Pitfalls in using statistical bias-correction methods to characterize climate change impacts

Nicolás A. Vásquez[1], Pablo A. Mendoza[1], Wouter Johannes Maria Knoben[2], Louise Arnal[2], Miguel Lagos-Zúñiga[1], Martyn P. Clark[3], and Ximena Vargas[1]

[1]Universidad de Chile
[2]University of Saskatchewan
[3]Centre for Hydrology, University of Saskacthewan Coldwater Laboratory

November 20, 2023

Abstract

Characterizing climate change impacts on water resources typically relies on Global Climate Model (GCM) outputs that are bias-corrected using observational datasets. In this process, two pivotal decisions are (i) the Bias Correction Method (BCM) and (ii) how to handle the historically observed time series, which can be used as a continuous whole (i.e., without dividing it into sub-periods), or partitioned into monthly, seasonal (e.g., three months), or any other temporal stratification (TS). Here, we examine how the interplay between the choice of BCM, TS, and the raw GCM seasonality may affect historical portrayals and projected changes. To this end, we use outputs from 29 GCMs belonging to the CMIP6 under the Shared Socioeconomic Pathway 5–8.5 scenario, using seven BCMs and three TSs (entire period, seasonal, and monthly). The results show that the effectiveness of BCMs in removing biases can vary depending on the TS and climate indices analyzed. Further, the choice of BCM and TS may yield different projected change signals and seasonality (especially for precipitation), even for climate models with low bias and a reasonable representation of precipitation seasonality during a reference period. Because some BCMs may be computationally expensive, we recommend using the linear scaling method as a diagnostics tool to assess how the choice of TS may affect the projected precipitation seasonality of a specific GCM. More generally, the results presented here unveil trade-offs in the way BCMs are applied, regardless of the climate regime, urging the hydroclimate community for a careful implementation of these techniques.

Pitfalls in using statistical bias-correction methods to characterize climate change impacts

Nicolás A. Vásquez[1], Pablo A. Mendoza[1,2], Wouter J. M. Knoben[3*], Louise Arnal[3†], Miguel Lagos-Zúñiga[4], Martyn Clark[3‡], Ximena Vargas[1]

[1]Department of Civil Engineering, Universidad de Chile
[2]Advanced Mining Technology Center, Universidad de Chile
[3]Centre for Hydrology, University of Saskatchewan, Canmore, Alberta, Canada
[4]Center for Climate and Resilience Research, Universidad de Chile

Key Points:

- The choice of temporal stratification for GCM bias correction is crucial for removing biases, even for GCMs with good raw seasonality.
- Different temporal stratifications used for GCM bias correction may yield different future seasonalities and signals in projected changes.
- The scaling factor method is effective to assess if the temporal stratification affects the precipitation seasonality projected by a GCM.

[*]now at the Department of Civil Engineering, Schulich School of Engineering, University of Calgary, Calgary, Alberta, Canada
[†]now at Ouranos, Montreal, Quebec, Canada
[‡]now at the Department of Civil Engineering, Schulich School of Engineering, University of Calgary, Calgary, Alberta, Canada

Corresponding author: Nicolás A. Vásquez, `nicolas.vasquez.pl@uchile.cl`

Abstract

Characterizing climate change impacts on water resources typically relies on Global Climate Model (GCM) outputs that are bias-corrected using observational datasets. In this process, two pivotal decisions are (i) the Bias Correction Method (BCM) and (ii) how to handle the historically observed time series, which can be used as a continuous whole (i.e., without dividing it into sub-periods), or partitioned into monthly, seasonal (e.g., three months), or any other temporal stratification (TS). Here, we examine how the interplay between the choice of BCM, TS, and the raw GCM seasonality may affect historical portrayals and projected changes. To this end, we use outputs from 29 GCMs belonging to the CMIP6 under the Shared Socioeconomic Pathway 5–8.5 scenario, using seven BCMs and three TSs (entire period, seasonal, and monthly). The results show that the effectiveness of BCMs in removing biases can vary depending on the TS and climate indices analyzed. Further, the choice of BCM and TS may yield different projected change signals and seasonality (especially for precipitation), even for climate models with low bias and a reasonable representation of precipitation seasonality during a reference period. Because some BCMs may be computationally expensive, we recommend using the linear scaling method as a diagnostics tool to assess how the choice of TS may affect the projected precipitation seasonality of a specific GCM. More generally, the results presented here unveil trade-offs in the way BCMs are applied, regardless of the climate regime, urging the hydroclimate community for a careful implementation of these techniques.

