows the K-S test results on the corrected model data.

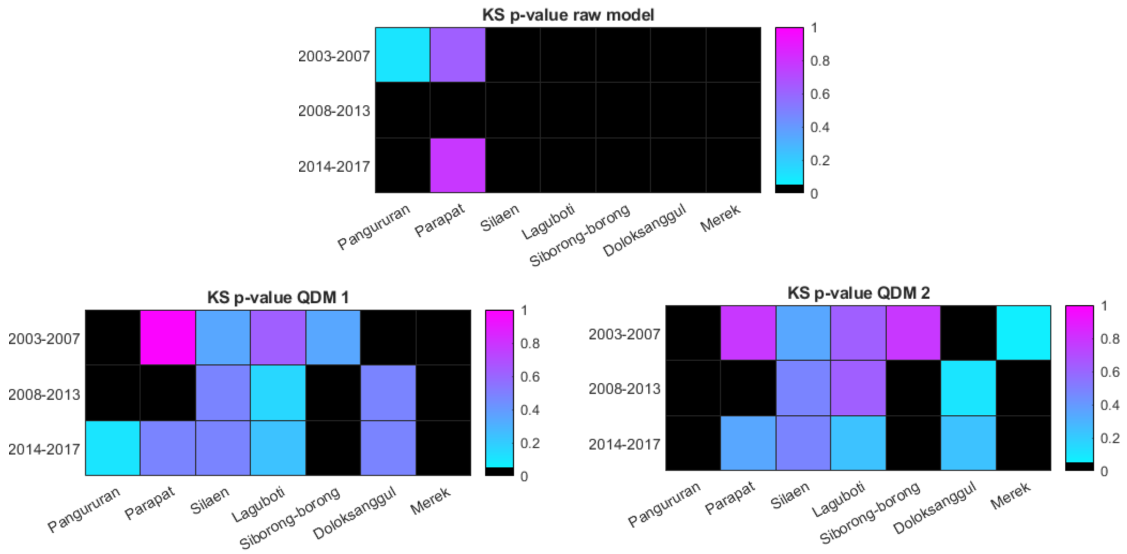

Figure 8. P-value of the Two-Sample Kolmogorov-Smirnov Test on the Distribution of Corrected Models with QDM Monthly Correction, Full-Period Correction, and Uncorrected Models

The Kolmogorov-Smirnov test is considered successful if the p-value of the observation is greater than 0.01. Based on Figure 8, it can be seen that QDM improves the distribution alignment between the model data and the observed data. For example, at the Laguboti station, the distribution before correction differed from the observation, as indicated by the p-value of the distribution being less than 0.01. However, after correction, the p-value of the distribution exceeds 0.01.

Before correction, many model data distributions significantly differed from the observed data. The model data only had significantly similar distributions based on the K-S test at the Parapat Station in the periods 2003–2007 and 2013–2017 and at the Pangururan Station in the period 2003–2007. The corrected data can improve the distributions at the Silaen, Laguboti, Siborong-borong, and Doloksanggul Stations in some periods, with QDM type 2 improving the distribution during the 2003–2007 period. QDM type 1 is more effective in improving the distribution at Doloksanggul, while QDM type 2 improves the distribution at Silaen, Laguboti, and Siborong-borong. Although not all distributions are corrected, significant improvements are observed between the models before and after correction.

Discussions

Quantile Delta Mapping (QDM) corrects model data more effectively during the equatorial dry season (May to August) as shown in MAE test. It can be happened due to several factors. First, rainfall patterns are more stable and predictable during the dry season, with less variability and fewer extreme events, allowing QDM to map the quantile distribution of observed data to the model more accurately. Additionally, the cumulative distribution function (CDF) of rainfall in the dry season is smoother, enabling more precise adjustments by QDM. The predictable and consistent seasonal patterns in the dry season allow QDM to leverage historical data more effectively, leading to improved performance in aligning model data with observations.

The effectiveness of the QDM method varied depending on the station and period analyzed. At Doloksanggul, the monthly QDM method outperformed the full-period QDM due to the station's exposure to more localized weather patterns, which change significantly between months. Conversely, in Silaen and Laguboti, the full-period QDM showed superior performance, potentially due to more consistent rainfall patterns throughout the year, making a holistic correction more suitable. These results highlight the importance of selecting appropriate bias correction methods based on local climate characteristics.

