








RESEARCH ARTICLE

Weather-type-conditioned calibration of Tropical Rainfall Measuring Mission precipitation over the South Pacific Convergence Zone

Óscar Mirones¹  | Joaquín Bedia^{1,2}  | Juan A. Fernández-Granja³  |
Sixto Herrera¹  | Sara O. Van Vloten⁴  | Andrea Pozo⁴ | Laura Cagigal⁴  |
Fernando J. Méndez⁴ 

¹Departamento de Matemática Aplicada y Ciencias de la Computación, Universidad de Cantabria, Santander, Spain

²Grupo de Meteorología y Computación, Universidad de Cantabria, Unidad Asociada al CSIC, Santander, Spain

³Santander Meteorology Group, Institute of Physics of Cantabria IFCA, CSIC-UC, Santander, Spain

⁴Geomatics and Ocean Engineering Group, Departamento de Ciencias y Técnicas del Agua y del Medio Ambiente, Universidad de Cantabria, Santander, Spain

Correspondence

Óscar Mirones, Santander Meteorology Group, Department of Applied Mathematics and Computer Science, University of Cantabria, 39005 Santander, Spain.

Email: oscar.mirones@unican.es

Funding information

AFRICULTURES, Grant/Award Number: 774652; Beach4Cast, Grant/Award Number: PID2019-107053RB-I00; CORDyS, Grant/Award Number: PID2020-116595RB-I00; INDECIS, Grant/Award Number: 690462

Abstract

The South Pacific region is an area affected by characteristic precipitation patterns undergoing extreme events such as tropical cyclones and droughts. First, a daily weather typing of precipitation is presented, based on principal component analysis and k -means clustering using precipitation and atmospheric circulation variables derived from sea-level pressure and wind reanalysis fields. As a result, five weather types (WTs) are presented, able to capture distinct precipitation spatiotemporal patterns, interpretable in terms of salient regional climate features. Second, we undertake the calibration of the TRMM precipitation product using a set of rain gauge stations as reference and scaling and empirical quantile mapping (eQM) as calibration techniques. Furthermore, we build upon the weather-type classification to compare the results with a WT-conditioned calibration approach. Overall, our results underpin the need of adjusting the existing TRMM biases, mostly relevant for the upper tail of their distribution, and advocate the use of correction techniques able to deal with quantile-dependent biases—such as eQM—instead of a simple scaling, in order to obtain a more realistic representation of extreme precipitation events. The conditioning has shown only a marginal added value over the simple approach, although this minor improvement may prove relevant for applications focused on extreme event analysis. Furthermore, the weather types created can be applied to a wide variety of conditioned analyses in this region.

KEYWORDS

conditioned calibration, extreme precipitation, k -means clustering, principal component analysis, quantile mapping

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *International Journal of Climatology* published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society.

1 | INTRODUCTION

Natural disasters related to extreme hydrological events result every year in economic, human and structural losses to the vulnerable populations of the South Pacific Islands (Johnson *et al.*, 2021). Many of these hazardous events are related with erosion and flooding and have a compound origin, during which the adverse effects of

waves and tides occur simultaneously with extreme precipitation (“compound events,” see, e.g., Anderson *et al.*, 2019), most often associated with tropical cyclone occurrence. The South Pacific region in this study is located between the Equator and 30°S and 160°E and 150°W (Figure 1).

Here, the spatiotemporal characteristics of precipitation are driven by a number of processes operating at

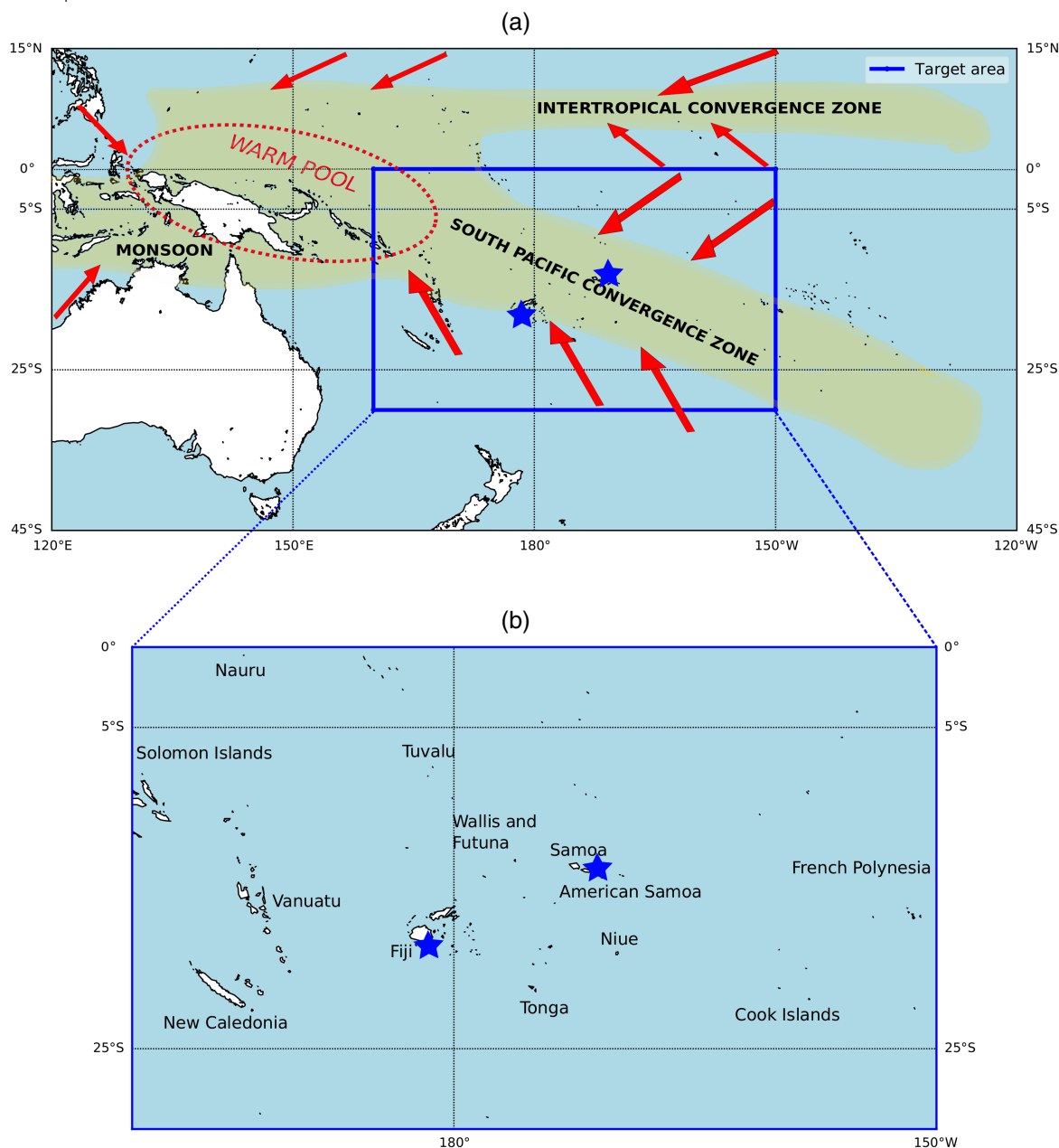


FIGURE 1 (a) Map of the South Pacific Ocean basin and (b) location of the study region, encompassing a rectangular domain between 0°N and –30°S and 160°E and 150°W. In (a), the arrows provide an approximated indication of the near-surface prevalent winds; the shading represents the bands of rainfall. The dashed ellipse locates the West Pacific Warm Pool (WPWP). Adapted from Australian Bureau of Meteorology and CSIRO (2011). In addition, the stars mark the locations of Suva (Fiji) and Apia (Samoa), used as reference for the calculation of the South Pacific Convergence Zone Index (SPI, Appendix A), whose SLP differences are envisaged as a quantitative synthetic descriptor of the SPCZ state (see, e.g., Salinger *et al.*, 2014) [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

multiple scales. For instance, the Intertropical Convergence Zone (ITCZ) and the South Pacific Convergence Zone (SPCZ) are prominent circulation patterns largely affecting precipitation in the area (Australian Bureau of Meteorology and CSIRO, 2011). These features form three extensive bands of large-scale wind convergence and associated rainfall (Figure 1a) leading in their confluence to the West Pacific Warm Pool (WPWP), partially overlapping in its northwestern part with the study area. The ITCZ is a band of precipitation located just in the north of the equatorial belt and is stronger between the months of June and August (Waliser and Gautier, 1993). In turn, the SPCZ is a band of high precipitation located between the Solomon Islands and the Cook Islands, strongest between December and February (Vincent, 1995) and exhibiting a significant temporal correlation with extreme precipitation events (Griffiths *et al.*, 2003). The western part of the SPCZ lies within the WPWP, determined by oceanic surface temperatures above 28°C (De Deckker, 2016), where precipitation is abundant and nearly constant throughout the year (Wyrtki, 1989). The Pacific Warm Pool provides an important source of moisture which serves as energy for the convective activity which drives the Walker circulation and a large part of the Hadley circulation (Wang, 2004). Similarly, the SPCZ position induces variability from interannual to multidecadal timescales (Salinger *et al.*, 2001). In addition, the West Pacific monsoon (WPM) stands out because its arrival marks a change from very dry to very wet conditions in the far west region. Furthermore, a major source of interannual variability is induced by El Niño–Southern Oscillation (ENSO; Trenberth, 1976), often described in terms of a displacement of the SPCZ diagonal axis to the northeast/southwest depending on positive(Niño)/negative(Niña) sea-surface temperature anomalies (Folland *et al.*, 2002). As a result, the SPCZ position is considered as a reliable indicator of atmospheric circulation changes in the South Pacific (Vincent *et al.*, 2011).