Plain Language Summary

Global Climate Models (GCMs) are useful tools to characterize the historical and future evolution of the Earth's climate and its impacts on water resources. Because these models contain errors and their horizontal resolution is too coarse for local impact assessments, spatial downscaling and bias correction are required steps. In particular, bias correction methods can be trained and applied using all the available historical data or by splitting the time series (e.g., by season or months). Since there is no guideline on selecting a temporal stratification, we analyze bias-corrected GCM outputs obtained with three types of strategy (entire period, seasons, and months) and seven bias-correction techniques over continental Chile. We show that the choice of bias correction method and the temporal stratification applied can modify the projected precipitation signal and seasonality. We also propose a simple statistical technique to identify if, for a given climate model, the temporal stratification may be a relevant decision for climate impact assessments.

1 Introduction

Understanding and quantifying climate change impacts is crucial for long-term water resources planning and management. Such characterization typically involves hydrologic model simulations forced by an ensemble of scenario-driven meteorological time series obtained from Statistically Downscaled Bias-Corrected (SDBC) Global Climate Model (GCM) outputs (e.g., Addor et al., 2014; Hattermann et al., 2018; Her et al., 2019; Chen et al., 2021; Hanus et al., 2021; Vicuña et al., 2021). This approach usually requires the choice of emission scenario (e.g., Vano et al., 2015; Chegwidden et al., 2019), the choice of GCM (e.g., Hakala et al., 2018; Di Virgilio et al., 2022), the selection of Bias Correction Method (BCM) (e.g., Werner & Cannon, 2016; Gutiérrez et al., 2019; Hess et al., 2023), and the choice of observational (or reference) dataset (e.g., Wootten et al., 2021; Rastogi et al., 2022).

Among the above decisions, the selection and configuration of BCMs is a critical step given the risk of introducing artificial perturbations in GCM outputs (Hagemann et al., 2011; Maurer & Pierce, 2014; Wootten et al., 2021), generating a mismatch between simulated (i.e., obtained from bias-corrected GCMs) and observed (i.e., obtained from a reference dataset) annual cycles of climate variables (e.g., precipitation; Teutschbein & Seibert, 2010; Alder & Hostetler, 2019; Chen et al., 2021), with potential effects on projected climate change impacts and subsequent interpretations and adaptation strategies. A somewhat overlooked step is the strategy for handling the time series when applying BCMs, hereafter referred to as temporal stratification (TS). For example, the bias correction of simulated daily time series can be performed using all the historical period (i.e., a single application of the BCM; e.g., Ghimire et al., 2019) or sub-periods of the historical time series, such as seasons (e.g., four applications of the BCM; e.g., Ruffault et al., 2014; Teng et al., 2015), months (i.e., twelve applications of the BCM; e.g., Pierce et al., 2015; Switanek et al., 2017; Matiu & Hanzer, 2022; Wu et al., 2022; J. Guo et al., 2023), or any other temporal window (e.g., Haerter et al., 2011; Reiter et al., 2018).

Despite the large body of work exploring modeling decisions at the top of the 'cascade of uncertainty' (Wilby & Dessai, 2010), climate impact studies have typically relied on subjectively selected TSs. For example, Teng et al. (2015) compared four BCMs (applied with a seasonal TS) for hydrological projections in southeastern Australia, concluding that the hydrological model amplifies biases in precipitation after applying the BCMs, and that the large spread in the projected signal of changes in precipitation extremes yields different impacts on runoff. Hakala et al. (2018) applied the quantile mapping (QM) method (using a seasonal TS) to assess whether a hydrological model, forced by SDBC GCMs, can replicate the hydrological climatology observed during a historical reference period, obtaining that, even after bias correction, biases in precipitation and streamflow seasonality persist. To analyze the effects of different observational datasets and BCMs on climate projections, Wootten et al. (2021) used three observational datasets to apply two BCMs: (i) the 'Delta' approach with a 3-month moving window, and (ii) the quantile delta mapping (QDM) method over four periods consisting of three non-overlapping months. They concluded that the selection of BCMs and observational datasets have different impacts on historical and projected time series for different variables, although they did not isolate the effect of the TS.

