https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

Refining Remote Sensing precipitation Datasets in the South Pacific: An Adaptive Multi-Method Approach for Calibrating the TRMM Product

Óscar Mirones[1], Joaquín Bedia[2,3], Sixto Herrera[2], Maialen Iturbide[1], and Jorge Baño Medina[1]

[1]Santander Meteorology Group. Instituto de Física de Cantabria IFCA, CSIC-UC, 39005 Santander, Spain
[2]Dept. Matemática Aplicada y Ciencias de la Computación (MACC), Universidad de Cantabria, 39005 Santander, Spain
[3]Grupo de Meteorología y Computación, Universidad de Cantabria, Unidad Asociada al CSIC, Santander (Spain)

Correspondence: O. Mirones (oscar.mirones@unican.es)

Abstract.

Calibration techniques are gaining popularity in climate research for refining numerical model outputs, favored for their relative simplicity and fitness-for-purpose in many climate impact applications. Their range of applicability goes beyond numerical model outputs and can be applied to calibrate remote sensing datasets that can exhibit important biases as compared to in situ meteorological observations. This study presents an adaptive calibration approach specifically designed for calibrating the Tropical Rainfall Measuring Mission (TRMM) precipitation product across multiple stations in the South Pacific. The methodology involves the daily classification of the target series into five distinct Weather Types (WTs) capturing the diverse spatio-temporal precipitation patterns in the region. Various quantile mapping (QM) techniques, including empirical (eQM), parametric (pQM), and Generalized Pareto Distribution (gpQM), as well as an ordinary scaling, are applied for each WT. We perform a comprehensive validation by evaluating 10 specific precipitation-related indices that hold significance in impact studies, which are then combined into a single Ranking Framework (RF) score, which offers a comprehensive evaluation of the performance of each calibration method for every Weather Type (WT). These indices are assigned user-defined weights, allowing for a customized assessment of their relative importance to the overall RF score. Our 'adaptive' approach selects the best performing method for each WT based on the RF score, yielding an optimally calibrated series.

Our findings indicate that the adaptive calibration methodology surpasses standard and weather-type conditioned methods based on a single technique, yielding more accurate calibrated series in terms of mean a extreme precipitation indices consistently across locations. Moreover, this methodology provides the flexibility to customize the calibration process based on user preferences, thereby allowing for specific indices, such as extreme rainfall indicators, to be assigned higher weights. This ability enables the calibration to effectively address the influence of intense rainfall events on the overall distribution. Furthermore, the proposed adaptive method is highly versatile and can be applied to different scenarios, datasets, and regions, provided that a prior weather typing exists to capture the pertinent processes related to regional precipitation patterns. Open-source code and illustrative examples are freely accessible to facilitate the application of the method.

Keywords — weather types, quantile mapping, extreme precipitation, precipitation indices, bias adjustment

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

1 Introduction

Satellite rainfall products serve as crucial sources of information for various hydrological applications, offering continuous temporal coverage and consistent spatial estimates of precipitation in regions lacking sufficient rain gauge data. However, unlike direct observations, satellite measurements are prone to systematic errors originating from uncertainties in estimating precipitation amounts from radar reflectivity measurements (Simpson et al., 1996; Sekaranom and Masunaga, 2019) or the irregular timing of satellite overpass (Aghakouchak et al., 2009), among others. Consequently, these products deviate significantly from the statistical properties of observed series, particularly concerning extreme precipitation events (Mirones et al., 2022), thus requiring calibration before their application in impact studies.

Essentially, the calibration process entails adjusting a transfer function that relates the parameters of raw satellite precipitation distribution to observed rain gauge time series. The effectiveness of bias reduction through post-processing depends on the underlying mechanisms producing the bias (see e.g. Maraun et al., 2017), as well as the appropriateness and accurate implementation of the chosen technique. Moreover, it is crucial to accompany this process with a proper estimation of the associated uncertainty. In particular, the TRMM biases are not constant but associated with specific meteorological conditions, exhibiting a systematic overestimation during wet periods and underestimation during dry periods (Islam et al., 2010; Almazroui, 2011). Hence, it is reasonable to anticipate that incorporating explicit information regarding the synoptic-scale meteorological conditions into the calibration process would enhance the fitting of the transfer function. In this context, weather typing techniques (see Huth et al., 2008, for a comprehensive review), prove helpful in defining relevant weather patterns by summarizing distinct atmospheric configurations associated with different precipitation regimes (Baltaci et al., 2015; Hay et al., 1991; Riediger and Gratzki, 2014; Trigo and DaCamara, 2000). This approach would effectively situate the calibration within the context of significant atmospheric circulation processes that impact the target variable, as previously shown (Mirones et al., 2022), considering that the biases may be different depending on the prevailing atmospheric processes at each moment (Jury et al., 2019), so that a generalist adjustment may not allow to solve them efficiently in all cases. Moreover, although conditioning reduces the sample size, it has been shown that the calibration with adequate sub-samples can significantly enhance the reliability of the corrected series Reiter et al. (2018).

While atmospheric pattern classifications for conditioned transfer function calibration have been already used in statistical downscaling for climate change studies (Stehlik and Bardossy, 2002; Wetterhall et al., 2007, 2012) as well as in seasonal forecasting applications (Manzanas and Gutiérrez, 2019) and short-term forecast calibration refinement (Vuillaume and Herath, 2017), they have seldom been explored in the context of satellite product calibration. In a recent study, Mirones et al. (2022) proposed an innovative approach for calibrating TRMM data in the South Pacific region. The methodology incorporates scaling and empirical quantile mapping techniques, conditioned to the dominant modes of interannual variability captured by specific precipitation types. This region encompasses the South Pacific Convergence Zone (SPCZ), characterized by a distinct band of low-level convergence and enhanced cloudiness extending across the South Pacific (Australian Bureau of Meteorology and CSIRO, 2011). The SPCZ is associated with notable meteorological phenomena such as heavy rainfall, convective storms, and the displacement of the intertropical convergence zone (ITCZ, Waliser and Gautier, 1993). The defined weather types were

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

thus designed to capture the key characteristics of the regional precipitation regime while ensuring a sufficient sample size for robust conditional model fitting.

Building upon this methodological framework presented by Mirones et al. (2022), this study aims to further explore the potential of conditioned calibration for improving the quality of TRMM precipitation data. We expand the range of calibration techniques by incorporating a broader selection of commonly used parametric and non-parametric methods, including scaling, empirical quantile mapping (eQM), parametric quantile mapping (pQM), and generalized Pareto Distribution quantile mapping (gpQM), the latter adapted for a more specific treatment of extreme values in the quantile adjustment. Furthermore, we investigate the feasibility of combining different calibration techniques for the same location, taking into account various weather types. Next, we optimize the performance of these calibration techniques by employing user-defined validation indices. These indices are globally assessed using a weighted Ranking Framework score, enabling us to identify the optimal combination of techniques for site-based calibration.

2 Data and methods

2.1 Data

The reference observations used as the predictand for calibration were obtained from the Pacific Rainfall Database (PACRAIN, Greene et al., 2008). The PACRAIN Database comprises daily and monthly rainfall records from a comprehensive collection of rain gauge stations situated across atolls and islands in the South Pacific region. These records are sourced from various institutions, including the National Institute of Water and Atmospheric Research of New Zealand (NIWA, www.niwa.cri.nz), the US National Centers for Environmental Information (NCEI, https://www.ncei.noaa.gov/), the French Polynesian Meteorological Service (https://meteo.pf), the Schools of the Pacific Rainfall Climate Experiment (SPaRCE, https://sparce.ou.edu), and the Atlas of Pacific Rainfall (Taylor, 1973). The set of stations used in this study is indicated in Table 1.

The calibrated dataset in this study is the Tropical Rainfall Measuring Mission 3B42 Daily product (TRMM TMPA Precipitation L3 1 day 0.25 $\times$ 0.25 degree V7, Huffman et al., 2016, https://disc.gsfc.nasa.gov/datasets/TRMM_3B42_Daily_7/summary). This dataset provides measurements of daily accumulated precipitation, covering the period from January 1, 1998, to January 1, 2020, with a temporal resolution of one day. The spatial coverage of the dataset ranges from $50.0°$N to $50.0°$S and $180.0°$E to $180.0°$W. For the calibration process, the TRMM data were extracted at the nearest grid point to each PACRAIN rain gauge location (Table 1).

2.2 Weather typing

The adaptive calibration approach in this study utilizes five weather types (WTs) derived from the study conducted by Mirones et al. (2022), based on principal component analysis and *k-means* clustering, and using precipitation and atmospheric circulation variables derived from sea-level pressure and wind reanalysis fields, over the South Pacific Convergence Zone (SPCZ). The weather typing employed in this study offers a valuable representation of the dominant synoptic patterns observed in the study

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

Table 1. Final set of rain gauge stations from the PACRAIN database used in this study. The columns provide information such as the PACRAIN ID, indicating the data source (NZ for NIWA, US for NCEI, and SP for SPaRCE), station name and location, longitude and latitude coordinates in degrees, time coverage of the time series (start and end dates; an asterisk indicates data outside the TRMM period, which were discarded in this study), percentage of missing data within the start-end period, and elevation in meters above sea level.