A better understanding of precipitation patterns in this vulnerable region is of utmost importance in order to develop adequate hydrological modelling and planning, allowing for the implementation of prevention plans and the design of adequate protection measures and infrastructures. To this aim, it is essential to have a historical record of reliable precipitation observations and the development of an adequate multiscalar characterization of rainfall variability (see, e.g., Pike and Lintner, 2020). The precipitation data sources to this aim include not only the local precipitation series from the rain gauge networks operating in the region (see, e.g., Greene *et al.*, 2008), but also the pseudo-observations derived

from different reanalysis products (Harvey *et al.*, 2019) and satellite measurement mission data (Huffman *et al.*, 2016). In contrast to rain gauge data, reanalysis and satellite estimates provide spatially continuous and homogeneous time series for an extended time period, as required for instance for running hydrological model simulation experiments at the catchment scale (e.g., Viviroli *et al.*, 2009; Usman *et al.*, 2022). Nevertheless, in the case of reanalysis data precipitation is not directly assimilated and largely depends on the numerical model associated to the assimilation system (Arakawa and Kitoh, 2004), which may hamper its direct application (see, e.g., Bedia *et al.*, 2012). In the same vein, satellite estimates offer a good alternative, but they can be affected by systematic biases, which can be problematic when applying the raw data in impact studies (see, e.g., Aghakouchak *et al.*, 2009), particularly when it comes to extreme event analysis (Sekaranom and Masunaga, 2019). As a result, it is often required a calibration of the data prior to their usage in impact modelling studies (Almazroui, 2011). In this study, we focus on the Tropical Rainfall Measuring Mission (TRMM; Huffman *et al.*, 2016), an observational dataset widely applied for hydrology studies in the Tropics.

In the framework of climate change and climate model calibration, several recent studies emphasize the importance of putting calibration in the context of relevant atmospheric processes influencing the target variables (see, e.g., Maraun *et al.*, 2017). Although the idea originally stems out from the potential pitfalls in climate model calibration, where the risk exists of artificially altering trends and the magnitude of the climate change signal, this approach may also help to deal with biased data in other applications (e.g., in seasonal forecasting; Manzanos and Gutiérrez, 2019). For instance, although biases are typically assumed to be time-independent, they may also vary in time, and this holds true also for remote sensing data of precipitation, for which large uncertainties exist regarding the estimation of precipitation amount from radar reflectivity measurements (Simpson *et al.*, 1996; Sekaranom and Masunaga, 2019) or infrequent satellite overpasses (Aghakouchak *et al.*, 2009). These biases are not constant but associated to particular meteorological situations, such as a systematic overestimation for wet periods (Almazroui, 2011). In particular, this work explores the potential of a reanalysis-based weather typing to improve the statistical adjustment of satellite series to adapt their statistical properties to those of in situ observations in the most reliable way possible. To this aim, prior to calibration we perform a weather typing of precipitation and other circulation variables using a clustering algorithm on reanalysis data.

Therefore, our approach includes synoptic-scale information from circulation fields together with their weather response (i.e., precipitation). As a result, we expect the weather types to be relevant for the description of local precipitation characteristics (Cannon, 2012). Using this technique, we extract the most representative spatiotemporal patterns, resulting in a classification of a few distinct synoptic atmospheric situations describing the main precipitation features. We also put this classification in the context of the most prominent elements of the regional rainfall climatology, affected by the varying position of the SPCZ and its influence on tropical cyclone tracks. To our knowledge, this is the first study addressing a weather-type-conditioned calibration of the TRMM precipitation product. “Conditioning” is here based on applying separate statistical corrections for each of the generated weather types (which act as data “blocks”) and then merging them together into a single calibrated time series. The main hypothesis supporting this approach is that biases might be dependent on specific atmospheric conditions which can be partially captured by weather types, thus adapting the correction factors to specific synoptic conditions. Even though sample size is reduced through conditioning, calibration with subsamples may result in a clear benefit improving the reliability of the corrected series (Reiter *et al.*, 2018). To this aim, the conditioned calibration approach is compared against an ordinary, nonconditioned calibration, considering a set of evaluation measures which take into account relevant distributional properties of the calibrated precipitation series, including extremes. We also assess the local adjustment undertaken through the analysis of their percentile adjustment functions (PAFs; Casanueva *et al.*, 2018), considering two popular bias adjustment techniques widely used by the climate impact communities, namely scaling and eQM.

Overall, our results underpin the need of adjusting the existing TRMM biases, mostly relevant for the upper tail of their distribution, and advocate the use of correction techniques able to deal with quantile-dependent biases (such as eQM) instead of a simple scaling, in order to obtain a more realistic representation of extreme precipitation events. The WTs generated offer a useful characterization of the spatiotemporal rainfall patterns building upon relevant synoptic-scale circulation fields, and proved useful for conditioned calibration. The improvements found with the conditioned calibration are marginal, and only relevant for weather types associated with extreme events, when tropical cyclone occurrence is concentrated. Furthermore, the weather typing may prove useful in other climate impact-relevant sectors beyond hydrological studies.

2 | DATA AND METHODS

2.1 | Reanalysis data

The weather types were constructed using the ERA5 reanalysis fields (Hersbach *et al.*, 2020). ERA5 provides hourly data for a number of atmospheric, land and oceanic variables since 1979 onwards at a 0.25° horizontal resolution grid for the entire globe. In particular, we retrieved the precipitation, mean sea-level pressure (SLP) and northward and eastward 10-m wind component fields for the period 1979–2019. The wind fields have been included for weather typing since it is more common to use wind as a circulation variable because of the inadequacy of the geostrophic approximation in the Tropics. We performed a daily mean aggregation (daily accumulated values for precipitation). Furthermore, we computed the SLP first-order time differences (SLP_{diff}) as the day-to-day difference of mean daily SLP at each grid cell. The SLP_{diff} is a relevant circulation variable which approximates the atmospheric circulation patterns for the characterization of cyclonic situations (<https://forecast.weather.gov/glossary.php?word=deepening>). The climatological maps of all input variables are included in Figure A1.

2.2 | Rain gauge data

The reference observations used as predictand for calibration were taken from the Pacific Rainfall Database (PACRAIN; Greene *et al.*, 2008). The PACRAIN Database consists of daily and monthly rainfall records from a comprehensive collection of rain gauge stations scattered across atolls and islands in the South Pacific region from different sources, such as 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). We extracted a subset of available locations within the study area (Figure 1), and from them, we retained a final set of stations compliant with the criteria of adequate temporal coverage and low number of missing data, ensuring their fitness-for-purpose for a robust calibration of the TRMM dataset. The final set of rain gauge stations is described in Table 1.

2.3 | TRMM data

In this study we focus on the calibration of the Tropical Rainfall Measuring Mission 3B42 Daily product (TRMM

TABLE 1 Final set of rain gauge stations from the PACRAIN Database used in this study

Station ID	Station name	Longitude	Latitude	Start	End	% missing data	Altitude
NZ75400	Kolopelu (Wallis and Futuna)	178.12°W	14.32°S	Jan 1, 1998	Jan 1, 2012	9.74	36
NZ82400	Alofi (Niue)	169.93°W	19.07°S	Jan 1, 1998*	Sep 2, 2010	2.68	59
NZ84317	Rarotonga (Cook Islands)	159.80°W	21.20°S	Sep 28, 1999	Jan 2, 2012	11.36	4
NZ99701	Raoul Island (New Zealand)	177.93°W	29.23°S	Jan 1, 1998*	Jan 1, 2012	0.72	49
SP00646	Port Vila (Vanuatu)	168.30°E	17.72°S	Jan 26, 2000	Jun 1, 2013	18.13	24
US14000	Aoloau (American Samoa)	170.77°W	14.30°S	Jan 1, 1998*	Dec 31, 2019*	21.72	408
US14690	Nu'uuli (American Samoa)	170.70°W	14.32°S	Jan 1, 1998*	Dec 31, 2019*	0.037	3

Note: The columns show the PACRAIN ID, which indicates the source from which the data are derived (NZ is related with NIWA, US with NCEI and SP with SPaRCE, see section 2.2), station name (and location), longitude and latitude coordinates in degrees, time coverage of the time series (start and end dates, the asterisk indicates that the original PACRAIN database contains data before/after the indicated calibration start/end period, outside the TRMM period, thus discarded in this study), percentage of missing data within the start–end period and elevation (meters above sea-level).

TPMA Precipitation L3 1 day $0.25^\circ \times 0.25^\circ$ V7; Huffman *et al.*, 2016, https://disc.gsfc.nasa.gov/datasets/TRMM_3B42_Daily_7/summary). This dataset provides daily accumulated precipitation with a temporal coverage from January 1, 1998 to January 1, 2020 with a daily resolution and a spatial coverage between 50.0°N to 50.0°S and 180.0°E to 180.0°W . The TRMM data were calibrated at each PACRAIN point, by extracting the nearest grid point to each rain gauge location (Table 1).

2.4 | Cyclone track data

In order to gain a closer insight into the tropical cyclone (TC) conditional probabilities of occurrence associated to each weather type, we retrieved the full TC track record within the study area from the International Best Track Archive for Climate Stewardship (IBTrACS v4.0 database; Knapp *et al.*, 2010), an open data transnational initiative to unify the TC data held by several climate agencies, storing a harmonized database of TCs from 1841 to present at a 0.1° spatial resolution and temporal resolutions ranging from 3-hourly to daily. In this work, we consider the position coordinates of TC tracks within the region for the whole analysis period 1998–2019, allowing for the calculation of TC frequency and their spatial distribution for selected time periods and/or weather types. All TC track data have been aggregated to a daily time resolution prior to TC frequency analyses.

2.5 | Weather-typing method

Principal component analysis (PCA) is extensively used in climate science (Preisendorfer and Mobley, 1988) to decompose space–time fields into a set of orthogonal spatial patterns (empirical orthogonal functions [EOFs]) and

their associated uncorrelated time indices (principal components [PCs]). The geometrical constraints characterizing EOFs and PCs can be very useful in practice since the covariance matrix of any subset of retained PCs is always diagonal and can drastically reduce the dimensionality of the data, coming at the cost of potential shortcomings in their interpretability (Hannachi *et al.*, 2007). In this work, a joint PCA of all five ERA5 daily-aggregated variables (precipitation, SLP, SLP_{diff} 10 m u and v wind components, section 2.1) was applied, in order to eliminate the linear dependence among the variables to be clustered; this may not happen if PCA is performed separately for each variable, which may result in unwanted data redundancy. After inspecting the explained variance by each PC, we kept all PCs explaining up to 80%, resulting in a 45 PC matrix for cluster analysis (further details on PCA procedure are provided in Figure A2). For illustration of the PCA results, a summary of the five first EOFs/PCs is displayed in Figure 2.