Other studies have focused on the ability of different BCMs to reproduce historically observed climate indices (e.g., Gutmann et al., 2014; François et al., 2020; Xavier et al., 2022), or the effects on climate projections (e.g., Maurer & Pierce, 2014; Melsen et al., 2018), without emphasizing the role of the TS and the evaluation timescale. More recently, Vogel et al. (2023) proposed a framework to evaluate downscaling and BCMs for climate change studies and demonstrated it over Australia using four GCMs, three BCMs and two downscaling methods, considering different TS (monthly, 3-month, and multi-time scales) for the BCMs. They suggested that the TS may influence the analysis (after bias correction) and should be adequately chosen after a careful bias assessment.

Although the preceding studies have covered domains with specific climate types, the trade-offs in selecting TS, BCMs, and GCMs for estimating historical biases (after applying BCMs) and projections across contrasting climates remain unclear. Hence, this paper seeks to disentangle the relative contribution of these decisions (especially TS) to the spread of bias-corrected time series at the annual, seasonal, and monthly timescales during historical and future periods rather than finding the 'best' configuration for the assessment of climate change impacts. Specifically, we address the following research questions:

1. To what extent does the choice of bias correction method and temporal stratification alter historical GCM simulations across different climate regions?
2. What are the effects of bias correction methods and temporal stratification on the projected signal and seasonality of different climate variables?

3. Are there any connections between the effects of TS (on historical biases and projections) and the capability of raw GCM output to replicate historically observed climatology?

To seek answers, we evaluate the performance of 29 SDBC GCMs from the sixth phase of the Coupled Models Intercomparison Project (CMIP6; O'Neill et al., 2016) over different climate groups in continental Chile. We use seven methods (three univariate and four multivariate) to correct biases in precipitation and maximum and minimum temperature. All BCMs are applied at three different TSs: (i) using the entire period (i.e., all daily data simultaneously used for one application of the BCM), (ii) seasonally (i.e., four applications of the BCM using four seasonally stratified time series), and (iii) monthly (i.e., twelve applications of the BCM for twelve monthly stratified time series).

2 Study area and datasets

2.1 Study area

Our study domain is continental Chile, which is suitable for a comprehensive assessment of the TS-BCM-GCM interplay in very different climate types. Figure 1 shows the spatial distribution of mean annual precipitation, mean annual temperature, and three climate indices. The snowfall fraction $SF = Sn/P$ (Figure 1d) is the fraction of mean annual precipitation (P, Figure 1b) falling as snow (Sn). The aridity index (Figure 1e) is the ratio between mean annual potential evapotranspiration (PET) and mean annual precipitation. Finally, the precipitation seasonality (p-seasonality, Figure 1f) indicates whether most precipitation falls during winter (negative values) or summer (positive values). In this paper, we use the season names within the context of the Southern Hemisphere (i.e., winter refers to months JJA, while summer to DJF).

In the northern area (17°S-25°S), two main climate zones can be identified: (i) the super-arid coastal area, with very low annual precipitation amounts (<50 mm/yr), and (ii) the Altiplano region, with lower temperatures due to increasing altitude and larger annual precipitation (∼200 mm/yr). The mean annual precipitation increases towards the south, although the Andes Cordillera generates a west-east gradient, with larger precipitation amounts and lower temperatures on the western slopes of the Andes Cordillera compared to the valleys. Moving south from ∼37°S, the altitude of Andean mountains progressively decreases, as well as the contribution of snowmelt to runoff, whereas precipitation increases. South from 45°S, a west-to-east precipitation gradient produces high precipitation amounts on the coast (>2500 mm/yr), whereas a dry climate develops in Patagonia a few kilometers to the east, with decreasing precipitation amounts. In summary: (i) most snowfall occurs in the Andes Cordillera, though snowfall events can also occur in the valleys of Austral Chile (<45°S); (ii) the hydroclimate is water-limited ($PET/P > 1$) in approximately half of the Chilean territory, especially from ∼35°S to the north, whereas the hydroclimate of the south is energy limited ($PET/P < 1$); and (iii) most precipitation in Chile falls during the winter (red color in panel f), being the Altiplano (northern Chile) and Patagonia (∼50-55°S) two notable exceptions. For a more comprehensive review of the climate and weather of Chile, readers are referred to Aceituno et al. (2021) and Vásquez et al. (2021).