Station ID	Station Name	Longitude	Latitude	Start	End	% Missing Data	Altitude
NZ75400	Kolopelu (Wallis and Futuna)	178.12 ° W	14.32° S	1998-01-01	2012-01-01	9.74	36
NZ82400	Alofi (Niue)	169.93° W	19.07° S	1998-01-01*	2010-09-02	2.68	59
NZ84317	Rarotonga (Cook Islands)	159.80° W	21.20° S	1999-09-28	2012-01-02	11.36	4
NZ99701	Raoul Island (New Zealand)	177.93° W	29.23° S	1998-01-01*	2012-01-01	0.72	49
SP00646	Port Vila (Vanuatu)	168.30° E	17.72° S	2000-01-26	2013-06-01	18.13	24
US14000	Aoloau (American Samoa)	170.77° W	14.30° S	1998-01-01*	2019-12-31*	21.72	408
US14690	Nu'uuli (American Samoa)	170.70° W	14.32° S	1998-01-01*	2019-12-31*	0.037	3

region in relation to precipitation. The identification of five distinct daily weather types and their relatively balanced sample sizes ensures a robust conditioning of the calibration process, thereby enhancing the reliability and stability of the calibration results.

2.3 Bias correction techniques

The calibration techniques used for the adaptive methodology include scaling, empirical quantile mapping (eQM), parametric quantile mapping (pQM), and generalized Pareto distribution quantile mapping (gpQM). A more detailed description of the methods is provided in Appendix A.

The scaling technique is applied to the raw TRMM data by multiplying it with a correction factor. This factor is computed as the ratio between the mean of the predictand (PACRAIN rain gauge measurements) and the mean of the raw TRMM measurements during the training period.

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

The eQM method is an adaptation of the approach presented in Themeßl et al. (2011), which utilizes empirical cumulative distribution functions (eCDFs) for calibration. In its parametric version (pQM), the QM method relies on the theoretical distribution rather than the empirical one, whose parameters are estimated based on the observed and TRMM data. In particular, here it is assumed that both the observed and simulated intensity distributions can be well approximated by the biparametric gamma distribution (Piani et al., 2010), and therefore both shape and scaling parameters need to be estimated for transfer function fitting.

The gpQM approach also utilizes quantile mapping but incorporates the generalized Pareto distribution (GPD) above a certain threshold (Gutjahr and Heinemann, 2013). The threshold, denoted as u, represents the percentile above which the GPD is used to adjust the wet-day distribution. Below the threshold, the distribution is adjusted to a gamma distribution following the pQM method. This method aims to improve the performance of pQM in the upper tail of the distribution, specifically for extreme events. In this work, two different thresholds are selected: the 95th and the 75th percentiles, resulting in the methods named gpQM-95 and gpQM-75, respectively.

2.4 Adaptive calibration methodology

Here, we introduce an adaptive methodology developed for the calibration of various calibration methods, namely scaling, eQM, pQM, gpQM-95, and gpQM-75. This methodology involves applying the calibration methods individually to each weather type (WT). Subsequently, the best calibration method for each WT is selected, and the calibrated series are combined to form a unified time series spanning the entire calibration period. For each model fit, the calibrated series are obtained following a cross validation scheme. Afterwards, a number of precipitation-derived indices are computed following the validation framework of the Action Cost VALUE (Maraun et al., 2015, see Table 2), and the TRMM-calibrated indices are then compared against the reference observations using simple measures such as relative/absolute bias or correlation. In order to facilitate a comprehensive evaluation and inter-comparison of these methods, we have employed a standardized score calculation methodology able to integrate into one single score the different aspects of the validation, for each calibration method and WT, as outlined next.

To determine the best method for each WT, we utilize a Ranking Framework (RF) score, which is based on the methodology described in Kotlarski et al. (2019). The computation of this score involves several steps. Firstly, we calculate the bias of each calibration method with respect to the reference observations by taking the absolute differences between each of the index values (Table 2) of the reference observations (X_i) and the calibrated TRMM series for each method ($Y_{i,j}$):

$$Z_{i,j} = |X_i - Y_{i,j}|. \tag{1}$$

Next, we normalize the bias values obtained from all calibrations (j) for a given index (i), such that lower values are considered better by the normalization:

$$Z'_{i,j} = 1 - \frac{Z_{i,j} - Z_{i,\text{min}}}{Z_{i,\text{max}} - Z_{i,\text{min}}}. \tag{2}$$

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

Finally, the RF score for each method is calculated as the average of the normalized values for all indices:

$$RF_j = \frac{1}{N}\sum_{i=1}^{N} w_i \cdot Z'_{i,j} \quad \text{where} \quad \sum_{i=1}^{N} w_i = 1 \tag{3}$$

and N represents the total number of indices evaluated ($N = 10$, see Table 2). Thus, in the calculation of the score, it is possible to incorporate arbitrary (unit normalized) weights w_i for the normalized climate indices. This allows for explicit consideration of validation aspects that may carry greater importance, with higher weights assigned to those aspects to determine the final score. A common example in many hydrological applications is the significance of accurately representing extreme precipitation events following calibration. Therefore, specific extreme indices (e.g., *P98Wet* or *P98WetAmount*, as shown in Table 2) can be given a higher relative weight in their contribution to the overall score, thereby reflecting their increased relevance in the calibration method ranking process.

To ensure robustness and avoid artificial skill, we employ a cross-validation scheme for model fitting. This scheme enables us to assess the consistency of the calibration results beyond the training period by using a separate test period for prediction (Efron and Gong, 1983). Specifically, we employ the classical k-fold cross-validation, ensuring that each fold contains a minimum of 275 samples for a robust training.

All the calibration methods have been run using the implementation available in the package *downscaleR* (Bedia et al., 2020) from the open-source *climate4R* framework for climate data analysis and visualization (Iturbide et al., 2019). The different evaluation indices presented in Table 2 have been computed using the standard definitions of the VALUE Framework (Maraun et al., 2015), which are implemented in the R package *VALUE*[1].

Table 2. Summary of the validation indices and measures used in the study, along with their corresponding codes as defined in the VALUE reference list (`http://www.value-cost.eu/validationportal/app/#!indices`). The indices and measures serve as evaluation metrics for assessing the performance and accuracy of the calibration techniques in the study.

Code	Description	Type
Skewness	Skewness	index
Mean	Mean	index
SDII	Mean wet-day ($\geq$ 1mm) precipitation	index
R10	Relative frequency of days with precip $\geq$ 10mm	index
R10p	Precipitation amount falling in days with precip $\geq$ 10mm	index
R20	As R10, but considering a 20mm threshold	index
R20p	As R10p, but considering a 20 mm threshold	index
P98Wet	98th percentile of wet ($\geq$1 mm) days	index
P98WetAmount	Total amount above 98th percentile of wet ($\geq$1 mm) days	index
RV20_max	Maximum Daily precipitation for a 20-year Return Value	index
absolute bias		measure
relative bias		measure
Spearman's rank correlation		measure

[1] https://github.com/SantanderMetGroup/VALUE

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

3 Results and discussion

3.1 Standard and weather-type conditioned calibration method intercomparison

To obtain a comprehensive evaluation of the method's overall performance, we initially focus on an unconditioned intercomparison, referred to as "standard calibration" hereafter. This evaluation involves assessing the method across the entire time series without considering different weather types. The preliminary findings indicate the presence of low to moderate (negative) biases in the TRMM product. As an illustrative example, we present the results obtained at the Kolopelu station in Fig. 1 (the upper triangle represents the standard calibration results), which are representative of the overall outcomes observed at other locations (the corresponding plots for the remaining rain gauge locations are included in the Appendix). While certain TRMM indices show negligible biases compared to the rain gauge stations (such as skewness and SDII), others exhibit significant relative biases, particularly for representing high precipitation events (such as R10p, R20p, or P98WetAmount). These findings highlight the necessity of applying some form of calibration to enhance the accuracy of TRMM for impact studies. In the same vein, we include the ERA5 precipitation series to highlight the strong biases associated with this reanalysis product.

In contrast to expectations, scaling, which is a common and straightforward technique for TRMM correction, was found to be ineffective in mitigating biases in most specific indices related to intense precipitation events and some others related to mean precipitation such as SDII. Instead of improving the situation, scaling had an overall deleterious effect, underscoring the critical importance of considering alternative calibration techniques better suited to TRMM adjustment (Fig. 1).

In the analysis of the remaining (quantile-mapping based) techniques, Figure 1 illustrates that there are no significant variations in their performance, although this may depend on the specific index or technique being analyzed. Likewise, the relative bias of the indices between the standard calibration and the conditioned calibration generally exhibits minimal to moderate differences. This finding confirms and extends the results previously reported by Mirones et al. (2022) for scaling and eQM, to include the parametric quantile mapping variants pQM or gpQM.

The final column in Figure 1 compares the best-performing conditioned technique in each case with the newly developed adaptive methodology in this study, which combines the optimal calibration technique for each weather type individually. In general, the adaptive methodology surpasses the results achieved by the best-conditioned calibration. The most notable reduction in relative bias is observed in indices that measure high rainfall amounts, such as R10p, R20p, or P98WetAmount. This improvement is significant because conditioned calibration alone did not exhibit substantial enhancements, except for specific cases involving techniques like gpQM. However, only three indices (mean, SDII, and P98Wet) did not show improvement with the adaptive approach and remained nearly unchanged. In summary, the overall results indicate that the adaptive calibration method offers improved adjustment in the upper tail of the distribution, which is where TRMM exhibits the most significant biases. This calibration methodology facilitates enhanced adjustment for extreme precipitation events, with a specific focus on high precipitation indices. Next, in Sec. 3.2, we present a more in-depth analysis of the detailed results obtained from the adaptive calibration approach.