The k -means clustering method was then followed for obtaining representative patterns from the PCA analysis. This is a classical method for partitioning the feature space into a predefined number of clusters (k) based on an iterative search of group centroids aimed at the maximization of cluster distances while minimizing within-cluster dispersion (see, e.g., Vrac and Yiou, 2010, for an application to regional precipitation regimes). A similar classification approach was followed in previous studies in the South Pacific region for the analysis of satellite-derived precipitation (Pike and Lintner, 2020) or to investigate the structure and spatiotemporal variability of the SPCZ (Vincent *et al.*, 2011; Matthews, 2012), for instance. In order to select a suitable value for k , different models have been analysed for k values ranging from 4 to 8 (see, e.g., Pike and Lintner, 2020). The final choice of $k = 5$ was a compromise between representativity of each group (at least ~ 2 years of data in each group) and a

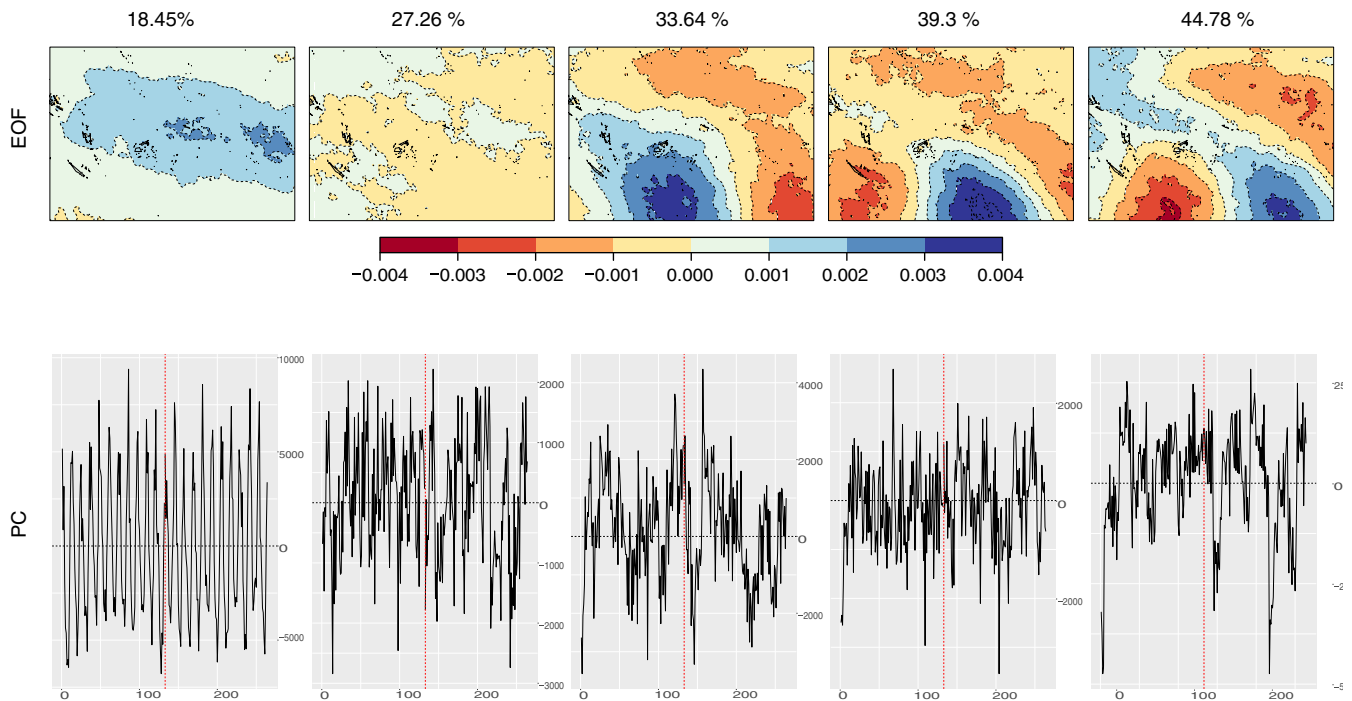


FIGURE 2 Upper row: first five EOFs and accumulated total explained variance (in %). Lower row: monthly aggregated first five principal components (PCs) [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/joc.7905)]

minimization of the total within-cluster variance (expressed as the total sum of squared distances to the centroid of each group instance). Further groups did not provide relevant reductions of within-group variance and yielded less representative WTs (less than 600 days for the 21-year period 1998–2019). Further detail is provided in Figure A3.

The robustness of the classification was assessed by partitioning the data in four temporal blocks (or *folds*, namely 1979–1988, 1989–1998, 1999–2008 and 2009–2019) following a fourfold procedure. As a result, PCA was trained on each fold and the resulting EOF was projected onto the remaining folds. This procedure was repeated four times, one for each training fold, yielding consistent results in terms of WT climatologies and precipitation seasonality (Figure 3). This analysis served as an evidence of the robustness of the weather typing, whose results were largely independent of the 10-year training period chosen. All the results presented hereafter are referred to the period 1998–2019, whose PCA has been performed using the 2009–2019 (the most recent period), thus greatly alleviating the computational cost of performing PCA on the entire 40-year period, without affecting the final WT classification. Furthermore, this period is consistent with the TRMM database availability used for the conditioned calibration study (section 2.6).

Principal component analysis and clustering have been undertaken using the functions available in the R

package *transformeR* of the *climate4R* open source framework for climate data analysis (Iturbide *et al.*, 2019; <https://github.com/SantanderMetGroup/climate4R>).

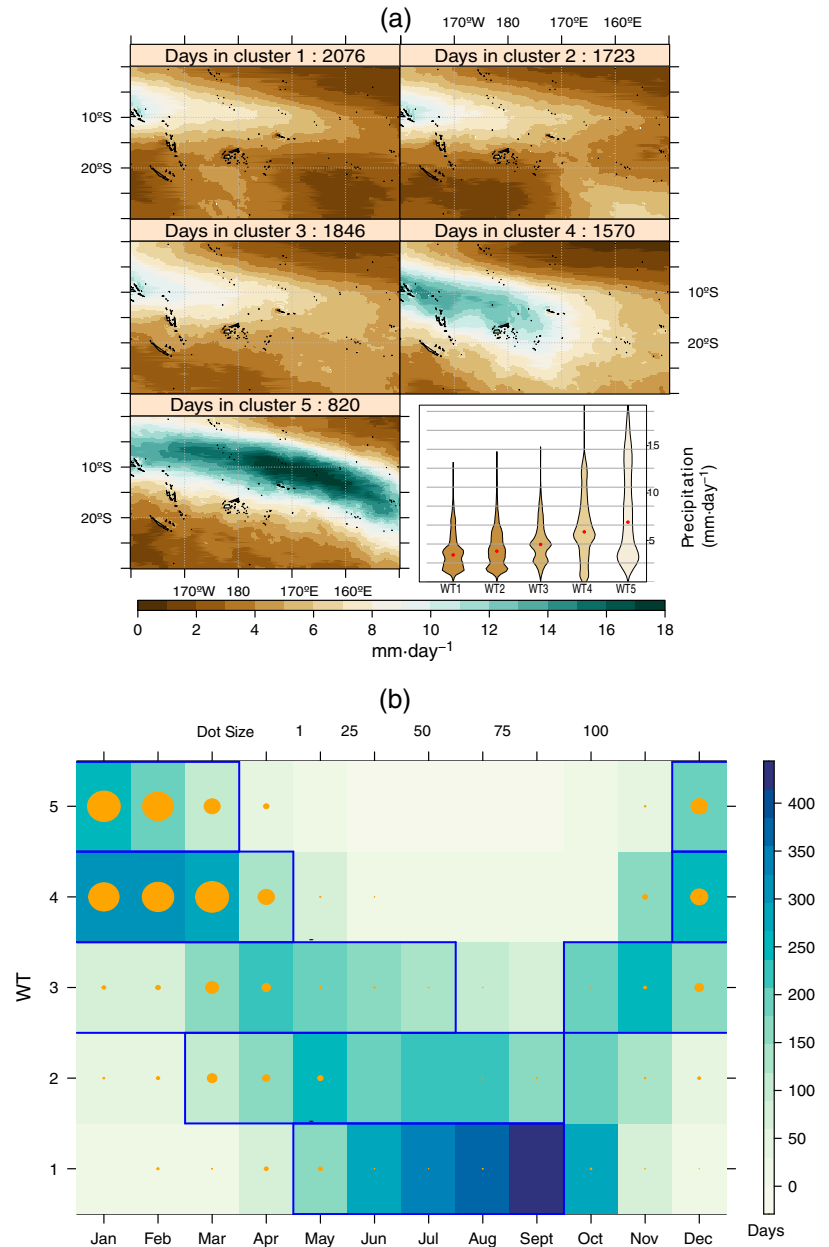
2.6 | Bias correction and WT-conditioned methodology

Here, we analyse two contrasting calibration techniques, namely empirical quantile mapping (eQM) and scaling, next described. A simple scaling of the data is the most common approach for TRMM calibration (e.g., Almazroui, 2011), being here used as a benchmarking choice. Scaling is performed on the TRMM raw data using a correction factor given by the quotient between the mean of the predictand (PACRAIN rain gauge measurements, p_{rg} ; Table 1) and the raw TRMM measurements p_{trmm} for a training period. As a result, the calibrated TRMM series \hat{p}_{trmm} are computed as

$$\hat{p}_{trmm} = p_{trmm} \frac{\bar{p}_{rg}}{\bar{p}_{trmm}}. \quad (1)$$

Our eQM implementation is the adaptation of Themeßl *et al.* (2011), based on using empirical cumulative distribution functions (ECDFs), calibrated on climatological distributions and with the predictor and the

FIGURE 3 (a) Precipitation climatology (1998–2019, $\text{mm}\cdot\text{day}^{-1}$) conditioned to each weather type (WT). The number of days falling in each WT is also indicated. The precipitation distribution for each WT is displayed in the violin plot, where the median is represented by the dots. (b) Number of days for each WT and month. The wet annual periods (calculated as the months encompassing 80% of total annual precipitation for that WT) are delimited by lines. The dot size is proportional to the total number of days affected by TCs (1998–2019) [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]



predictand the same parameters. We denote $\hat{X}_{t,i}$ as the corrected time series after applying quantile mapping on daily basis (t) and each grid cell (i). The resulting corrected $\hat{X}_{t,i}$ is computed as

$$\hat{X}_{t,i} = \text{ECDF}_{\text{doy},i}^{\text{rg,cal}^{-1}} \left(\text{ECDF}_{\text{doy},i}^{\text{trmm,cal}}(X_{t,i}) \right), \quad (2)$$

where $X_{t,i}$ is the uncorrected value for the corresponding day and grid cell, and $\text{ECDF}_{\text{doy},i}^{\text{trmm,cal}}$ and $\text{ECDF}_{\text{doy},i}^{\text{rg,cal}}$ are the ECDFs for TRMM and PACRAIN, respectively, corresponding to the given calibration period (cal) and day of the year (doy).