A study by [4] has corrected precipitation in ERA5 data for the Lake Toba region using a different approach. That study applied Quantile Mapping (QM) to correct precipitation and temperature data, focusing on grid data for the entire Lake Toba area. In contrast, this study uses data from seven specific weather stations to correct precipitation. Despite these differences in methodology, both studies found that the corrected data is more accurate than the uncorrected data.

The improvements in rainfall prediction accuracy can have significant practical applications. Local governments can use these findings to enhance disaster preparedness and response plans for extreme weather events. Natural resource managers can better plan water resource management during dry periods, while the agricultural sector can use these predictions to optimize planting schedules and mitigate the impacts of both droughts and floods. Such targeted interventions can help reduce the vulnerability of communities around Lake Toba to extreme weather conditions.

Recommendations for Future Research: Future research is recommended to use daily data as the basis for BMKG observations. Daily data provides a larger dataset compared to monthly data, especially for a short observation period like 1973–2017. Using daily data could result in a better distribution, providing more detailed information. Additionally, future studies could explore combining QDM with other bias correction methods or applying machine learning approaches to further improve model accuracy. Evaluating long-term trends could provide deeper insights into climate change impacts.

CONCLUSIONS

The bias correction method can identify rainfall prediction biases quite well based on observational data and models. This is demonstrated by the rainfall bias correction process at weather stations around Lake Toba using two distribution search methods. The results of both methods, namely QDM with a monthly period and QDM with a full period, effectively correct the model bias at the Silaen, Laguboti, and Doloksanggul Stations.

The MAE difference results show that the corrected data's MAE is smaller than the model data before correction in 41 out of 84 months using the QDM monthly period method and in 45 out of 84 months using the QDM full period method at the seven stations. Additionally, both QDM methods improve model data more effectively during the equatorial climate dry season in May, June, July, and August. The QDM full period method better corrects the rainfall data distribution at Silaen and Laguboti during the entire prediction period and at Siborong-borong in the initial period compared to the QDM monthly period method. Meanwhile, the QDM monthly period method better corrects the data distribution at Doloksanggul Station. The results of this study suggest that QDM can be a useful tool for bias correction in rainfall prediction models, especially in tropical regions like Lake Toba. However, its application to different datasets or in regions with different climate patterns should be explored.

REFERENCES

[1] M. T. Astuti, "Strengthening the Tourism Promotion Strategy inPriority Destinations for Lake Toba," *Int. Conf. Gov. Educ. Manag. Tour.*, 2021.

[2] J. D. Guballa, J. Bollmann, K. Schmidt, and A. Lückge, "The Toba Eruption 74,000 Years ago Strengthened the Indian Winter Monsoon-Evidence From Coccolithophores," *Paleoceanogr. Paleoclimatology*, vol. 39, no. 4, p. e2023PA004823, 2024, doi: https://doi.org/10.1029/2023PA004823.

[3] R. Lukman, "Kajian Kondisi Morfometri Dan Beberapa Parameter Stratifikasi Perairan Danau Toba," *J. Limnotek*, vol. 17, no. 2, pp. 158–170, 2010.

[4] H. Irwandi, M. S. Rosid, and T. Mart, "Effects of Climate change on temperature and precipitation in the Lake Toba region, Indonesia, based on ERA5-land data with quantile mapping bias correction," *Sci. Rep.*, vol. 13, no. 1, pp. 1–11, 2023, doi: 10.1038/s41598-023-29592-y.

[5] H. Irwandi, J. Y. Ariantono, Q. A. Kartika, A. C. P. Siregar, C. A. Tari, and A. Sudrajat, "Pengaruh Iklim Terhadap Penurunan Tinggi Muka Air Danau Toba," *Semin. Nas. Sains Atmos. 2017*, pp. 105–110, 2017.

[6] N. S. N. Golkhatmi and M. Farzandi, "Enhancing Rainfall Data Consistency and Completeness: A Spatiotemporal Quality Control Approach and Missing Data Reconstruction Using MICE on Large Precipitation Datasets," *Water Resour. Manag.*, 2024, doi: 10.1007/s11269-023-03567-0.