2.2 Datasets

We use the gridded meteorological product CR2MET v2.5 (Boisier et al., 2018; DGA, 2022) as the observational baseline (hereafter reference dataset). CR2MET precipitation estimates (pr) are obtained through a combination of (i) logistic regression models and (ii) multiple linear regression models that use ERA5 reanalysis outputs (Hersbach et al., 2020) and geomorphological attributes as predictors and daily precipitation from meteorological stations as predictands. For daily extreme temperatures (tmax and tmin),

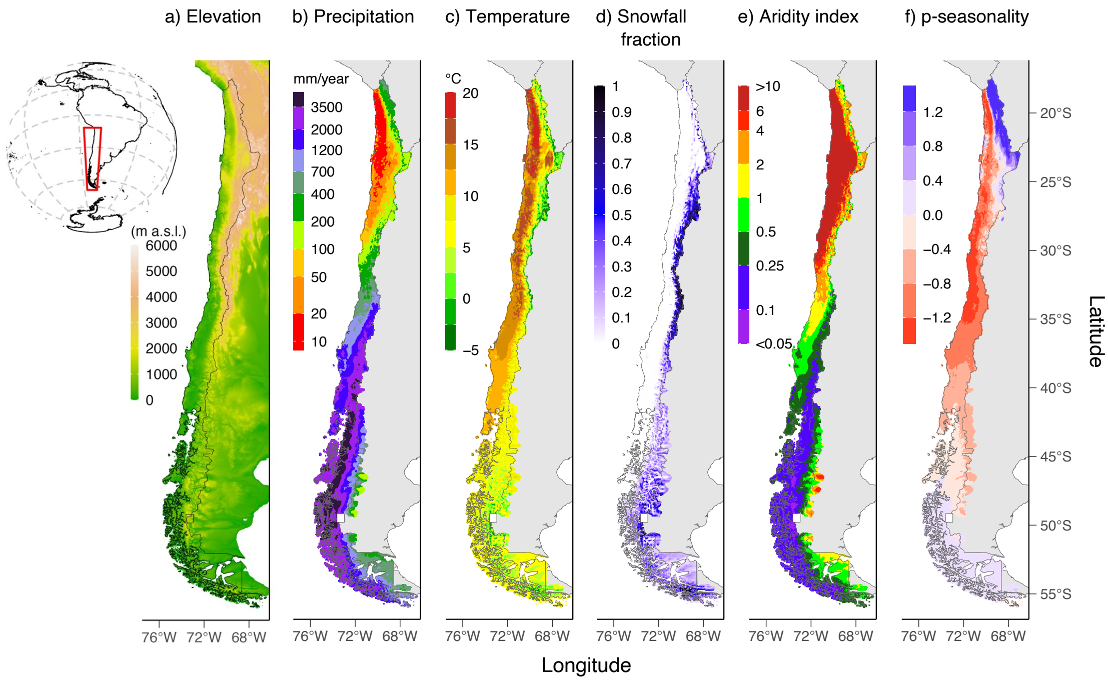

Figure 1. Main physiographic and climate attributes of continental Chile for the period 1980-2014 (34 water years): (a) elevation, (b) mean annual precipitation, (c) mean annual temperature, (d) snowfall fraction, (e) aridity index, and (f) p-seasonality.

land surface temperature from MODIS AQUA and TERRA (Wan, 2014) are also included as predictors. All variables (pr, tmax, and tmin) are available at a daily time step for the period January/1979-March/2020, covering continental Chile at a horizontal resolution of 0.05° x 0.05°. The mean daily temperature is computed as the average between tmax and tmin. It should be noted that CR2MET is, arguably, the most accurate meteorological dataset for continental Chile since its development incorporated local meteorological stations.

We use outputs from 29 GCMs from the CMIP6 (O'Neill et al., 2016), based on the data availability for pr, tmax and tmin during the historical and projected periods, and the SSP5-8.5 scenario for being the worst in terms of greenhouse emissions and the 'business as usual' development case. The name and horizontal resolution of each GCM are included in Table A1.

3 Methodology

Figure 2 shows the main steps of our approach. First, we delineate climate zones across Chile using cluster analysis (step 1), with the aim to examine possible relationships between climate types and the BCM-TS-GCM interplay. Step 2 considers different strategies for correcting biases in GCM outputs (i.e., seven bias-correction methods are applied using three different stratification periods). In step 3, we compute several climate indices derived from precipitation and temperature at different time scales (e.g., annual, seasonal, and monthly mean values), for a historical and a future period. Finally, we conduct an Analysis of Variance (ANOVA) to quantify the relative contribution of different decisions to the spread of historical estimates. More details can be found in the following sections.