In the case of the weather-type-conditioned calibration, both techniques are applied separately for each WT,

being the final calibrated series the result of joining each independent calibration into a single time series encompassing the entire calibration period.

2.7 | Evaluation of calibrated series

2.7.1 | Cross-validation scheme

For a proper evaluation and intercomparison of different calibration methods (scaling and eQM) and choices (WT-conditioned/nonconditioned), we apply a cross-validation scheme for model fitting, in order to avoid spurious results due to artificial skill. The cross-validation method allows to properly evaluate whether the calibration

results are consistent outside the training period, in this case using a separate test period for prediction (Efron and Gong, 1983). Here, we consider a leave-one-year out validation setup, a variation of the classical k -fold cross-validation in which each of the k test folds is constructed with all data from a particular year, while the remaining years are used for training. The process is repeated k times ($k = 22$ years of the 1998–2019 period), yielding 22 different independent predictions which are joined together into a complete calibrated series. The same procedure was used in the case of conditioned calibration, in this case considering, for each year, the daily records belonging to each different WT separately (section 2.6).

2.7.2 | Evaluation indices and measures

The performance of the calibration, understood as the ability of the different calibration methods/approaches to bring the TRMM precipitation series to the PACRAIN observed reference in terms of their statistical properties, is measured as the relative bias between the raw and calibrated series, using to this aim a number of specific indices aimed at the characterization of mean and extreme properties of precipitation, consistent the validation framework proposed in the VALUE Framework for method intercomparison (Gutiérrez *et al.*, 2019). These indices are summarized in Table 2.

The calibration methods and the leave-one-year out scheme are implemented in the calibration and empirical statistical downscaling tools available in the R package *downscaleR* (Bedia *et al.*, 2020) of the *climate4R* framework. The different evaluation indices displayed in Table 2 have been computed with the standard definitions of the VALUE Framework (Maraun *et al.*, 2015) implemented in the R package *VALUE* (<https://github.com/SantanderMetGroup/VALUE>).

TABLE 2 Summary of the validation indices and measures used in this work

Code	Description	Type
SDII	Mean wet-day (≥ 1 mm) precipitation	Index
R10	Relative frequency of days with precip ≥ 10 mm	Index
P98Wet	98th percentile of wet (≥ 1 mm) days	Index
R98p_TOT	Quotient of total amount above 98th percentile of wet (≥ 1 mm) days and total precipitation	Index
Relative bias		Measure

Note: Their codes are consistent with the VALUE reference list (<http://www.value-cost.eu/validationportal/app/\#!indices>).

2.7.3 | Percentile adjustment functions

In order to gain an insight into the way the calibration is operating on the raw TRMM data series, we constructed the Percentile Adjustment Functions PAFs associated to each calibrated series. The PAF is envisaged as a diagnostic tool for eQM displaying the magnitude of the correction applied to each of the 99 percentiles of the training data (Casanueva *et al.*, 2018).

3 | RESULTS AND DISCUSSION

3.1 | Spatiotemporal characterization of precipitation by weather types

The resulting weather-type (WT) classification depicts five different states of activity of the South Pacific Convergence Zone (SPCZ) and its interaction with the West Pacific Warm Pool (WPWP; section 1), leading to distinct precipitation patterns over the SPCZ band: low precipitation states (WT1 and WT2), an intermediate type (WT3) and high activity SPCZ types, characterized by the highest rainfall amount and tropical cyclone frequency (WT4 and WT5; Figure 3a). As a result, the WTs describe the characteristic SPCZ seasonal cycle (Figure 3b), related to the salient regional climate features, such as droughts and floods. Further details on SPCZ activity and tropical cyclone affection conditioned to each WT occurrence is given in Appendix (Figure A3). All WTs (particularly WT4 and WT5) exhibit the characteristic SPCZ band, extending from SE to NW of the study area. Therefore, there is a west–east-oriented part interacting with the WPWP, and a diagonally oriented part extending into the subtropics (Vincent, 1995), driven by the convergence between the northeasterly trade winds and the southeasterly circulation ahead of the Australian anticyclones (Trenberth, 1976). While the western zone is in contact with the WPWP, the eastern part is characterized by the interactions with the troughs of the mid-latitude circulation, leading to short-term variability of regional climate features (Kiladis *et al.*, 1989; Vincent, 1995). During the austral winter, the WPWP displaces eastwards and the SPCZ retracts (as captured by WTs 1 and 2); in turn, during summer the SPCZ expands southeastwards and eastern WPWP retreats westwards (Linsley *et al.*, 2008, reflected in WTs 4 and 5). As a result, austral summer is characterized by a more intense rainfall and tropical cyclone (TC) activity (captured by WTs 4 and 5; Figure 3), triggered by higher sea-surface temperatures (SST) and deep convection (Takahashi and Battisti, 2007). On the other hand, during the dry season the SPCZ weakens and the spatial pattern of precipitation is

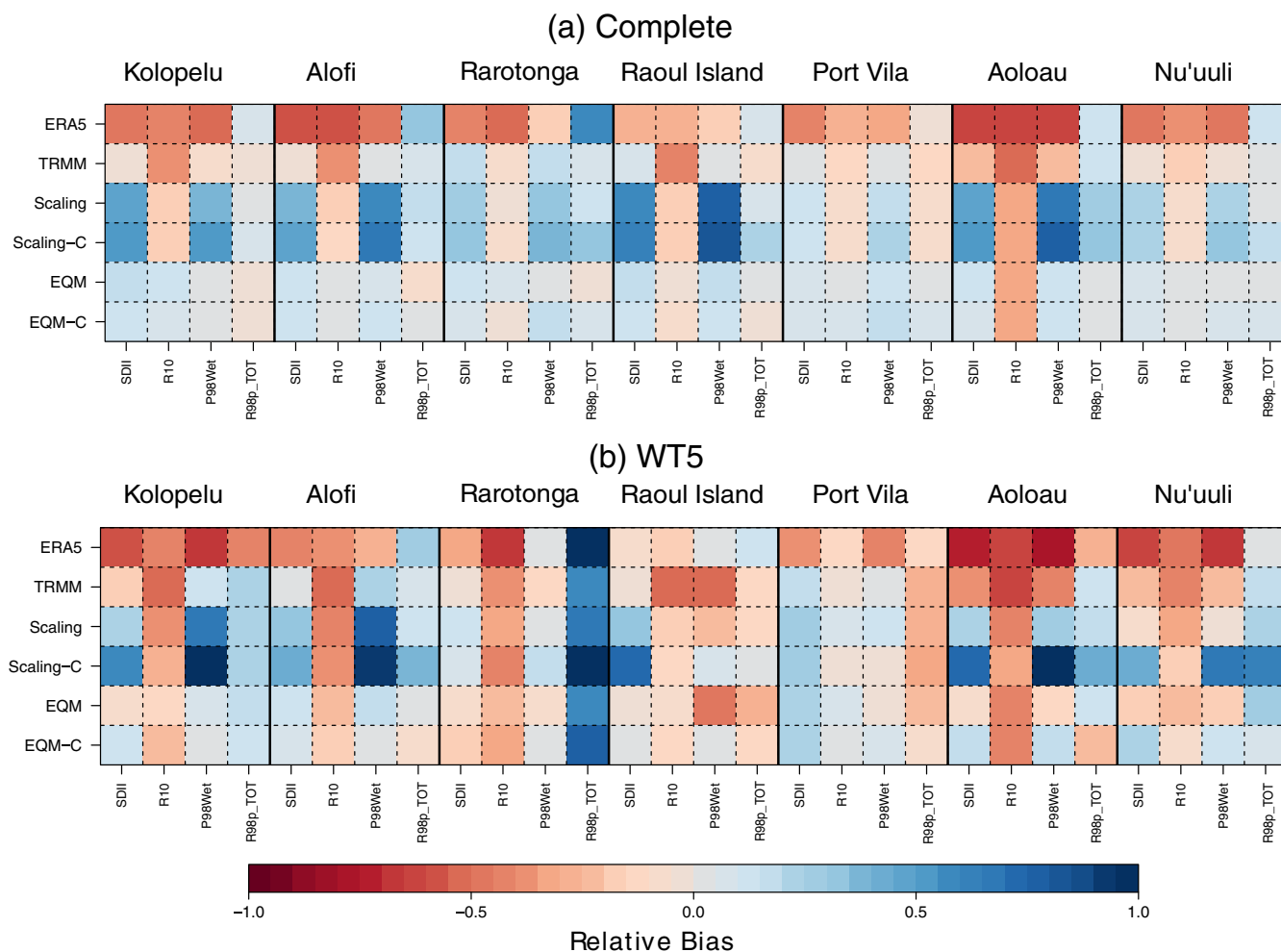


FIGURE 4 Relative bias of (raw) precipitation products (ERA5, TRMM) and calibrated TRMM w.r.t. the reference PACRAIN rain gauge observations, for the seven locations indicated in Table 1. The calibration techniques are eQM, scaling, and their respective WT-conditioned counterparts (indicated with the suffix -C). The different validation indices (Table 2) are arranged in columns, and their relative bias indicated by the colorbar (reds for negative, blue for positive). In (a), the results are presented for the complete 1998–2019 period. The lower panel (b) shows the same information but extracting all data values belonging to WT5 only [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/terms-and-conditions)]

more homogeneous across the region, reaching the dry peak between June and August (Vincent, 1995), when precipitation is mostly restricted to the northwestern part, over the Solomon Islands, while Fiji, Tonga and Cook Islands experience dry conditions as they are located in the widening zone of the SPCZ (Figure 1), weak and sparse at this point (WTs 1–3; Figure 3). The characteristic “diagonal” shape of the SPCZ band is the result of the Pacific zonal SST gradient forcing and the accumulation of wave energy over the central South Pacific (Widlansky *et al.*, 2011), as highlighted by the spatial pattern of WT5. This pattern gradually weakens in WT4 and WT3 and it is lost in WTs 1 and 2, corresponding to quiescent SPCZ conditions. In the same vein, TC activity goes in hand with the SPCZ state and concentrates mostly in WTs 4 and 5 (Figure A4).