[7] M. Mamenun, H. Pawitan, and A. Sopaheluwakan, "Validasi Dan Koreksi Data Satelit Trmm Pada Tiga Pola Hujan Di Indonesia," *J. Meteorol. dan Geofis.*, vol. 15, no. 1, 2014, doi: 10.31172/jmg.v15i1.169.

[8] F. Oriani, S. Stisen, M. C. Demirel, and G. Mariethoz, "Missing Data Imputation for Multisite Rainfall Networks: A Comparison between Geostatistical Interpolation and Pattern-Based Estimation on Different Terrain Types," *J. Hydrometeorol.*, 2020, doi: 10.1175/jhm-d-19-0220.1.

[9] H. Hersbach *et al.*, "The ERA5 global reanalysis," *Q. J. R. Meteorol. Soc.*, vol. 146, no. 730, pp. 1999–2049, 2020, doi: 10.1002/qj.3803.

[10] M. K. Najib and S. Nurdiati, "Koreksi Bias Statistik Pada Data Prediksi Suhu Permukaan Air Laut Di Wilayah Indian Ocean Dipole Barat Dan Timur," *Jambura Geosci. Rev.*, vol. 3, no. 1, pp. 9–17, 2021, doi: 10.34312/jgeosrev.v3i1.8259.

[11] S. Nurdiati, A. Sopaheluwakan, and M. K. Najib, "Statistical Bias Correction for Predictions of Indian Ocean Dipole Index With Quantile Mapping Approach," *Int. MIPAnet Conf. Sci. Math.*, no. April 2021, 2019, doi: 10.31219/osf.io/7dq2j.

[12] M. Cucchi *et al.*, "WFDE5: Bias-adjusted ERA5 reanalysis data for impact studies," *Earth Syst. Sci. Data*, 2020, doi: 10.5194/essd-12-2097-2020.

[13] J. Muñoz-Sabater *et al.*, "ERA5-Land: A state-of-the-art global reanalysis dataset for land applications," *Earth Syst. Sci. Data*, 2021, doi: 10.5194/essd-13-4349-2021.

[14] X. Li and Z. Li, "Evaluation of bias correction techniques for generating high-resolution daily temperature projections from CMIP6 models," *Clim. Dyn.*, vol. 61, no. 7–8, pp. 3893–3910, 2023, doi: 10.1007/s00382-023-06778-8.

[15] M. Shah, A. Thakkar, and H. Shastri, "Comparative analysis of bias correction techniques for future climate assessment using {CMIP6} hydrological variables for the Indian subcontinent," *Acta Geophys.*, 2024.

[16] A. J. Cannon, "Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables," *Clim. Dyn.*, vol. 50, no. 1–2, pp. 31–49, 2018, doi: 10.1007/s00382-017-3580-6.

[17] A. J. Cannon, S. R. Sobie, and T. Q. Murdock, "Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes?," *J. Clim.*, vol. 28, no. 17, pp. 6938–6959, 2015, doi: 10.1175/JCLI-D-14-00754.1.

[18] B. Gumus, S. Oruc, I. Yucel, and M. T. Yilmaz, "Impacts of Climate Change on Extreme Climate Indices in Türkiye Driven by High-Resolution Downscaled CMIP6 Climate Models," *Sustain.*, vol. 15, no. 9, 2023, doi: 10.3390/su15097202.

[19] M. S. Reboita, G. W. de S. Ferreira, J. G. M. Ribeiro, R. P. da Rocha, and V. B. Rao, "South American Monsoon Lifecycle Projected by Statistical Downscaling with CMIP6-GCMs," *Atmosphere (Basel).*, vol. 14, no. 9, p. 1380, 2023, doi: 10.3390/atmos14091380.

[20] S. Mohita Anand and S. Jai Bhagwan, "Use of Probability Distribution in Rainfall Analysis," *New York Sci. J.*, vol. 3, no. 9, pp. 40–49, 2010.

[21] C. Piani, J. O. Haerter, and E. Coppola, "Statistical bias correction for daily precipitation in regional climate models over Europe," *Theor. Appl. Climatol.*, vol. 99, no. 1–2, pp. 187–192, 2010, doi: 10.1007/s00704-009-0134-9.

[22] Tukidi, "Karakter Curah Hujan Di Indonesia," *J. Geogr.*, vol. 7, no. 2, pp. 136–145, 2010, [Online]. Available: http://journal.unnes.ac.id/nju/index.php/JG/article/view/84