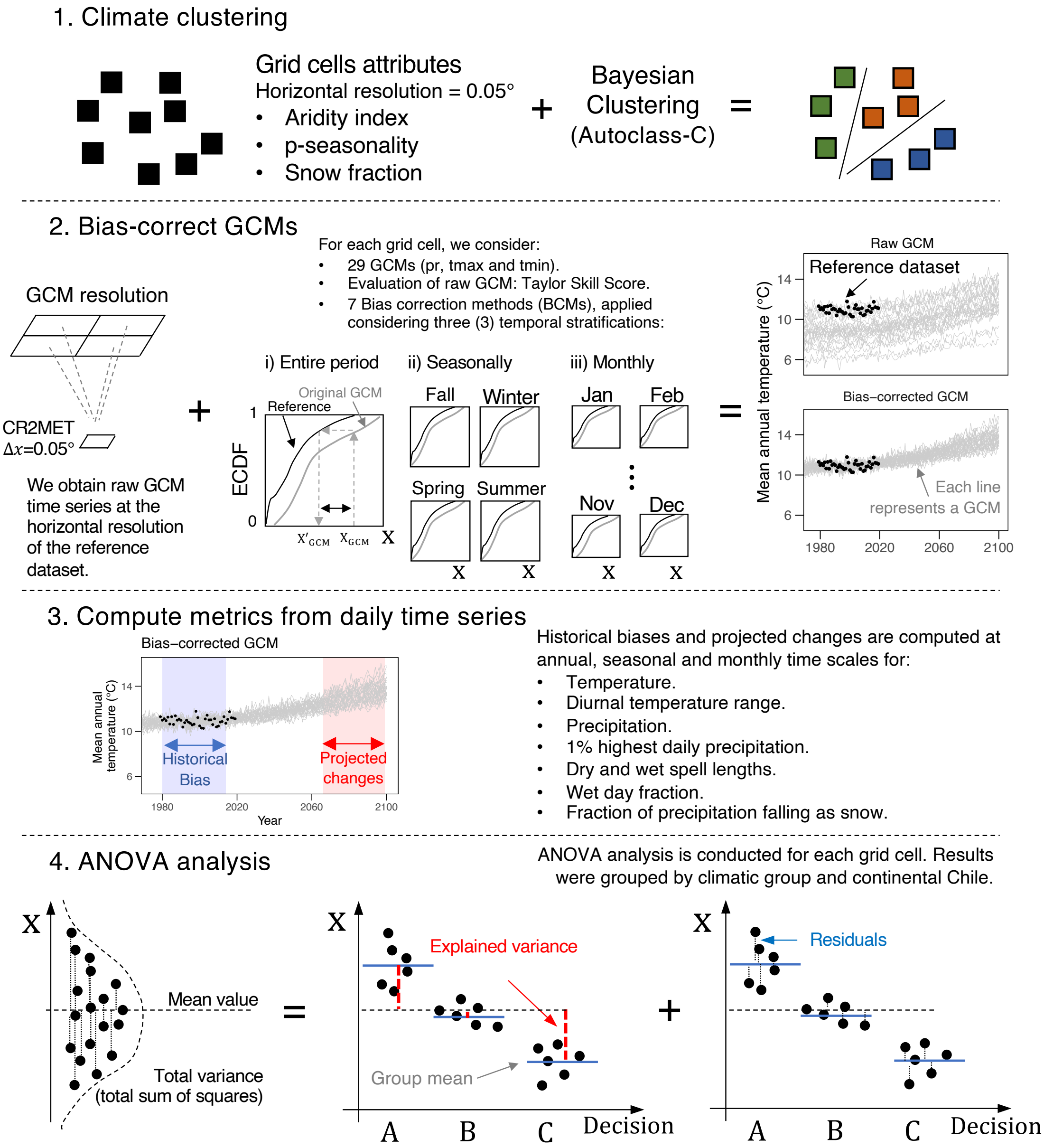

Figure 2. Diagram of the methodology used in this study

3.1 Climate clustering

We perform a Bayesian clustering to identify climate zones across Chile. To this end, we use the aridity index (PET/P), the p-seasonality, and the fraction of precipitation falling as snow as explanatory variables, since they reflect observed hydrological behaviors (Knoben et al., 2018). PET is computed using the Oudin et al. (2005) formula - available in the R Package airGR (Coron et al., 2017) - which requires air temperature (provided at daily time steps here) and latitude as inputs. To estimate Sn, we consider that snowfall occurs when the mean daily temperature is below 2°C (Jennings et al., 2018; Han et al., 2019; Sepúlveda et al., 2022), and p-seasonality is computed with the formula proposed by Woods (2009).
Prior climate groups are defined with the Autoclass-C software (Cheeseman et al., 1988, 1996), which has been previously used in hydrological applications (e.g., Sawicz et al., 2011). We subsequently refined the clustering results through visual inspection, grouping small clusters based on spatial proximity and climate similarity.

3.2 Raw GCM performance

We use the Taylor Skill Score (TSS; Taylor, 2001) to evaluate the role of the raw GCM performance and its interplay with BCM and TS for SDBC-biases and projections at different time scales. The TSS is computed at the grid cell level (0.05° x 0.05°) for the period 1980-2014, contrasting downscaled GCM outputs against the reference dataset, as is commonly done for local climate impact assessments (e.g., Lafon et al., 2013). In this study, TSS is computed for precipitation, as shown in Eq. 1.

$$TSS = \frac{4(1+R)}{\left(\hat{\sigma} + \frac{1}{\hat{\sigma}}\right)^2 (1+R_o)} \tag{1}$$