In order to gain a deeper insight into the temporal component of each WT, we characterize their annual

cycles of precipitation. To this aim, we calculated the characteristic “wet” season of each WT, defined as the continuous monthly interval in which at least 80% of precipitation occurs, relative to the mean total precipitation for that particular WT (Figure 3b). The results underpin the adequate feature separation performed by the clustering algorithm also on the temporal aspect, able to consistently capture the precipitation seasonality.

3.2 | Bias correction results

The intercomparison of precipitation datasets reveals an overall low bias of the TRMM product considering the complete series 1998–2019 (Figure 4a), exhibiting a few moderate negative biases (−0.5 or lower) in the representation of the R10 index for some of the sites (Kolopelu, Alofi, Raoul Island and Aoloau). Even though the biases

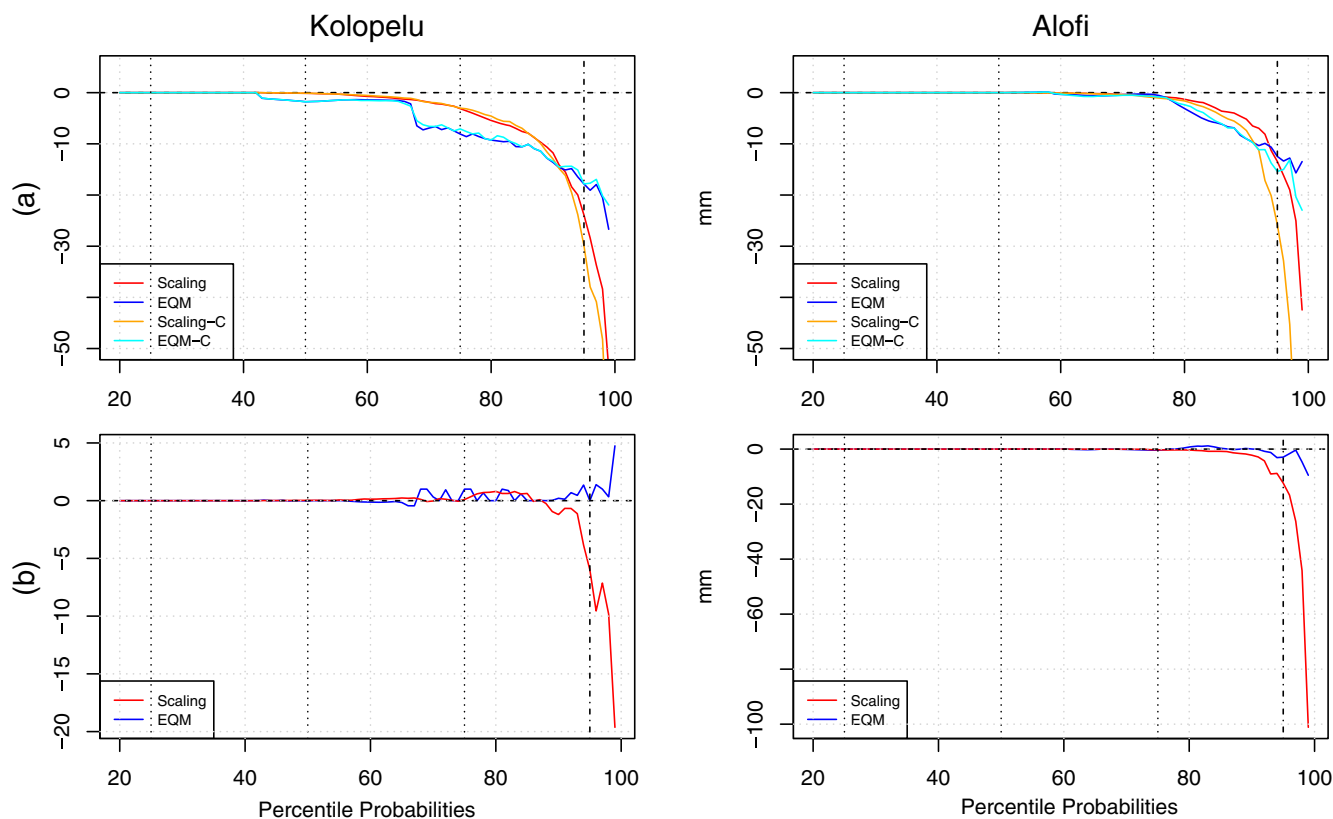


FIGURE 5 (a) Percentile adjustment functions (PAFs) for scaling, eQM and their respective conditioned calibration versions, considering the last 79 percentiles of Kolopelu (left) and Alofi (right) rain gauge stations. (b) Percentile adjustment differences between nonconditioned and conditioned calibration for scaling and eQM. The dotted vertical lines represent the first quartile (Q1), the median and the third quartile (Q3); the dot-dashed vertical line represents the 95th percentile. In all cases, the y-axis units are mm, as a measure of the magnitude of the correction applied in each corresponding percentile of x-axis [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/joc.7905)]

found are low for the complete period, these significantly increase when considering high rainfall situations, such as those encompassed by WT5 (Figure 4b). In all cases, the biases found for TRMM precipitation are much lower than those from ERA5, which systematically under-represents all indices analysed, excepting the indicator of torrentiality R98p_TOT, which is slightly overestimated in most locations, notably in Rarotonga (Figure 4a,b). This preliminary analysis allows to conclude that TRMM is by far a preferable precipitation database than ERA5 for application in impact studies in the region. However, TRMM is not free from biases, particularly under the very wet conditions predominant in the cyclonic-prone, active SPCZ conditions represented by WTs 4 and 5.

Scaling is the most usual and straightforward technique for TRMM correction (see, e.g., Almazroui, 2011). However, our results indicate that overall, scaling was not useful to alleviate biases, and even had a deleterious effect on most indices analysed, with the exception of R10, for which a slight improvement is attained after calibration at some locations (Figure 4a,b). Similar results are obtained by conditioned scaling, which was not able to satisfactorily adjust the TRMM bias neither considering the complete period (Figure 4a), nor considering the

WT5 alone (Figure 4b). While in theory scaling is a good choice for correcting the mean and frequency of precipitation, our results, produced under a leave-one-year-out calibration scheme, reveal that the method has some drawbacks for its robust application to daily data.

Unlike scaling, eQM was able to effectively correct most biases. Only R10 on Aolou, exhibiting a large negative bias, was not totally corrected by quantile mapping, not even after WT conditioning (Figure 4a). The same holds true for the large overestimation of R98p_TOT by TRMM in the case of the WT5 training set at Rarotonga, which eQM failed to correct. In all cases, eQM and eQM-C attained consistent correction results and proved a preferable choice than scaling for calibration. However, the benefits of WT conditioning are marginal, and only noticeable when looking at more extreme WTs, such as WT5, and particular locations and indices (see Figure A5 for further details on the validation results for each WT individually).

The analysis of the percentile adjustment functions (PAFs) allows to further investigate how the calibration is operating. At a first glance, the advantages of eQM over scaling seem to lie in its ability to apply a different correction for different percentiles of the distribution

(Figure 5). While this has a negligible effect on most of the distribution, it makes a difference when it comes to correct the higher percentiles (i.e., extreme precipitation events), which are mostly responsible for the overall better performance of eQM over scaling. For higher percentiles, the magnitude of the correction is significantly different for eQM and scaling (Figure 5a). This may have an important effect on the representation of extreme events by the calibrated series, such as the case of P98Wet index in the Kolopelu site (Figure 4), in which eQM is able to reduce the (already low) TRMM bias while scaling adds more error after its application. In the same vein, PAFs unveil minor differences between the conditioned/unconditioned variations of scaling and eQM methods until the upper tail of the distribution, where the magnitude of the correction largely differs (Figure 5b). In the case of eQM, the conditioning has shown just a marginal added value over the simple approach, although this minor improvement may prove relevant for applications sensitive to extreme events.

Therefore, we have found the largest TRMM biases in the upper tail of the distribution, affecting extreme rainfall events mainly. These results suggest further exploration of other parametric, nonstandard techniques beyond the classical eQM, such as parametric quantile mapping (pQM; Piani *et al.*, 2010) or generalized-Pareto quantile mapping (GPQM; Gutjahr and Heinemann, 2013), the latter tailored to better deal with extremes using a generalized Pareto distribution above a user-defined percentile, or alternative calibration techniques based on a specific adjustment for extreme values (e.g., Tye *et al.*, 2014).

4 | CONCLUSIONS

A weather typing analysis of precipitation has been undertaken for a South Pacific region under the influence of the South Pacific Convergence Zone (SPCZ). To this aim, precipitation, sea-level pressure, daily sea-level pressure difference and 10 m northward and eastward wind field components have been considered after PCA transformation and *k*-means clustering. We show that the resulting WTs are able to consistently describe the main spatiotemporal patterns of precipitation, including extreme events related to tropical cyclone activity, and pose a straightforward interpretation in terms of salient regional climate features such as the SPCZ seasonal displacement. As a result, the daily classification produced can serve as a basis for further process-based oriented hydrological studies in the region.

To illustrate their potential application, we investigate the calibration of the TRMM precipitation product using

two commonplace bias adjustment techniques, namely scaling and empirical quantile mapping (eQM), and comparatively test their performance when their application is conditioned to the weather types. Overall, our results indicate that TRMM exhibits some biases, particularly in the upper tail of the distribution, that the scaling is not able to properly handle, but eQM can significantly reduce or eliminate. The analysis of the percentile adjustment functions has helped to gain an insight into how the calibration operates, revealing minor differences between the conditioned/unconditioned variations of scaling and eQM methods excepting for the higher percentiles, where the correction factors largely differ among methods. Therefore, the WT-conditioning has shown only a marginal added value over the unconditioned approach, although this minor improvement can make a difference for a better correction of extreme rainfall events, critical in many impact studies. In addition, this WT-conditioning approach can be extended to compound extreme events, in which several parameters are involved (e.g., precipitation, sea level, wind, etc.), in order to preserve the consistency between the different variables.

An associated dataset to this study is available in Zenodo (<https://doi.org/10.5281/zenodo.6940093>), containing the input data at the target locations in Table 1, and the raw ERA5, TRMM and calibrated precipitation series, as well as the corresponding daily weather-type classification.