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

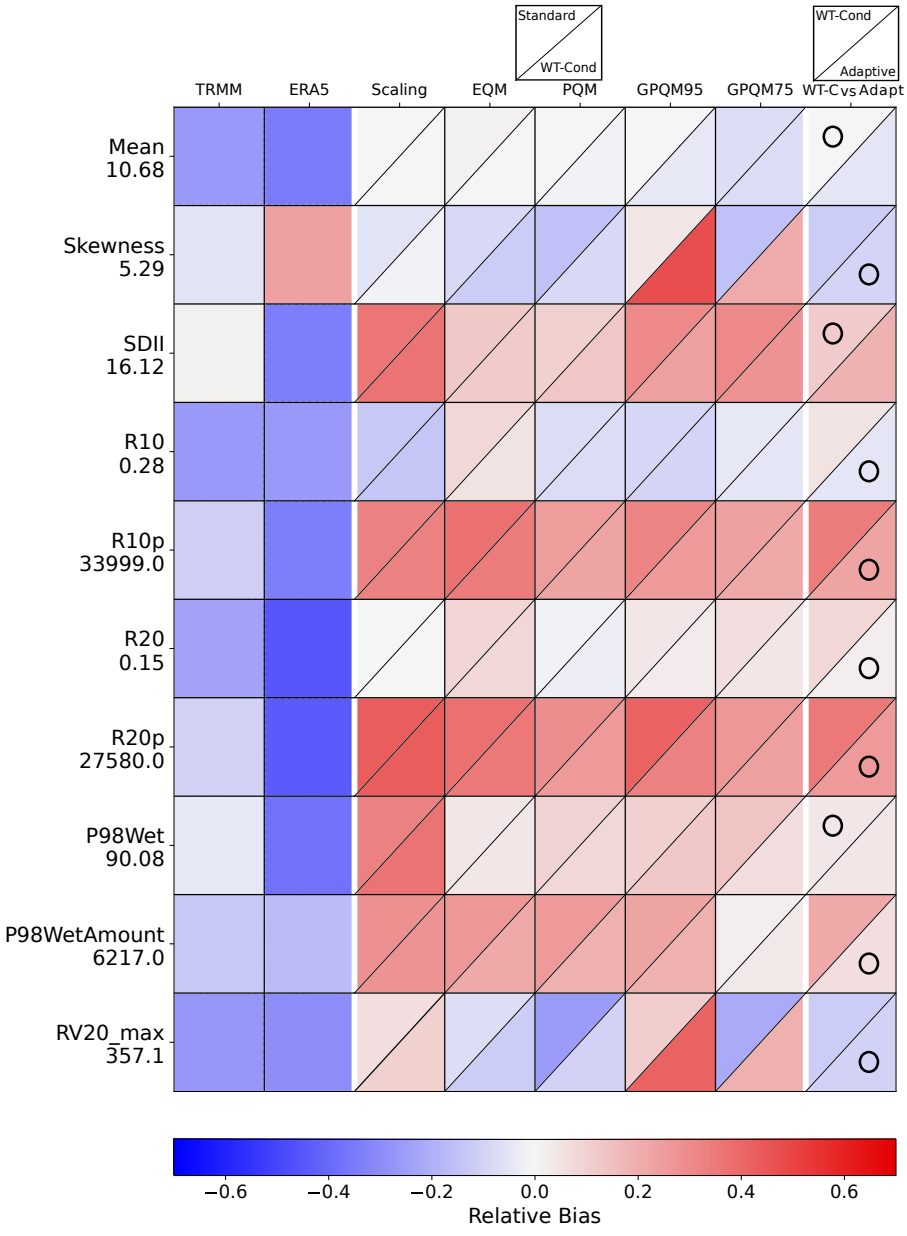

Figure 1. Relative Biases of the climate indices used for validation (Table 2 of raw TRMM data and TRMM-calibrated data, at the Kolopelu Station grid box (Table 1). As an additional reference, we include in the second column the biases of the ERA5 reanalysis raw precipitation data (Hersbach et al., 2020). The calibration techniques are scaling, eQM, pQM, GPQM-95, and GPQM-75 (Sec. 2.3). For each method plot cell, the upper triangle displays the relative bias of the standard calibration, while the lower triangle represents the WT-conditioned approach. The last column presents shows a comparison between the relative bias of the best WT-conditioned technique (eQM at the Kolopelu site) vs. the adaptive calibration. The circle indicates the best-performing approach with the lowest relative bias. The Y-axis labels show the actual index values from the rain gauge observations beneath the index names.

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

3.2 Adaptive calibration

To evaluate the performance of the adaptive approach in comparison to the WT-conditioned method, we introduce the RF score as a comprehensive measure that accounts for the various indicators described earlier (Table 2). This global performance measure allows for an easier ranking of methods. The summarized results, considering both unweighted and weighted RF values (the latter emphasizing the significance of extreme indicators in the evaluation process as discussed in Section 2.4), are presented in Figure 2.

First of all, we undertake a method intercomparison using the unweighted RF score as a ranking measure (Fig. 2a). In most stations the adaptive calibration outperforms the standard and WT-conditioned techniques, with some exceptions (Port Vila and Nuu'uli), in which the adaptive approach has a similar performance than the simpler ones. On the other hand, the adaptive method is clearly superior at some other locations like Kolopelu, Alofi, or Raoul Island, where the RF scores obtained with the adaptive calibration are significantly higher than the best scores obtained with standard and WT-conditioned calibration methods. While the score for the best standard technique at these stations exceeds 0.60, the adaptive calibration achieves values between 0.80 and 0.90, representing an improvement of 33-50%. This significant enhancement in calibration, based on the climate indices utilized in the adaptive approach, demonstrates an overall improvement that justifies the application of the adaptive method. Furthermore, it is important to note that in the worst-case scenario, the adaptive approach will nearly match (and never significantly impair) the performance of the calibration.

As mentioned earlier, it is also possible to assign arbitrary weights to the indices involved in the RF score, giving more importance to specific precipitation characteristics, such as the representation of extremes. In this study, we selected index weights that prioritize high rainfall indices (Table A2). This weighting aims to guide the calibration towards better adjustment in the upper tail of the distribution, thereby achieving improved correction for extreme precipitation events beyond a certain threshold. In this way, the influence of high rainfall indices benefits methods like gpQM, which specialize in adjusting the upper tail using a GPD (Generalized Pareto Distribution). The findings are illustrated through the boxplots presented in Figure 2. It is evident that the scores of gpQM95 and gpQM75 exhibit higher values in the weighted version (Figure 2b) compared to the unweighted version (Figure 2a).

The analysis of the RF weighting configuration demonstrates its dual impact: not only does it affect the overall score, but it also influences the selection of techniques for each weather type. In our weighting scheme, which prioritizes superior performance in extreme precipitation indices, the gpQM approaches emerge sometimes as the favored choice after applying these weights, for instance at Aoloau site for WTs 1 to 3 (Fig. 3). This outcome establishes the superiority of the gpQM technique over the conventional quantile mapping methods, namely eQM and pQM when the representation of extreme indices is prioritized. Consequently, the adaptive calibration score at this station improves from approximately 0.65 to 0.85, representing a 30% increase, with the best standard calibration method changing from eQM and pQM to gpQM75 in this particular case.