where R is the Pearson correlation coefficient between the raw GCM and the reference mean seasonality, and $\hat{\sigma} = \sigma_{GCM}/\sigma_{REF}$ is the ratio between the standard deviation of raw monthly values (σ_{GCM}) and the reference (σ_{REF}). R, and $\hat{\sigma}$ are computed using simulated and observed mean monthly values of each variable (i.e., 12 values of GCMs vs. 12 reference values). R_o is the maximum achievable Pearson correlation coefficient for a specific GCM, which is assumed to be $R_o \cong 1$ to simplify the analysis. When $R \rightarrow R_o$ and $\hat{\sigma} \rightarrow 1$, the $TSS \rightarrow 1$. Alternatively, $TSS \rightarrow 0$ when R decreases or $\hat{\sigma}$ approaches zero or infinity. Hence, TSS ranges between 0 and 1. Further, we compute the TSS for each climate group, estimating the mean group climatology through spatial averages.

3.3 Bias correction of GCMs

3.3.1 Bias correction methods

We downscale the raw GCM outputs to the CR2MET grid using inverse distance weighting, considering the four closest GCM grid cells. We use seven bias correction methods, including three univariate and four multivariate techniques, listed in Table 1 and briefly reviewed here. The quantile delta mapping (QDM) preserves the projected change for each quantile while correcting the bias. Empirical cumulative density functions are estimated for the historical reference ($F_{h,ref}$), the raw historical GCM ($F_{h,GCM}$), and the raw projected GCM ($F_{p,GCM}$) to relate (X) with the cumulative probability (τ). For a specific value during the historical period $X_{h,GCM}$, the correction (for pr) is given by $X^{'}_{h,GCM} = F^{-1}_{h,ref}\left(F_{h,GCM}\left(X_{h,GCM}\right)\right)$, while for a projected raw GCM value $X_{p,GCM}$, the corrected value is $X^{'}_{p,GCM} = \Delta \cdot F^{-1}_{h,GCM}\left(F_{p,GCM}\left(X_{p,GCM}\right)\right)$, where Δ is computed as $\Delta = X_{p,GCM}/F^{-1}_{h,GCM}\left(F_{p,GCM}\left(X_{p,GCM}\right)\right)$ for precipitation.
The asynchronous regression (AR) relies on a piecewise linear regression calibrated with sorted raw GCM and reference data during a historical period (i.e., $F_{h,ref}$ is a function of $F_{h,GCM}$). Although a simple linear regression could be used, the error in the tails of