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. Óscar Mirones and Sixto Herrera acknowledge partial funding from AFRICULTURES (H2020-SFS-2017-1) (Grant No. 774652). Joaquín Bedia acknowledges partial funding from the project INDECIS, part of the European Research Area for Climate Services Consortium (ERA4CS) with co-funding by the European Union (Grant No. 690462). Sara O. Van Vloten, Andrea Pozo, Laura Cagigal and Fernando J. Méndez acknowledge Beach4Cast: project Beach4cast PID2019-107053RB-I00, Spanish Ministry of Science and Innovation. We thank two anonymous referees for their insightful comments which have helped us to significantly improve the initial manuscript.

ORCID

Óscar Mirones  <https://orcid.org/0000-0001-7446-9440>

Joaquín Bedia  <https://orcid.org/0000-0001-6219-4312>

Juan A. Fernández-Granja  <https://orcid.org/0000-0002-9763-7811>

Sixto Herrera  <https://orcid.org/0000-0002-5384-179X>

Sara O. Van Vloten  <https://orcid.org/0000-0001-7832-955X>

Laura Cagigal  <https://orcid.org/0000-0001-5384-6382>

Fernando J. Méndez  <https://orcid.org/0000-0002-5005-1100>

REFERENCES

- Aghakouchak, A., Nasrollahi, N. and Habib, E. (2009) Accounting for uncertainties of the TRMM satellite estimates. *Remote Sensing*, 1, 606–619.
- Almazroui, M. (2011) Calibration of TRMM rainfall climatology over Saudi Arabia during 1998–2009. *Atmospheric Research*, 99, 400–414.
- Anderson, D., Rueda, A., Cagigal, L., Antolinez, J.A.A., Mendez, F. J. and Ruggiero, P. (2019) Time-varying emulator for short and long-term analysis of coastal flood hazard potential. *Journal of Geophysical Research: Oceans*, 124, 9209–9234.
- Arakawa, O. and Kitoh, A. (2004) Comparison of local precipitation–SST relationship between the observation and a reanalysis dataset. *Geophysical Research Letters*, 31, L12206.
- Australian Bureau of Meteorology and CSIRO. (2011) *Climate Change in the Pacific: Scientific Assessment and New Research. Volume 1: Regional Overview*. Aspendale, Vic: Pacific Climate Change Science Program. Available at: <https://www.pacificclimatechangescience.org/publications/reports/report-climate-change-in-the-pacific-scientific-assessment-and-new-research>.
- Bedia, J., Baño-Medina, J., Legasa, M.N., Iturbide, M., Manzananas, R., Herrera, S., Casanueva, A., San-Martín, D., Cofiño, A.S. and Gutiérrez, J.M. (2020) Statistical downscaling with the downscaleR package (v3.1.0): contribution to the VALUE intercomparison experiment. *Geoscientific Model Development*, 13, 1711–1735.
- Bedia, J., Herrera, S., Gutiérrez, J.M., Zavala, G., Urbieto, I.R. and Moreno, J.M. (2012) Sensitivity of fire weather index to different reanalysis products in the Iberian Peninsula. *Natural Hazards and Earth System Sciences*, 12, 699–708.
- Cannon, A.J. (2012) Regression-guided clustering: a semisupervised method for circulation-to-environment synoptic classification. *Journal of Applied Meteorology and Climatology*, 51, 185–190.
- Casanueva, A., Bedia, J., Herrera, S., Fernández, J. and Gutiérrez, J. M. (2018) Direct and component-wise bias correction of multi-variate climate indices: the percentile adjustment function diagnostic tool. *Climatic Change*, 147, 411–425.
- De Deckker, P. (2016) The Indo-Pacific Warm Pool: critical to world oceanography and world climate. *Geoscience Letters*, 3, 20.
- Deo, A., Chand, S., Ramsay, H., Holbrook, N., Mcgree, S., Magee, A., Bell, S., Titimaea, M., Haruhiru, A., Malsale, P., Mulitalo, S., Daphne, A., Prakesh, B., Vainikolo, V. and Koshiha, S. (2021) Tropical cyclone contribution to extreme rainfall over Southwest Pacific Island nations. *Climate Dynamics*, 56, 3967–3993.
- Efron, B. and Gong, G. (1983) A leisurely look at the bootstrap, the jackknife, and cross-validation. *American Statistician*, 37, 36–48.
- Folland, C.K., Renwick, J.A., Salinger, M.J. and Mullan, A.B. (2002) Relative influences of the Interdecadal Pacific Oscillation and ENSO on the South Pacific Convergence Zone. *Geophysical Research Letters*, 29, 1–4.
- Greene, J.S., Klatt, M., Morrissey, M. and Postawko, S. (2008) The comprehensive Pacific rainfall database. *Journal of Atmospheric and Oceanic Technology*, 25, 71–82.
- Griffiths, G.M., Salinger, M.J. and Leleu, I. (2003) Trends in extreme daily rainfall across the South Pacific and relationship to the South Pacific Convergence Zone. *International Journal of Climatology*, 23, 847–869.
- Gutiérrez, J.M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., Roessler, O., Wibig, J., Wilcke, R., Kotlarski, S., San Martín, D., Herrera, S., Bedia, J., Casanueva, A., Manzananas, R., Iturbide, M., Vrac, M., Dubrovsky, M., Ribalaygua, J., Pórtoles, J., Rätty, O., Räisänen, J., Hingray, B., Raynaud, D., Casado, M.J., Ramos, P., Zerenner, T., Turco, M., Bosshard, T., Štěpánek, P., Bartholy, J., Pongracz, R., Keller, D. E., Fischer, A.M., Cardoso, R.M., Soares, P.M.M., Czernecki, B. and Pagé, C. (2019) An intercomparison of a large ensemble of statistical downscaling methods over Europe: results from the VALUE perfect predictor cross-validation experiment. *International Journal of Climatology*, 39, 3750–3785.
- Gutjahr, O. and Heinemann, G. (2013) 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.
- Hannachi, A., Jolliffe, I.T. and Stephenson, D.B. (2007) Empirical orthogonal functions and related techniques in atmospheric science: a review. *International Journal of Climatology*, 27, 1119–1152.
- Harvey, T., Renwick, J.A., Lorrey, A.M. and Ngari, A. (2019) The representation of the South Pacific Convergence Zone in the twentieth century reanalysis. *Monthly Weather Review*, 147, 841–851.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. and Thépaut, J. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049.
- Huffman, G., Bolvin, D., Nelkin, E. and Adler, R. (2016) In: Savtchenko, A. (Ed.) *TRMM (TMPA) Precipitation L3 1 day 0.25° × 0.25° V7*. Goddard Earth Sciences Data and Information Services Center (GES DISC). Available at: https://disc.gsfc.nasa.gov/datasets/TRMM_3B42_Daily_7/summary.
- Iturbide, M., Bedia, J., Herrera, S., Baño-Medina, J., Fernández, J., Frías, M., Manzananas, R., San-Martín, D., Cimadevilla, E., Cofiño, A. and Gutiérrez, J. (2019) The R-based climate4R open framework for reproducible climate data access and post-processing. *Environmental Modelling & Software*, 111, 42–54.
- Johnson, F., Higgins, P. and Stephens, C. (2021) Climate change and hydrological risk in the Pacific: a humanitarian engineering perspective. *Journal of Water and Climate Change*, 12, 647–678.

- Kiladis, G.N., von Storch, H. and Loon, H. (1989) Origin of the South Pacific Convergence Zone. *Journal of Climate*, 2, 1185–1195.
- Knapp, K.R., Kruk, M.C., Levinson, D.H., Diamond, H.J. and Neumann, C.J. (2010) The international best track archive for climate stewardship (ibtracs): unifying tropical cyclone data. *Bulletin of the American Meteorological Society*, 91, 363–376.
- Linsley, B.K., Zhang, P., Kaplan, A., Howe, S.S. and Wellington, G. M. (2008) Interdecadal-decadal climate variability from multi-coral oxygen isotope records in the South Pacific Convergence Zone region since 1650 A.D.: Pacific IPO-PDO variability. *Paleoceanography*, 23, PA2219.
- Manzanas, R. and Gutiérrez, J.M. (2019) Process-conditioned bias correction for seasonal forecasting: a case-study with ENSO in Peru. *Climate Dynamics*, 52, 1673–1683.
- 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. (2017) Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, 7, 764–773.
- Maraun, D., Widmann, M., Gutiérrez, J.M., Kotlarski, S., Chandler, R.E., Hertig, E., Wibig, J., Huth, R. and Wilcke, R.A. (2015) VALUE: a framework to validate downscaling approaches for climate change studies. *Earth's Future*, 3, 1–14.
- Matthews, A.J. (2012) A multiscale framework for the origin and variability of the South Pacific Convergence Zone. *Quarterly Journal of the Royal Meteorological Society*, 138, 1165–1178.
- Piani, C., Haerter, J.O. and Coppola, E. (2010) Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and Applied Climatology*, 99, 187–192.
- Pike, M. and Lintner, B.R. (2020) Application of clustering algorithms to TRMM precipitation over the tropical and South Pacific Ocean. *Journal of Climate*, 33, 5767–5785.
- Preisendorfer, R. and Mobley, C. (1988) *Principal Component Analysis in Meteorology and Oceanography*. *Developments in Atmospheric Science*. Amsterdam: Elsevier.
- Reiter, P., Gutjahr, O., Schefczyk, L., Heinemann, G. and Casper, M. (2018) 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.
- Salinger, M.J., McGree, S., Beucher, F., Power, S.B. and Delage, F. (2014) A new index for variations in the position of the South Pacific Convergence Zone 1910/11–2011/2012. *Climate Dynamics*, 43, 881–892.
- Salinger, M.J., Renwick, J.A. and Mullan, A.B. (2001) Interdecadal Pacific Oscillation and South Pacific climate. *International Journal of Climatology*, 21, 1705–1721.
- Sekaranom, A.B. and Masunaga, H. (2019) 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.
- Simpson, J., Kummerow, C., Tao, W. and Adler, R.F. (1996) On the Tropical Rainfall Measuring Mission (TRMM). *Meteorology and Atmospheric Physics*, 60, 19–36.
- Takahashi, K. and Battisti, D.S. (2007) Processes controlling the mean tropical Pacific precipitation pattern. Part II: the SPCZ and the Southeast Pacific dry zone. *Journal of Climate*, 20, 5696–5706.
- Taylor, R.C. (1973) *An Atlas of Pacific Islands Rainfall*. Manoa, HI: Department of Meteorology, University of Hawaii at Manoa. Technical report: 25.
- Themeßl, J., Gobiet, A. and Leuprecht, A. (2011) Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *International Journal of Climatology*, 31, 1530–1544.
- Trenberth, K.E. (1976) Spatial and temporal variations of the Southern Oscillation. *Quarterly Journal of the Royal Meteorological Society*, 102, 639–653.
- Tye, M.R., Stephenson, D.B., Holland, G.J. and Katz, R.W. (2014) A Weibull approach for improving climate model projections of tropical cyclone wind-speed distributions. *Journal of Climate*, 27, 6119–6133.
- Usman, M., Ndehedehe, C.E., Ahmad, B., Manzanas, R. and Adeyeri, O.E. (2022) Modeling streamflow using multiple precipitation products in a topographically complex catchment. *Modeling Earth Systems and Environment*, 8, 1875–1885.
- Vincent, D.G. (1995) The South Pacific Convergence Zone (SPCZ): a review. *Oceanographic Literature Review*, 42, 1949–1970.
- Vincent, E.M., Lengaigne, M., Menkes, C.E., Jourdain, N.C., Marchesiello, P. and Madec, G. (2011) Interannual variability of the South Pacific Convergence Zone and implications for tropical cyclone genesis. *Climate Dynamics*, 36, 1881–1896.
- Viviroli, D., Mittelbach, H., Gurtz, J. and Weingartner, R. (2009) Continuous simulation for flood estimation in ungauged meso-scale catchments of Switzerland—part II: parameter regionalisation and flood estimation results. *Journal of Hydrology*, 377, 208–225.
- Vrac, M. and Yiou, P. (2010) Weather regimes designed for local precipitation modeling: application to the Mediterranean basin. *Journal of Geophysical Research: Atmospheres*, 115, D12103.
- Waliser, D.E. and Gautier, C. (1993) A satellite-derived climatology of the ITCZ. *Journal of Climate*, 6, 2162–2174.
- Wang, C. (2004) ENSO, Atlantic climate variability, and the Walker and Hadley circulations. In: *The Hadley Circulation: Present, Past and Future*. *Advances in Global Change Research*, vol 21. Wangs: Springer, pp. 173–202.
- Widlansky, M.J., Webster, P.J. and Hoyos, C.D. (2011) On the location and orientation of the South Pacific Convergence Zone. *Climate Dynamics*, 36, 561–578.
- Wyrтки, K. (1989) Some thoughts about the West Pacific Warm Pool. In: *Picaut, Lukas & Delcroix (Eds.), Proceedings of the Western Pacific International Meeting and Workshop on TOGA COARE*, 24–30 May 1989 ORSTOM Nouméa, pp. 99–109.