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

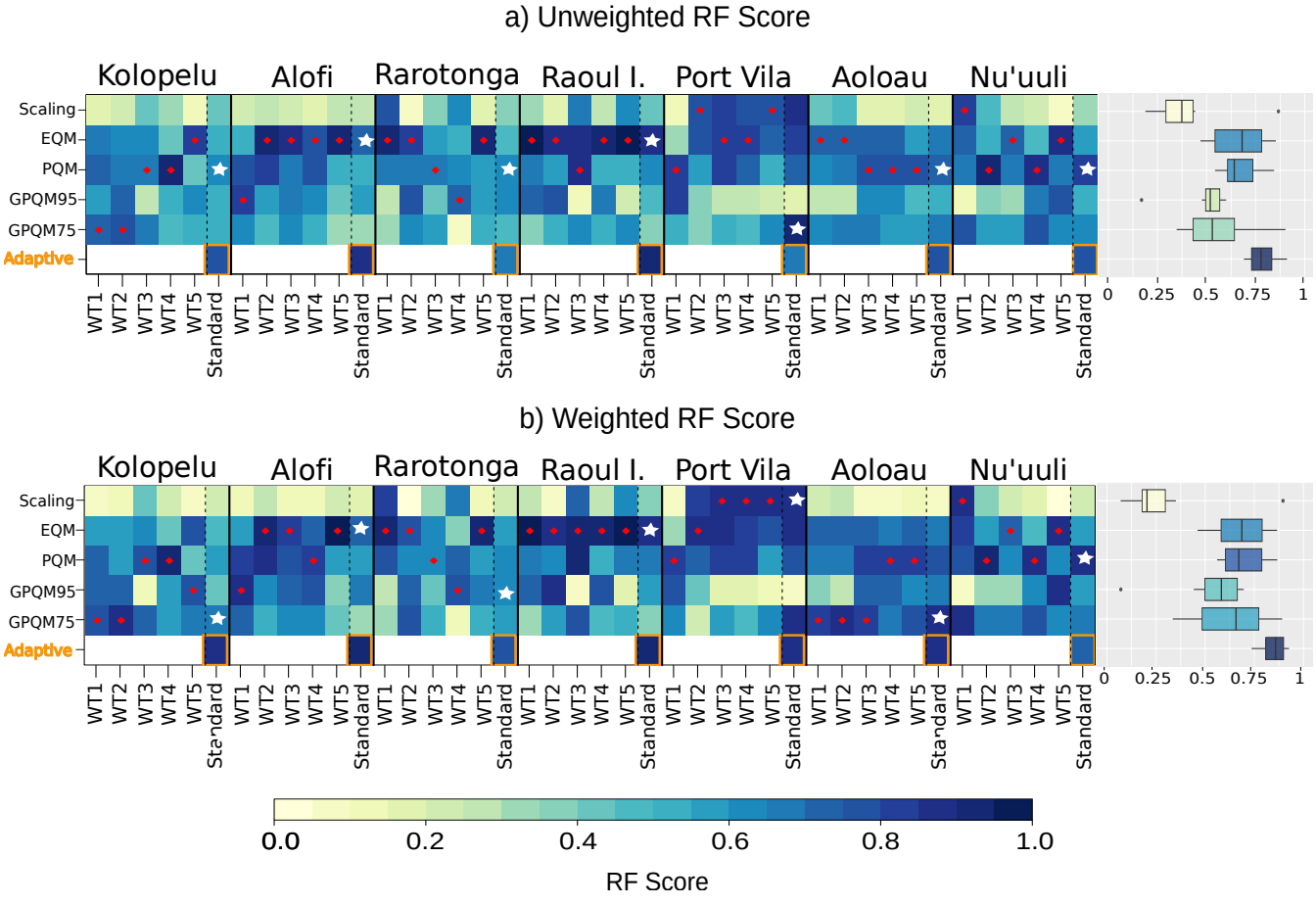

Figure 2. **(a)**: Ranking Framework (RF) score results of the adaptive calibration for each WT and site. The red dots indicate the best method for the corresponding WT, while the white stars represent the best method for the standard calibration (the same technique applied over the entire period without conditioning). On the right side, the standard approach score for each method is represented in the box plot (their mean represented by the color of the boxes). **(b)**: Similar to **(a)**, but with the addition of different weights (refer to A2) in the computation of climate indices for the RF score.

At the Port Vila station, we observe another interesting situation. In Figure 2a, the adaptive calibration score is lower compared to the score of gpQM75. However, when applying weighting with a focus on high rainfall indices (Figure 2b), the adaptive calibration undergoes enhancements and achieves a competitive score higher than the unweighted version. It is worth noting that while the inclusion of weights leads to changes and improvements in the adaptive calibration for certain stations, it has no effect on others. For instance, stations like Rarotonga or Nu'uli exhibit no changes in the composition of the adaptive calibration, regardless of the weighting applied.

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

Therefore, the results show that adaptive method consistently performs better than the rest in all stations, as highlighted in the boxplots in Figure 2, attaining higher scores. Additionally, the adaptive calibration demonstrates a narrower interquartile range (IQR) compared to the other methods in Figure 2a, indicating lower variability. Only gpQM95 calibration shows a narrower IQR range, but with significantly lower scores. We attribute this poor result to the limited robustness of the fit of the extreme function due to the high percentile threshold, which greatly reduces the sample size (see Table A1). The other methods individually considered exhibit wider RF variability ranges and lower values as compared to the adaptive approach.

In conclusion, the adaptive calibration method improves upon the results obtained with the WT-conditioned methodology presented in Mirones et al. (2022) or at least, in the worst case, it maintains the calibration performance. The adaptive calibration method showcases competitive performance in effectively calibrating the TRMM data at the target stations, thereby promoting consistency in the results across diverse locations. Furthermore, the capability to customize the calibration by applying arbitrary weights to specific indices offers increased flexibility in determining the optimal combination of methods that align with the unique characteristics of each site. This adaptability further enhances the overall calibration process.

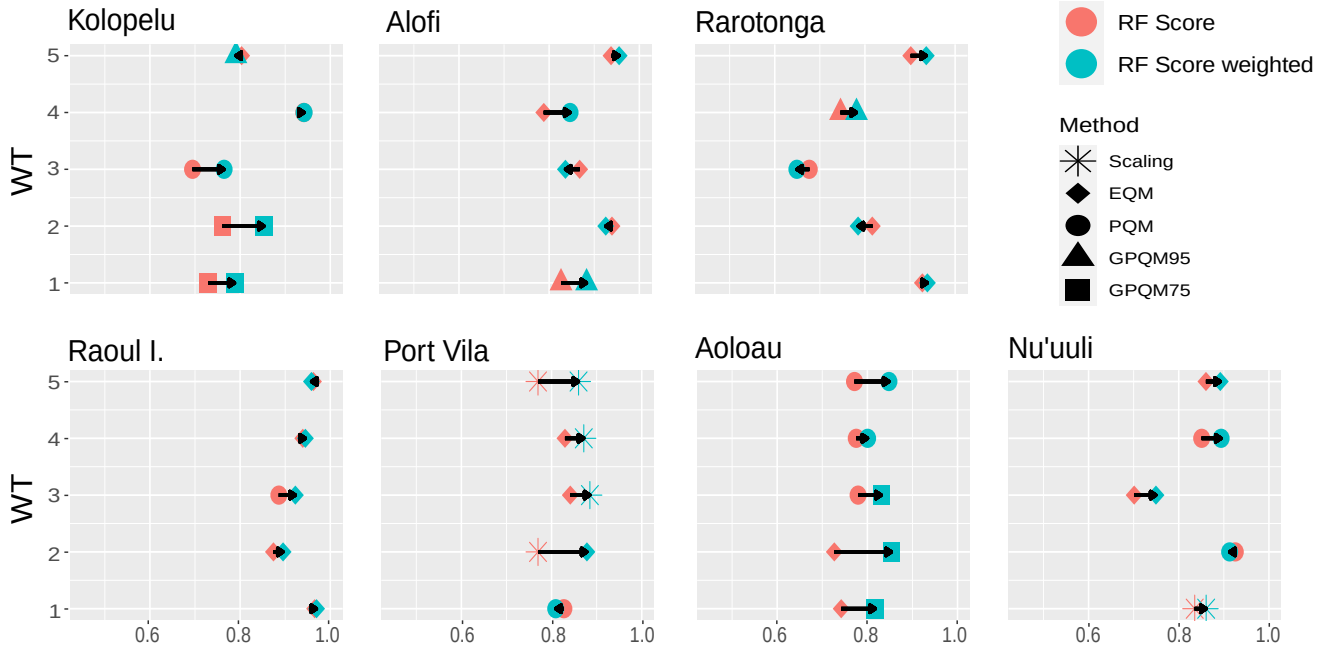

Figure 3. Differences between unweighted and weighted RF scores attained for each weather type at each of the target locations. The technique associated to the highest score is also indicated by the key of symbols.

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

4 Conclusions

We intercompare a range of bias-adjustment techniques for the calibration of daily Tropical Rainfall Measuring Mission (TRMM) precipitation data, building upon a set of rain-gauge stations scattered across the South Pacific region, spanning the period 1998–2019. The calibration techniques evaluated in this study include empirical quantile mapping (eQM), parametric quantile mapping (pQM), generalized Pareto distribution quantile mapping (gpQM) and scaling, the latter used as a benchmark since it is the most common approach for this task. An adaptive calibration methodology has been developed based on weather type (WT) conditioning, which selects the best-performing calibration technique for each specific WT.

Building upon the methodology proposed by Mirones et al. (2022), we extend it to encompass an expanded set of calibration techniques and adopt a more adaptable approach to suit the distinct characteristics of each weather type (WT). This extension results in a calibrated series that capitalizes on the individual strengths of each technique, tailored to specific situations. The methodology entails the selection of an optimal method for each WT, guided by a comprehensive performance measure (RF score) that encompasses various precipitation indices. The minimal spatial variability observed in the results obtained through the adaptive method enhances its potential for a robust application across different locations. This characteristic holds importance in hydrological studies where spatial consistency between locations at the basin level is typically desired. Moreover, the adaptive approach allows users to define and configure the weights assigned to different indices, affording flexibility in the assessment of each method. This comprehensive approach ensures the utilization of the most suitable techniques for each WT, resulting in an enhanced final calibrated series.

Our findings unequivocally establish the superiority of the adaptive calibration methodology over the best WT-conditioned calibration technique in terms of relative bias. Notably, the most substantial improvements are observed in accumulated precipitation indices, specifically R10p, R20p, and P98WetAmount. These indices hold great significance within the realm of hydrological modeling and climate impact studies, making the observed enhancements particularly relevant. The adaptive calibration methodology offers a promising avenue for refining and improving the accuracy of precipitation data from indirect measures such as the TRMM database, thereby enhancing the reliability of subsequent hydrological and climate impact assessments.

Code and data availability. An interactive notebook associated with this study is available at the following link: https://github.com/SantanderMetGroup/notebooks/tree/2023_TRMM_adaptiveCal/2023_adaptiveCalibration. This notebook provides a comprehensive illustration of the entire adaptive calibration process, including available data download from an open repository and the computation of both standard and adaptive calibration RF scores.