How to cite this article: Mirones, Ó., Bedia, J., Fernández-Granja, J. A., Herrera, S., Van Vloten, S. O., Pozo, A., Cagigal, L., & Méndez, F. J. (2023). Weather-type-conditioned calibration of Tropical Rainfall Measuring Mission precipitation over the South Pacific Convergence Zone. *International Journal of Climatology*, 43(2), 1193–1210. <https://doi.org/10.1002/joc.7905>

APPENDIX A: CLUSTERING METHODS A

In this section, information is provided regarding the clustering methodology, summarized in Figures A1–A3.

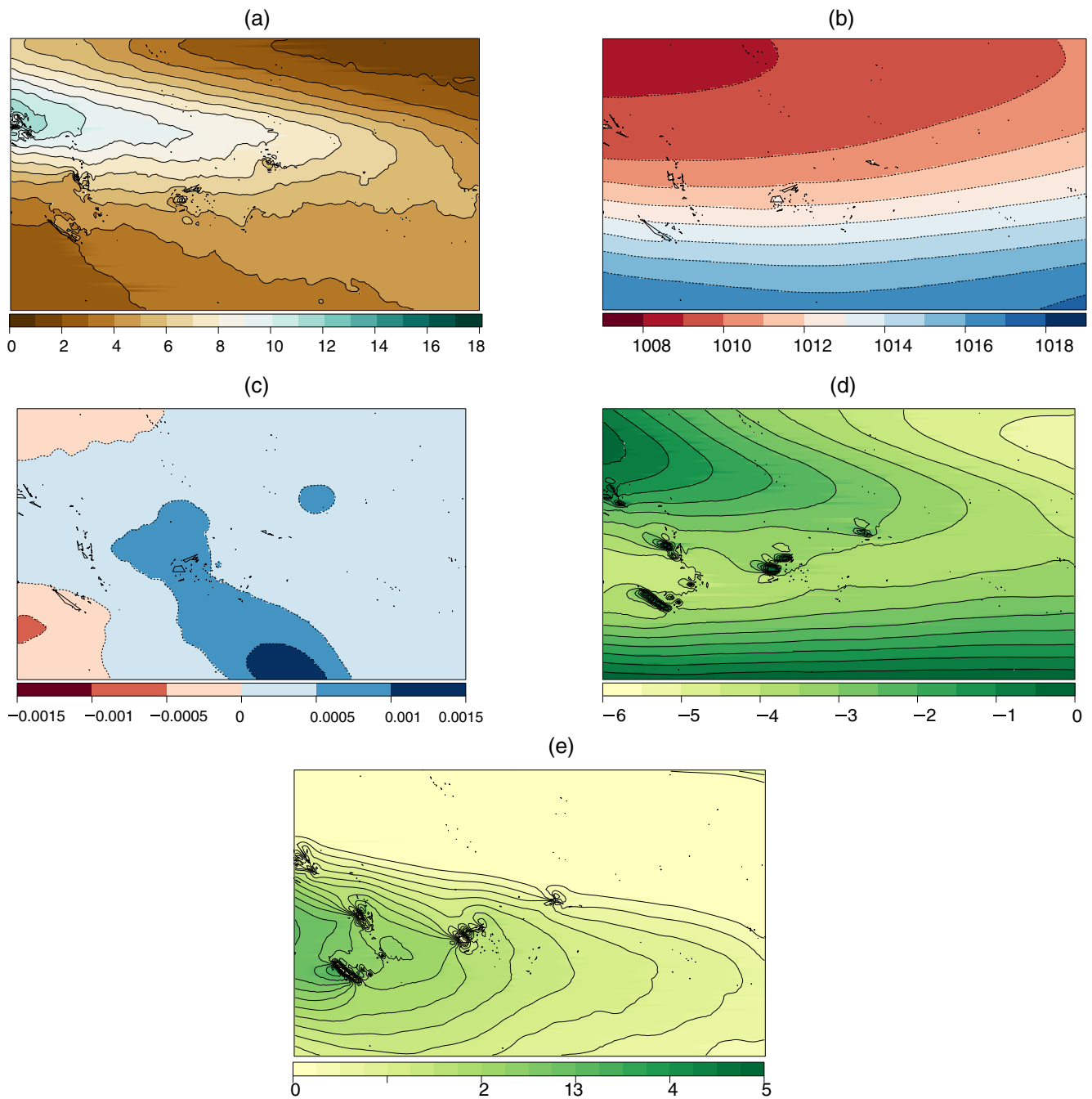


FIGURE A1 ERA5 climatologies (1998–2019) of (a) precipitation ($\text{mm}\cdot\text{day}^{-1}$), (b) mean sea-level pressure (SLP, in hPa), (c) SLP_{diff} (hPa), (d, e) 10 m eastward and northward components of wind (u and v , $\text{m}\cdot\text{s}^{-1}$) [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

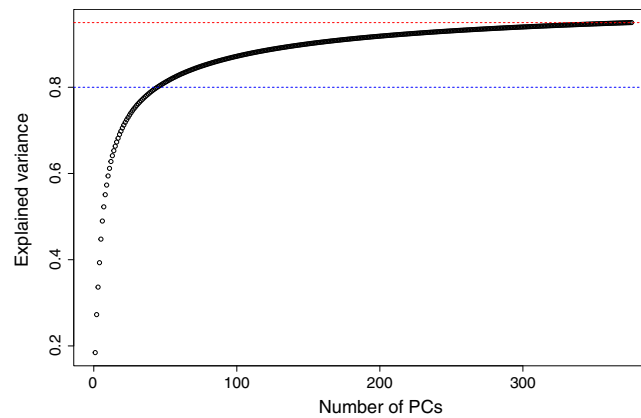


FIGURE A2 Cumulative explained variance of principal components (section 2.5). The lower and upper horizontal dotted lines indicate the 80% (chosen in this study) and 95%, respectively [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/joc.7905)]

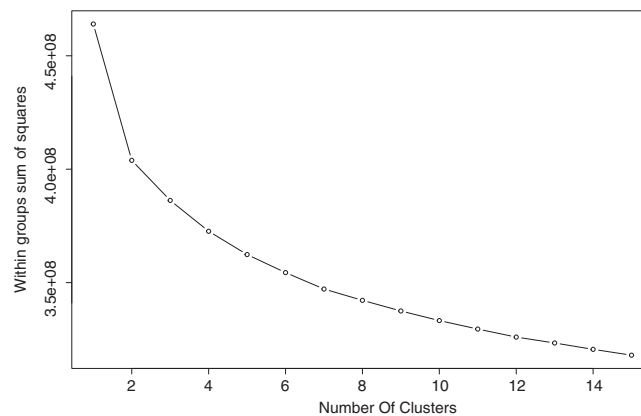


FIGURE A3 Total within-group sum of squares error as a function of the number of clusters for the training sample in the *k*-means clustering

APPENDIX B: RELATIONSHIP OF WEATHER TYPES WITH TROPICAL CYCLONE ACTIVITY AND SPCZ STATE B

In this section we further investigate the physical interpretation of the weather types generated in light of the most relevant climatic features of the South Pacific region. As previously described in section 1, warm-phase (El Niño) sea-surface temperature (SST) conditions produce a northeastward displacement of the SPCZ, while cold SST anomalies are associated with a southwestward displacement of the SPCZ (Folland *et al.*, 2002), having an effect on precipitation patterns and TC activity within the study region as it has been shown. In order to analyse the SPCZ position and its relationship with the WTs generated and TC frequency, we will use the SPCZ Position Index (SPI), envisaged as a quantitative synthetic descriptor of the SPCZ state (Folland *et al.*, 2002; Salinger

et al., 2014). The SPI is computed as the normalized difference in mean sea level pressure during the November–April period between Suva (Fiji) and Apia (Samoa), located on opposite sides of the middle position of the SPCZ (Figure 1). We computed SPI from the ERA5 mean sea-level pressure field considering the nearest data grid points to Apia and Suva locations and the period 1998–2019, and calculated their normalized difference following the SPI definition by Folland *et al.* (2002). In order to gain a better insight into the relationship of our weather typing with large-scale climatic features, we characterize the SPI and analyse TC frequency and precipitation climatologies of each WT conditioned to the SPCZ state. It has been shown that TC activity is mostly concentrated in the period December–April, and mostly linked to WTs 4 and 5 (Figure 3b). However, when each WT is conditioned to SPI tertiles, different spatial patterns of TC