Acknowledgements. This paper is part of the R+D+i project CORDyS (PID2020-116595RB-I00) with funding from the Spanish Ministry of Science MCIN/AEI/10.13039/501100011033. O.M. has received research support from grant PID2020-116595RB-I00 funded by

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

MCIN/AEI/10.13039/501100011033. We also thank our colleague Dr. Javier Díez-Sierra for sharing code snippets used to develop some of the figures.

260 *Competing interests.* The authors declare no competing interests.

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

References

Aghakouchak, A., Nasrollahi, N., and Habib, E.: Accounting for Uncertainties of the TRMM Satellite Estimates, Remote Sensing, 1, 606–619, https://doi.org/10.3390/rs1030606, 2009.

Almazroui, M.: Calibration of TRMM rainfall climatology over Saudi Arabia during 1998–2009, Atmospheric Research, 99, 400–414, https://doi.org/10.1016/j.atmosres.2010.11.006, 2011.

Australian Bureau of Meteorology and CSIRO: Climate change in the Pacific: scientific assessment and new research. Volume 1: Regional Overview, Pacific Climate Change Science Program, Aspendale, Victoria, https://www.pacificclimatechangescience.org/publications/reports/report-climate-change-in-the-pacific-scientific-assessment-and-new-research, 2011.

Baltaci, H., Gokturk, O. M., Kındap, T., Unal, A., and Karaca, M.: Atmospheric circulation types in Marmara Region (NW Turkey) and their influence on precipitation, International Journal of Climatology, 35, 1810–1820, https://doi.org/https://doi.org/10.1002/joc.4122, 2015.

Bedia, J., Baño-Medina, J., Legasa, M. N., Iturbide, M., Manzanas, R., Herrera, S., Casanueva, A., San-Martín, D., Cofiño, A. S., and Gutiérrez, J. M.: Statistical downscaling with the downscaleR package (v3.1.0): contribution to the VALUE intercomparison experiment, Geoscientific Model Development, 13, 1711–1735, https://doi.org/10.5194/gmd-13-1711-2020, 2020.

Efron, B. and Gong, G.: A leisurely look at the bootstrap, the jackknife, and cross-validation, The American Statistician, 37, 36–48, 1983.

Greene, J. S., Klatt, M., Morrissey, M., and Postawko, S.: The Comprehensive Pacific Rainfall Database, Journal of Atmospheric and Oceanic Technology, 25, 71–82, https://doi.org/10.1175/2007JTECHA904.1, 2008.

Gutjahr, O. and Heinemann, G.: Comparing precipitation bias correction methods for high-resolution regional climate simulations using COSMO-CLM: Effects on extreme values and climate change signal, Theoretical and Applied Climatology, 114, 511–529, https://doi.org/10.1007/s00704-013-0834-z, 2013.

Hay, L. E., McCabe Jr., G. J., Wolock, D. M., and Ayers, M. A.: Simulation of precipitation by weather type analysis, Water Resources Research, 27, 493–501, https://doi.org/https://doi.org/10.1029/90WR02650, 1991.

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., De 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., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, 2020.

Huffman, G., Bolvin, D., Nelkin, E., and Adler, R.: TRMM (TMPA) Precipitation L3 1 day 0.25 degree x 0.25 degree V7, Edited by A. Savtchenko. Goddard Earth Sciences Data and Information Services Center (GES DISC), https://disc.gsfc.nasa.gov/datasets/TRMM_3B42_Daily_7/summary, dOI: 10.5067/TRMM/TMPA/DAY/7, 2016.

Huth, R., Beck, C., Philipp, A., Demuzere, M., Ustrnul, Z., Cahynová, M., Kyselý, J., and Tveito, O. E.: Classifications of atmospheric circulation patterns: recent advances and applications, Annals of the New York Academy of Sciences, 1146, 105–152, https://doi.org/10.1196/annals.1446.019, 2008.

Islam, M. N., Das, S., and Uyeda, H.: Calibration of TRMM Derived Rainfall Over Nepal During 1998-2007, The Open Atmospheric Science Journal, 4, 12–23, https://doi.org/10.2174/1874282301004010012, 2010.

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

Iturbide, M., Bedia, J., Herrera, S., Baño-Medina, J., Fernández, J., Frías, M., Manzanas, R., San-Martín, D., Cimadevilla, E., Cofiño, A., and Gutiérrez, J.: The R-based climate4R open framework for reproducible climate data access and post-processing, Environmental Modelling & Software, 111, 42–54, https://doi.org/10.1016/j.envsoft.2018.09.009, 2019.

Jury, M. W., Herrera, S., Gutiérrez, J. M., and Barriopedro, D.: Blocking representation in the ERA-Interim driven EURO-CORDEX RCMs, Climate Dynamics, 52, 3291–3306, https://doi.org/10.1007/s00382-018-4335-8, 2019.

Kotlarski, S., Szabó, P., Herrera, S., Räty, O., Keuler, K., Soares, P. M., Cardoso, R. M., Bosshard, T., Pagé, C., Boberg, F., Gutiérrez, J. M., Isotta, F. A., Jaczewski, A., Kreienkamp, F., Liniger, M. A., Lussana, C., and Pianko-Kluczyńska, K.: Observational uncertainty and regional climate model evaluation: A pan-European perspective, International Journal of Climatology, 39, 3730–3749, https://doi.org/https://doi.org/10.1002/joc.5249, 2019.

Manzanas, R. and Gutiérrez, J. M.: Process-conditioned bias correction for seasonal forecasting: a case-study with ENSO in Peru, Climate Dynamics, 52, 1673–1683, https://doi.org/10.1007/s00382-018-4226-z, 2019.

Maraun, D., Widmann, M., Gutiérrez, J. M., Kotlarski, S., Chandler, R. E., Hertig, E., Wibig, J., Huth, R., and Wilcke, R. A.: VALUE: A framework to validate downscaling approaches for climate change studies, Earth's Future, 3, 1–14, https://doi.org/https://doi.org/10.1002/2014EF000259, 2015.

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., Hagemann, S., Richter, I., Soares, P. M. M., Hall, A., and Mearns, L. O.: Towards process-informed bias correction of climate change simulations, Nature Climate Change, 7, 764–773, https://doi.org/10.1038/nclimate3418, 2017.

Mirones, O., Bedia, J., Fernández-Granja, J. A., Herrera, S., Van Vloten, S. O., Pozo, A., Cagigal, L., and Méndez, F. J.: Weather-type-conditioned calibration of Tropical Rainfall Measuring Mission precipitation over the South Pacific Convergence Zone, International Journal of Climatology, pp. 1–18, https://doi.org/https://doi.org/10.1002/joc.7905, 2022.

Piani, C., Haerter, J. O., and Coppola, E.: Statistical bias correction for daily precipitation in regional climate models over Europe, Theoretical and Applied Climatology, 99, 187–192, https://doi.org/10.1007/s00704-009-0134-9, 2010.

Reiter, P., Gutjahr, O., Schefczyk, L., Heinemann, G., and Casper, M.: Does applying quantile mapping to subsamples improve the bias correction of daily precipitation?: DOES QUANTILE MAPPING BENEFIT FROM SUBSAMPLING?, International Journal of Climatology, 38, 1623–1633, https://doi.org/10.1002/joc.5283, 2018.

Riediger, U. and Gratzki, A.: Future weather types and their influence on mean and extreme climate indices for precipitation and temperature in Central Europe, Meteorologische Zeitschrift, 23, 231–252, https://doi.org/10.1127/0941-2948/2014/0519, 2014.

Sekaranom, A. B. and Masunaga, H.: Origins of Heavy Precipitation Biases in the TRMM PR and TMI Products Assessed with CloudSat and Reanalysis Data, Journal of Applied Meteorology and Climatology, 58, 37–54, https://doi.org/10.1175/JAMC-D-18-0011.1, 2019.

Simpson, J., Kummerow, C., Tao, W., and Adler, R. F.: On the Tropical Rainfall Measuring Mission (TRMM), Meteorology and Atmospheric Physics, 60, 19–36, 1996.

Stehlik, J. and Bardossy, A.: Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation, Journal of Hydrology, 256, 120–141, https://doi.org/10.1016/S0022-1694(01)00529-7, 2002.

Taylor, R. C.: An atlas of Pacific islands rainfall., Tech. rep., HAWAII INST OF GEOPHYSICS HONOLULU, 1973.

Themeßl, M. J., Gobiet, A., and Leuprecht, A.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, International Journal of Climatology, 31, 1530–1544, https://doi.org/https://doi.org/10.1002/joc.2168, 2011.

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

Trigo, R. M. and DaCamara, C. C.: Circulation weather types and their influence on the precipitation regime in Portugal, International Journal of Climatology, 20, 1559–1581, https://doi.org/https://doi.org/10.1002/1097-0088(20001115)20:13<1559::AID-JOC555>3.0.CO;2-5, 2000.

Vuillaume, J.-F. and Herath, S.: Improving global rainfall forecasting with a weather type approach in Japan, Hydrological Sciences Journal, 62, 167–181, https://doi.org/10.1080/02626667.2016.1183165, 2017.