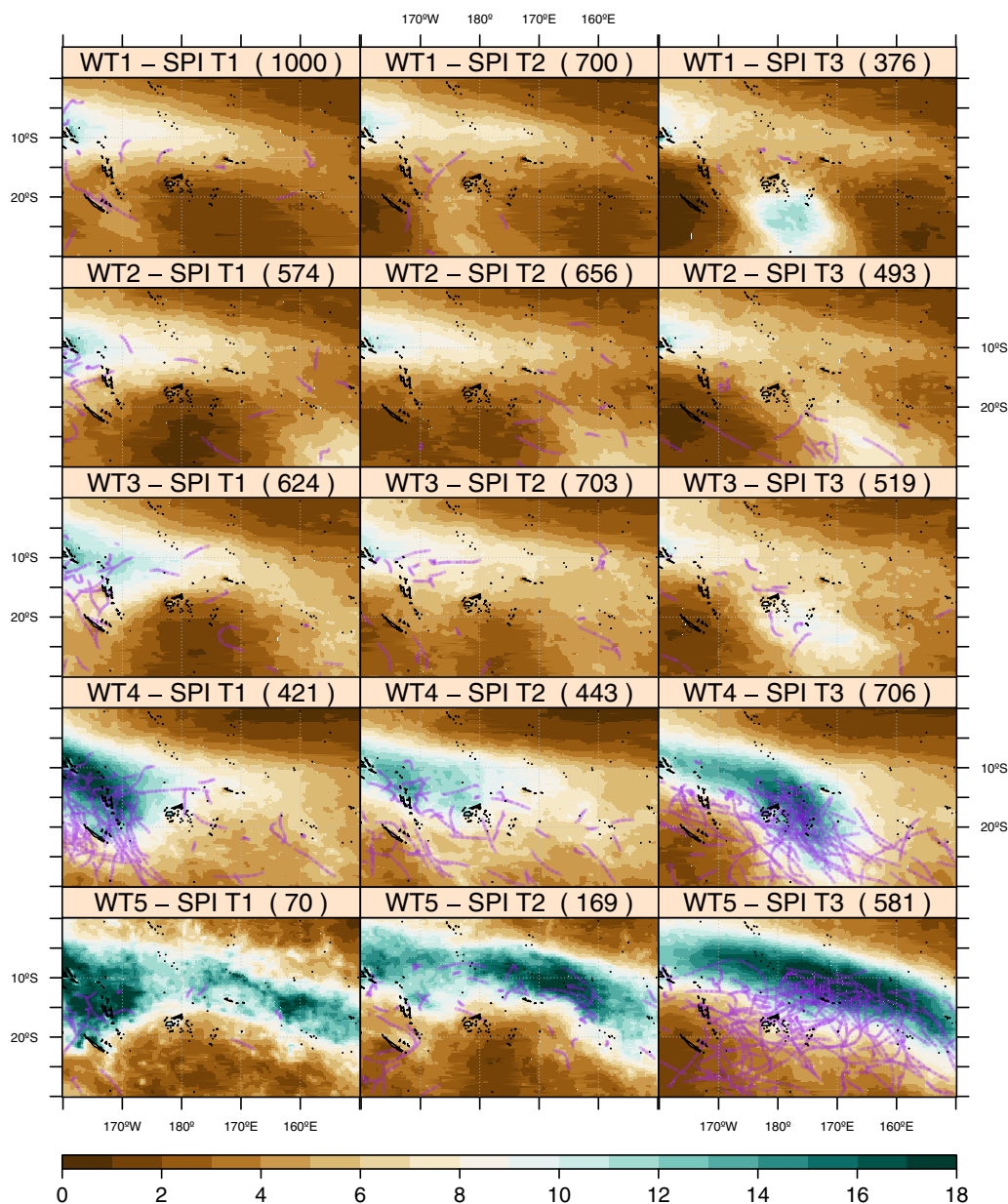


FIGURE A4 Precipitation climatology (1998–2019, mm-day⁻¹) conditioned to WT and South Pacific Convergence Zone Position Index (SPI) tertile (T1 upper tertile, T2 medium tertile, T3 lower tertile). The lines depict the TC tracks occurring in each case [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/joc.7905)]

tracks emerge, unveiling how the SPCZ modulates the TC frequency and trajectory within each WT (Figure A4). This analysis also helps to better understand the “within-cluster” variability, that could be captured through weather typing by defining a larger k number (at the cost of obtaining less likely groups and thus lower calibration robustness).

Considering the results for the WTs 4 and 5, highest precipitation and TC tracks are arranged in the mid-western part of the work, in line with the cyclone-prone area described by Deo *et al.* (2021). The TC

tracks tend to displace southwards starting from a generation point located in the area of highest precipitation. The bulk of TC activity concentrates within the area of strongest SPCZ influence, with a very few exceptions. Considering the medium SPI tertile, the arrangement of the traces becomes less evident, and TC frequency is relatively lower. An eastward broadening of the SPCZ causes a sparse distribution of the TC tracks within the region, which is common to both WTs. As a result, a SPI-conditioned probability of TC activity is found within the region within each

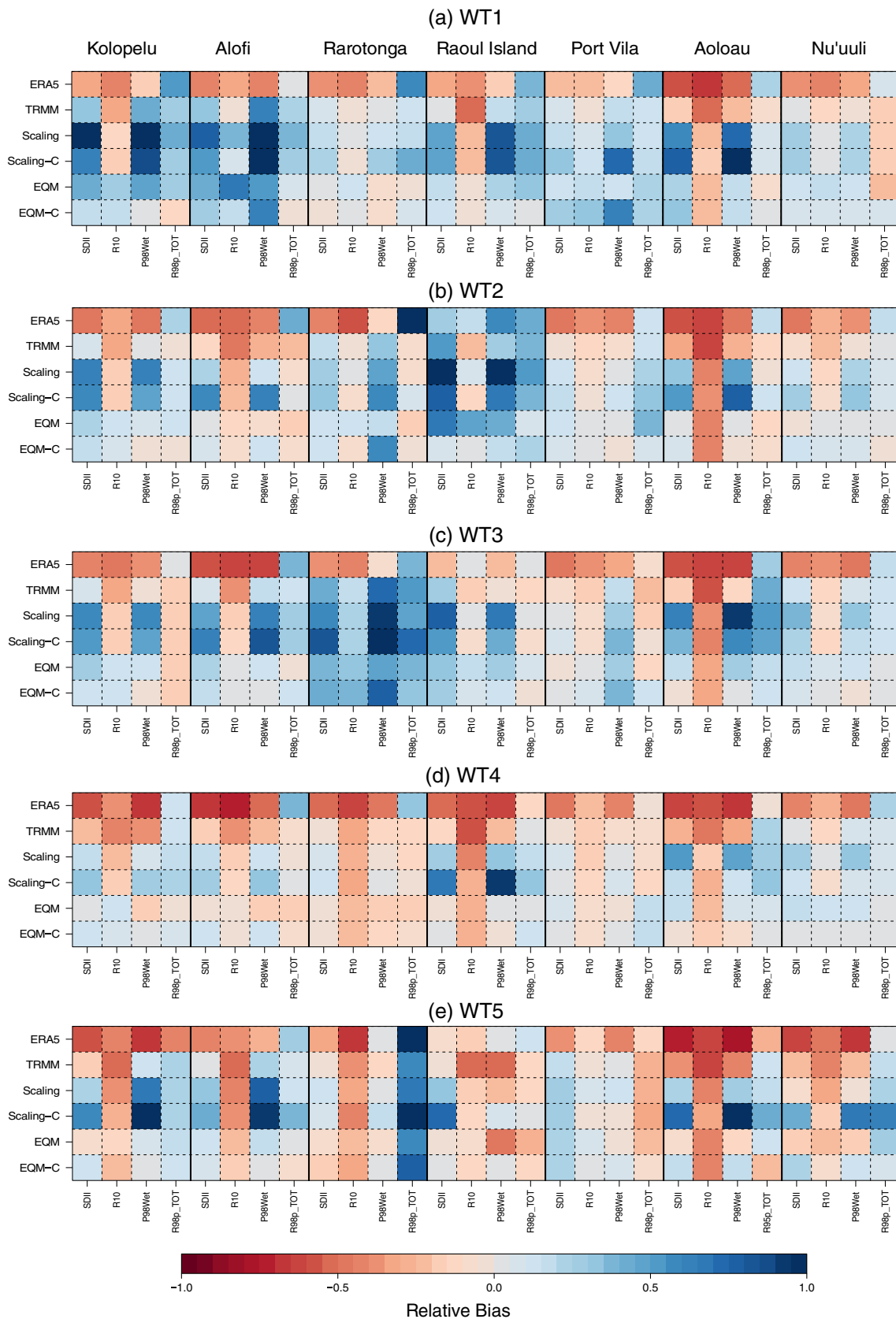


FIGURE A5 Relative bias of (raw) precipitation products (ERA5, TRMM) and calibrated TRMM w.r.t. the reference PACRAIN rain gauge observations (Table 1). The calibration techniques are eQM, scaling, and their respective WT-conditioned counterparts (indicated with the suffix -C). The different validation indices (Table 2) are arranged in columns, and their relative bias indicated by the colorbar. In each panel (a–e), the results are shown after extracting all data values belonging to a particular WT (WT1–WT5) only [Colour figure can be viewed at wileyonlinelibrary.com]

WT. Considering the most cyclone-prone WTs 4 and 5, a high probability of TC occurrence over the mid-western part of the domain (New Caledonia, Vanuatu) can be expected in below-average SPI situations, while a much higher probability of TC affection can be expected in these WTs under above-average SPI events in the central part of the domain, mostly affecting the land areas of Fiji and Tonga.

APPENDIX C: VALIDATION OF TRMM CALIBRATION FOR EACH WEATHER TYPE C

In this section we show the calibration evaluation results, disaggregated by weather types. The WT5 results (panel e) are already shown in Figure 4b of the paper, but it is here repeated for completeness and ease of comparison.