Waliser, D. E. and Gautier, C.: A Satellite-derived Climatology of the ITCZ, Journal of Climate, 6, 2162 – 2174, https://doi.org/10.1175/1520-0442(1993)006<2162:ASDCOT>2.0.CO;2, 1993.

Wetterhall, F., Halldin, S., and Xu, C.-Y.: Seasonality properties of four statistical-downscaling methods in central Sweden, Theoretical and Applied Climatology, 87, 123–137, https://doi.org/10.1007/s00704-005-0223-3, 2007.

Wetterhall, F., Pappenberger, F., He, Y., Freer, J., and Cloke, H. L.: Conditioning model output statistics of regional climate model precipitation on circulation patterns, Nonlinear Processes in Geophysics, 19, 623–633, https://doi.org/10.5194/npg-19-623-2012, 2012.

Appendix A: Bias correction methods formulas and complementary information

Here we provide a detailed description of the correction methods used in the study. These methods aim to improve the accuracy of the TRMM rainfall data by incorporating information from PACRAIN rain gauge measurements, used as predictand. The calibration techniques are next described:

Equation A1 presents the scaling method, where $\hat{p}_{trmm}$ represents the corrected TRMM rainfall, p_{rg} and p_{trmm} denote the PACRAIN rain gauge and raw TRMM measurements, respectively, and $\bar{P}_{rg}$ and $\bar{P}_{trmm}$ are the means of the p_{rg} and p_{trmm} series.

$$\hat{p}_{trmm} = p_{trmm} \frac{\bar{P}_{rg}}{\bar{P}_{trmm}} \tag{A1}$$

Equation A2 describes the empirical quantile mapping (eQM) method. Here, $\hat{X}_{t,i}$ represents the corrected value for a specific day and grid, $\hat{F}^{trmm}_{doy,i}$ and $\hat{F}^{rg}_{doy,i}$ are the empirical cumulative distribution functions (eCDFs) for TRMM and PACRAIN, respectively, corresponding to the given day of the year (doy), and $X_{t,i}$ is the uncorrected value.

$$\hat{X}_{t,i} = \hat{F}^{rg^{-1}}_{doy,i}(\hat{F}^{trmm}_{doy,i}(X_{t,i})), \tag{A2}$$

Equation A3 presents the parametric quantile mapping (pQM) method. $F^{trmm}_{doy,i}$ and $F^{rg}_{doy,i}$ represent the assumed theoretical distributions for TRMM and PACRAIN, respectively, and $X_{t,i}$ is the uncorrected value.

$$\hat{X}_{t,i} = F^{rg^{-1}}_{doy,i}(F^{trmm}_{doy,i}(X_{t,i})), \tag{A3}$$

Lastly, Equation A4 outlines the generalized parametric quantile mapping (gpQM) method. It uses a combination of gamma and generalized Pareto distribution (GPD) to correct the TRMM rainfall values based on their percentiles. $F^{trmm,gamma}_{doy,i}$ and $F^{rg,gamma}_{doy,i}$ are the gamma cumulative distributions for TRMM and PACRAIN, while $F^{trmm,GPD}_{doy,i}$ and $F^{rg,GPD}_{doy,i}$ represent the GPDs for TRMM and PACRAIN, respectively. The threshold of the 95th percentile is used to differentiate between the two

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

distributions.

$$\hat{X}_{t,i} = \begin{cases} F_{doy,i}^{rg,gamma^{-1}}(F_{doy,i}^{trmm,gamma}(X_{t,i})) & \text{if } X_{t,i} < 95\text{th percentile} \\ F_{doy,i}^{rg,GPD^{-1}}(F_{doy,i}^{trmm,GPD}(X_{t,i})) & \text{if } X_{t,i} \geq 95\text{th percentile} \end{cases} \tag{A4}$$

Table A1. Overview of observations (days) for each WT across multiple stations. Each row corresponds to a specific station, while the columns represent different WTs. The table displays the total number of observations recorded for each WT, along with the corresponding 75th and 95th percentiles.

Station	WT1			WT2			WT3			WT4			WT5		
	N	*P75*	*P95*	*N*	*P75*	*P95*	*N*	*P75*	*P95*	*N*	*P75*	*P95*	*N*	*P75*	*P95*
Kolopelu	966	242	48	775	194	39	851	213	43	826	206	41	306	76	15
Alofi	1140	285	57	932	233	47	1023	256	51	917	229	46	437	109	22
Rarotonga	1105	276	55	916	229	46	1055	264	53	957	239	48	391	98	20
Raoul Island	1327	332	66	1058	264	53	1158	290	58	1076	269	54	458	114	23
Port Vila	1084	271	54	884	221	44	1010	252	50	849	212	42	370	92	18
Aoloau	1091	273	55	874	218	44	999	250	50	826	206	41	428	107	21
Nu'uli	2076	519	104	1721	430	86	1846	462	92	1570	392	78	820	205	41

Table A2. Index weights w_i (see Eq. 3) for the weighted RF score calculation in the adaptive calibration technique selection.

Code	Weight
Skewness	0.05
Mean	0.05
SDII	0.05
R10	0.05
R10p	0.05
R20	0.15
R20p	0.15
P98Wet	0.15
P98WetAmount	0.2
RV20_max	0.1

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

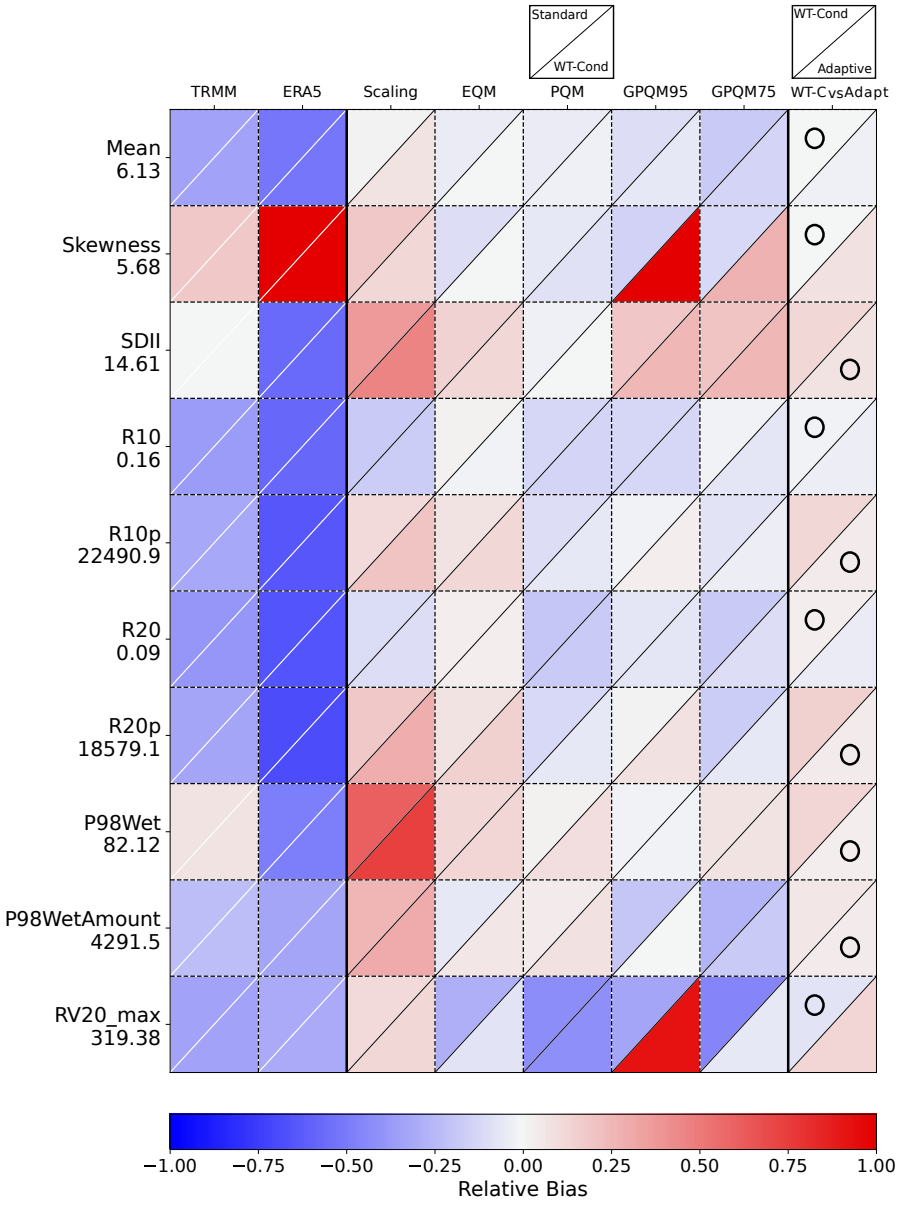

Figure A1. Relative Biases of the climate indices used for validation (Table 2 of raw TRMM data and TRMM-calibrated data, at the Alofi Station grid box (Table 1). As an additional reference, we include in the second column the biases of the ERA5 reanalysis raw precipitation data (Hersbach et al., 2020). The calibration techniques are scaling, eQM, pQM, GPQM-95, and GPQM-75 (Sec. 2.3). For each method plot cell, the upper triangle displays the relative bias of the standard calibration, while the lower triangle represents the WT-conditioned approach. The last column presents shows a comparison between the relative bias of the best WT-conditioned technique vs. the adaptive calibration. The circle indicates the best-performing approach with the lowest relative bias. The Y-axis labels show the actual index values from the rain gauge observations beneath the index names.

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

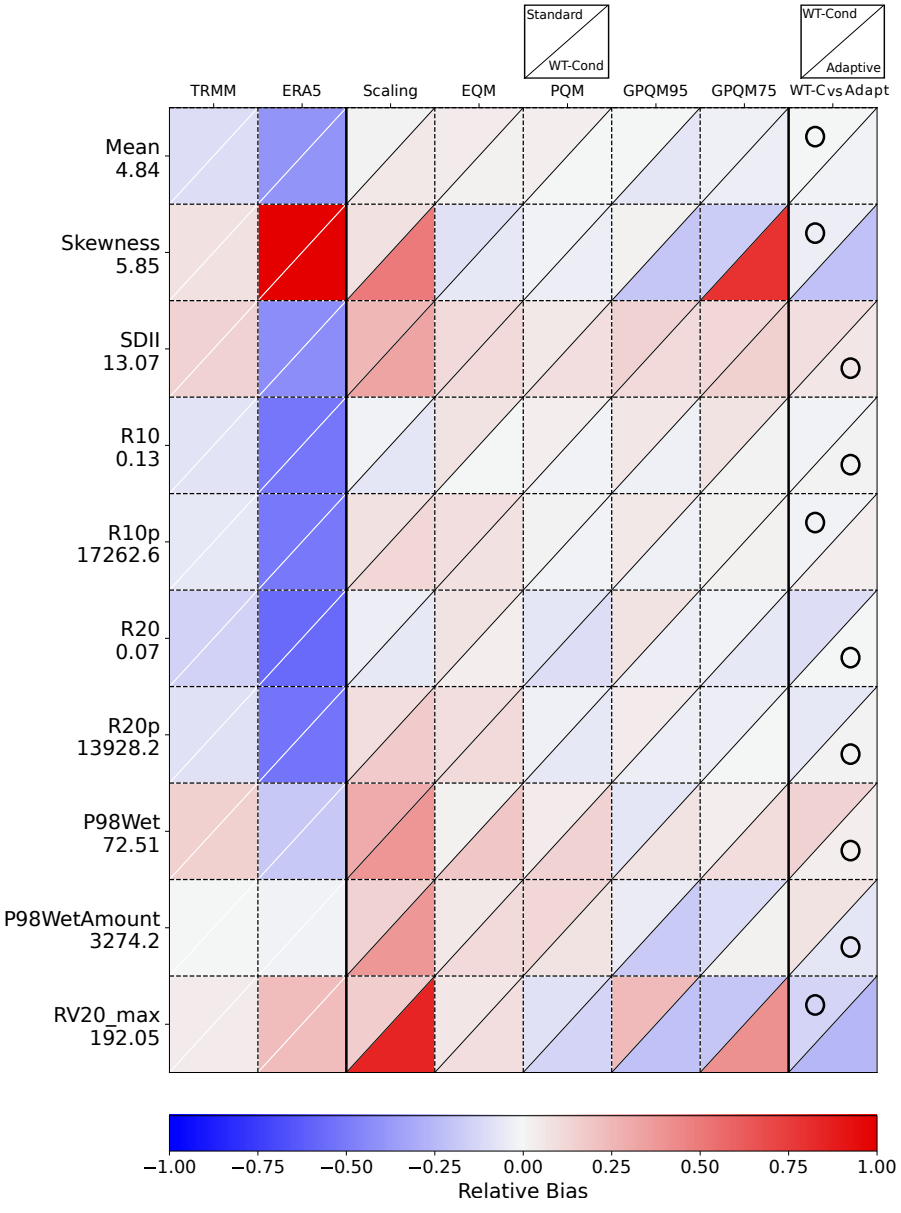

Figure A2. Same as Fig. A1, but for the Rarotonga rain gauge location.

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

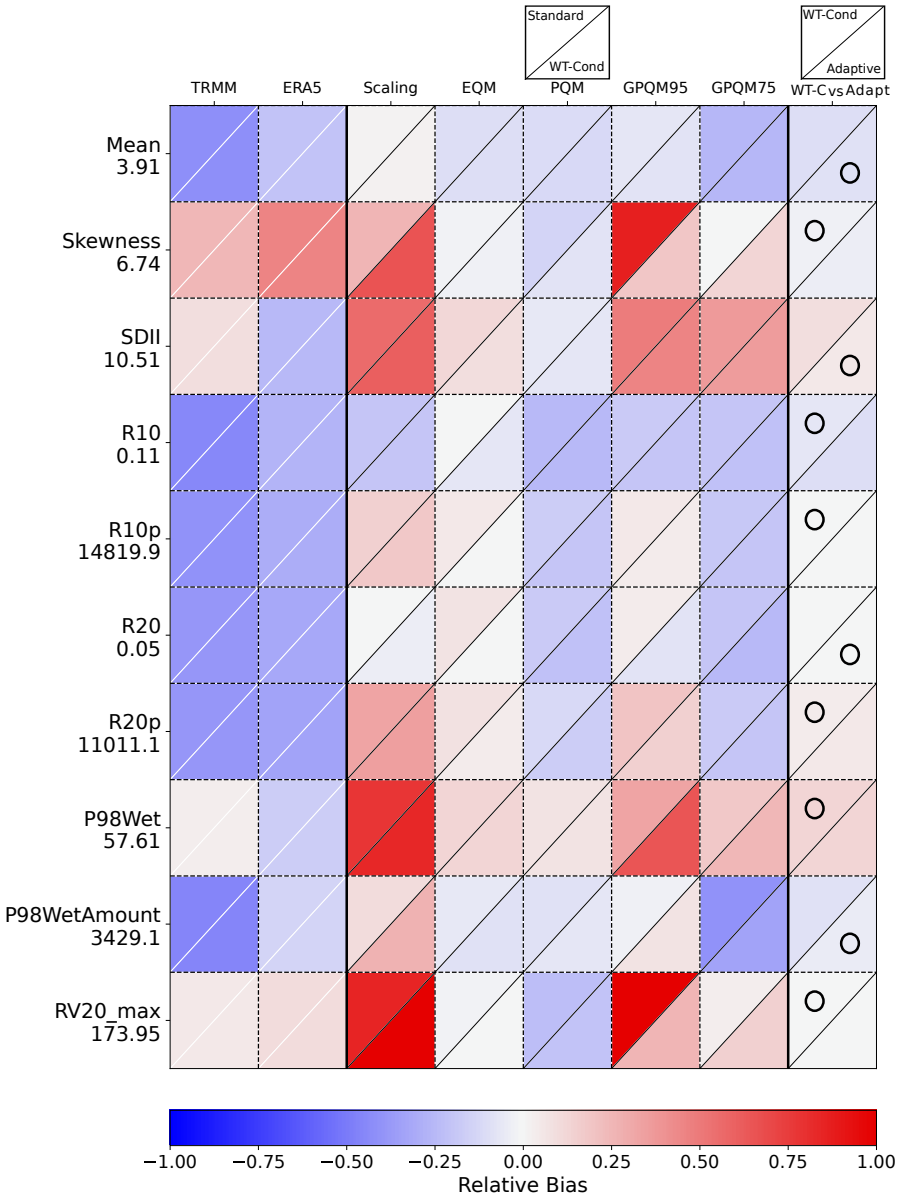

Figure A3. Same as Fig. A1, but for the Raoul Island rain gauge location.

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

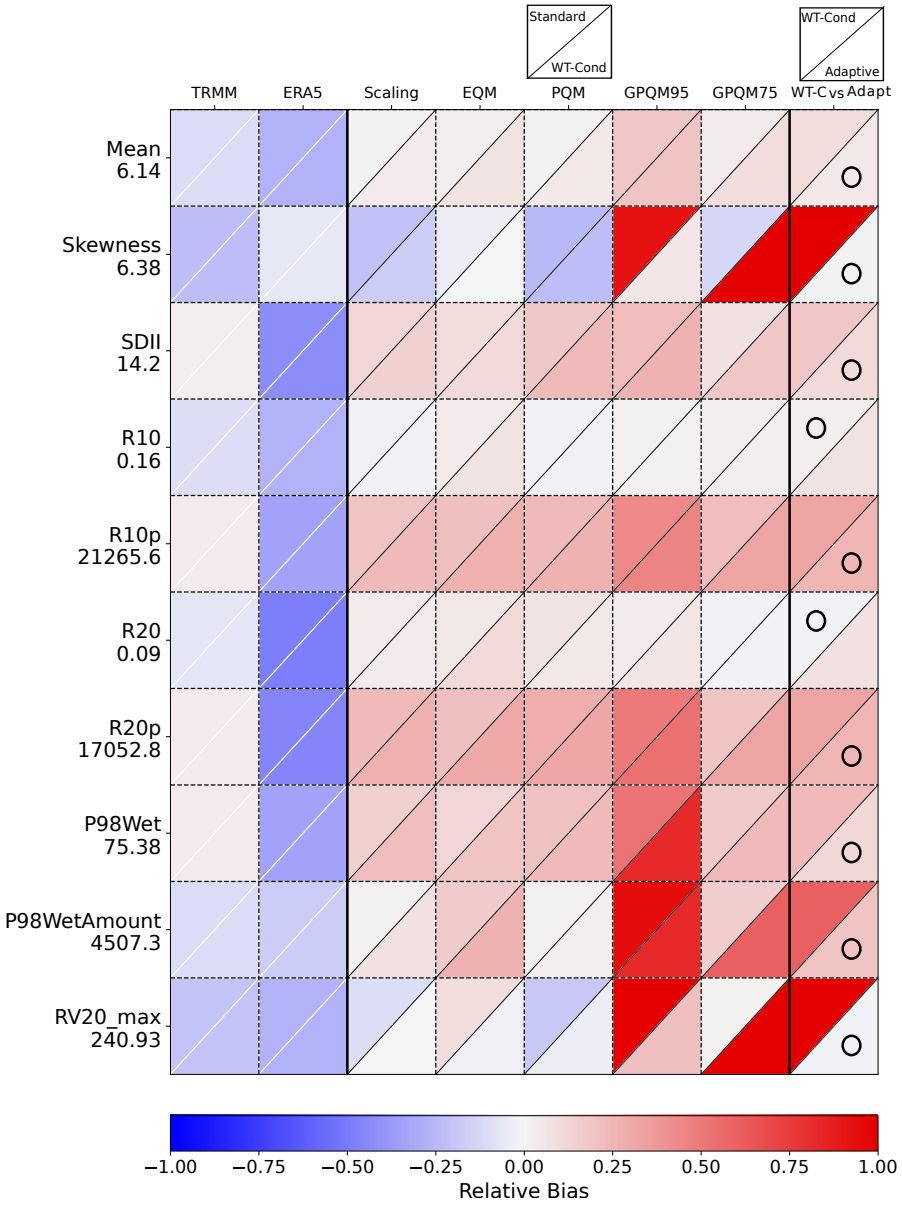

Figure A4. Same as Fig. A1, but for the Port Vila rain gauge location.

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

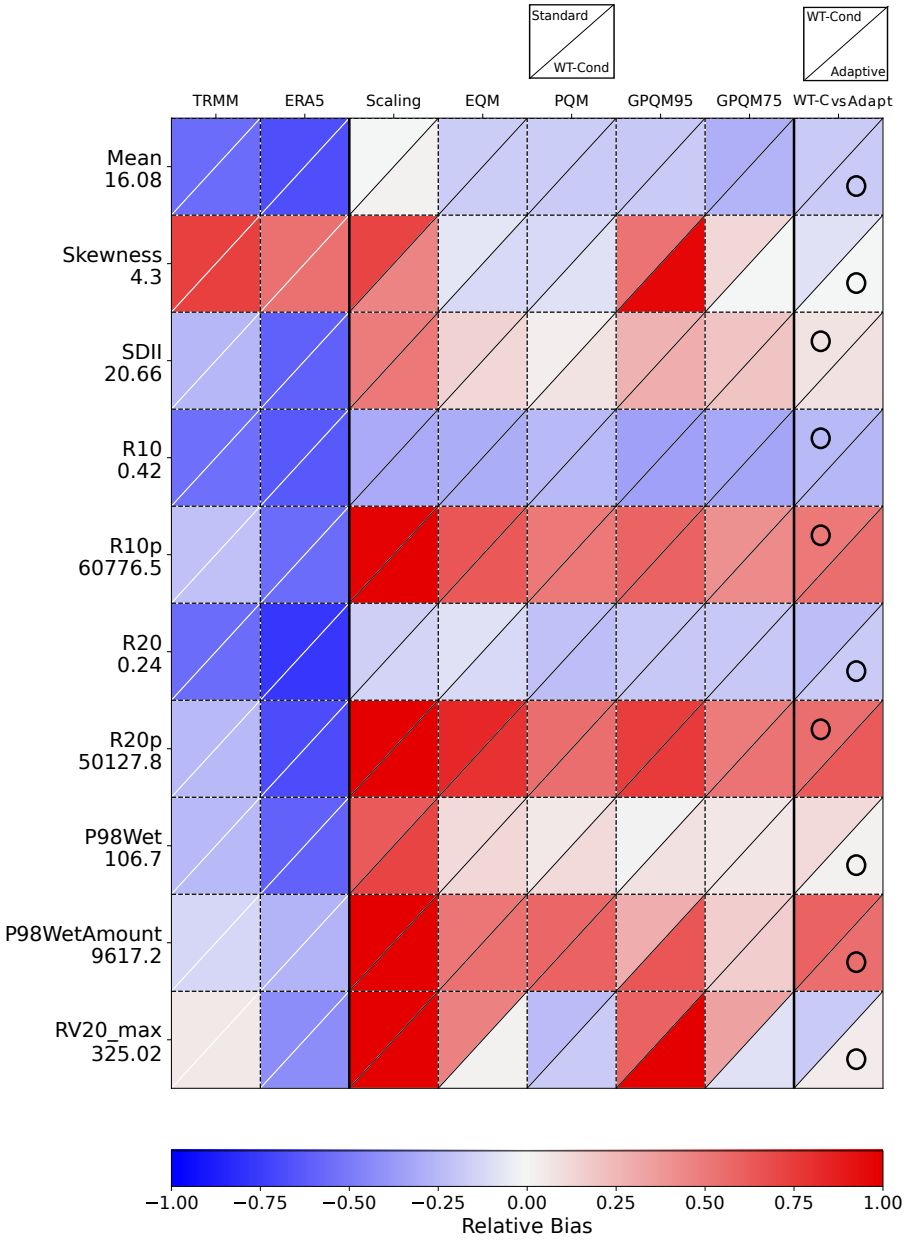

Figure A5. Same as Fig. A1, but for the Aoloau rain gauge location.

https://doi.org/10.5194/egusphere-2023-1402
Preprint. Discussion started: 21 July 2023
© Author(s) 2023. CC BY 4.0 License.

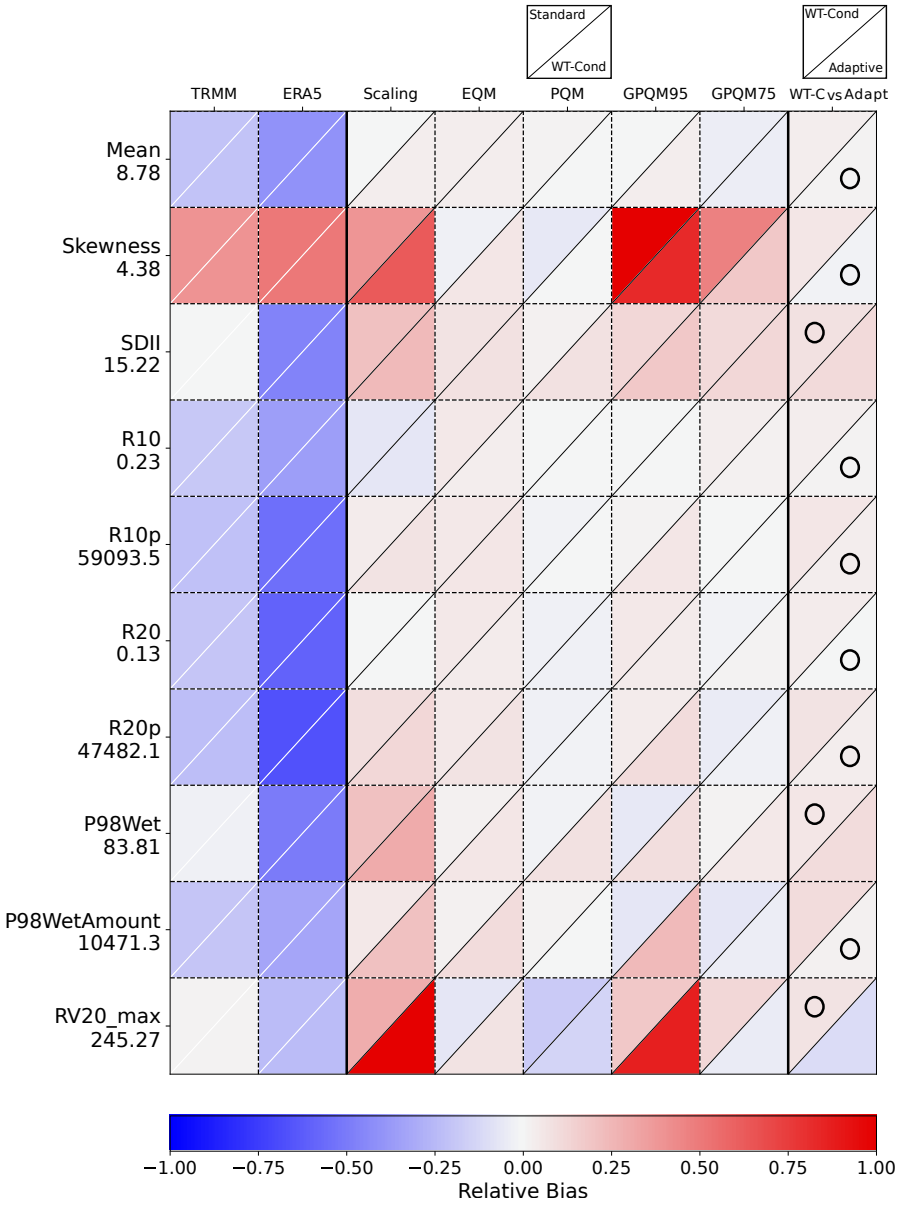

Figure A6. Same as Fig. A1, but for the Nuu'uli rain gauge location.