the regression can be large and, therefore, the data is split by including different knots (up to six) to reduce errors in low and high values. To bias-correct projected values, the calibrated piecewise linear regression is applied. The quantile regressions neural network (QRNN) uses neural networks to bias correct the sorted data (i.e., quantiles) from simulations and the reference. QRNN is a flexible model since it does not assume a specific relationship between the raw GCM and the reference data.
The rank resampling for distributions and dependences (R^2D^2) corrects the covariance among sites and/or variables through four steps: (i) the univariate bias correction of each variable/site separately, (ii) the selection of one variable/site and the computation of the ranking for all variables/sites, (iii) for a specific date, select the same ranking in the reference period for the dimension selected, and (iv) the shuffling of the other variables/sites to maintain rank structure.
The 'multivariate bias correction' family (MBC) includes three different methods using the Pearson correlation coefficient (MBCp), the Spearman rank correlation coefficient (MBCr), and an N-dimensional probability density function (MBCn) to transform the raw correlated GCM data (i.e., the intervariable dependence structure) through consecutive iterations. For MBCp and MBCr, the transformation relies on the Cholesky matrix decomposition and the correction of the covariance matrix. Conversely, MBCn relies on an orthogonal rotation, the application of QDM to these orthogonal variables, and, finally, the application of an inverse matrix (the one used to compute the orthogonal variables) to obtain the resulting data. The reader is referred to the studies listed in Table 1 for more details on the methods.

Table 1. Methods considered in this study to bias-correct GCMs outputs (pr, tmax, and tmin).

Acronym	Name	Type	Reference
QDM	Quantile Delta Mapping	Univariate	Cannon et al. (2015)
AR	Asynchronous Regression		Dettinger et al. (2004); Stoner et al. (2013)
QRNN	Quantile Regression Neural Network		Cannon (2011)
R^2D^2	Rank Resampling for Distributions and Dependences	Multivariate	Vrac and Thao (2020)
MBCp	Multivariate Bias Correction method - Pearson		Cannon (2016)
MBCr	Multivariate Bias Correction method - Rank		
MBCn	Multivariate Bias Correction method – QDM		Cannon (2018)

We stress that it is not our aim to perform detailed comparisons among different bias correction techniques but to quantify the impact of this and other methodological choices on historical biases and projected changes in climate indices. All bias correction methods were applied using the statistical software 'R' (http://www.r-project.org/). The QDM, MBCp, MBCr, MBCn, and R^2D^2 methods were applied using the library 'MBC' (Cannon, 2018). QRNN was implemented using the 'qrnn' library (also available in R), while the AR method was implemented following Stoner et al. (2013). To reduce the computational effort, we randomly select 100 grid cells within each climate group, and all subsequent analyses are conducted at these grid cells ($100 \cdot N_{clusters}$).

3.3.2 Choice of the temporal stratification

Bias correction methods can be applied using different stratification strategies. For example, a BCM can be applied at daily time steps using all the data in the historical period (usually 30 years), which means that all ∼10,950 days (∼365 days · 30 years) are simultaneously bias-corrected. For a seasonal TS, BCMs are applied four times, each one considering ∼2730 days (∼91 days · 30 years), whereas for a monthly TS, the BCM is applied 12 times considering ∼900 days (∼30 days · 30 years). Note that other temporal stratifications could be considered. Here, we applied BCMs to daily time series of pr, tmax, and tmin (e.g., Rastogi et al., 2022) using the entire time series in the historical period (1980-2014), and stratifying the data seasonally and monthly, since these TSs are typically considered for climate change impact assessments. For all combinations of BCM and TS, we obtained daily time series from 1980 to 2100.

3.4 Climate indices

We consider several climate indices that are relevant to reproduce historically observed hydrological responses (e.g., Gutmann et al., 2014), including (i) mean annual, seasonal, and monthly total precipitation, (ii) highest 1% daily precipitation, (iii), wet-day fraction, (iv) wet and dry-spell lengths, (v) fraction of precipitation falling as snow, and (vi) annual, seasonal and monthly averages of mean daily temperature and diurnal temperature ranges. To estimate the mean annual snowfall, we add all precipitation amounts for days with a mean daily temperature below 2°C. Wet-spell and dry-spell lengths (mean consecutive rainy and non-rainy days, respectively), as well as the wet-day fraction (mean fraction of rainy days) are computed as in Gutmann et al. (2014), considering 0.1 mm/d as a threshold. To examine the capability of BCMs to replicate historically observed climate indices, we computed the difference between SDBC-GCM outputs and the reference dataset during the historical period 1980-2014 as a percent bias (hereafter referred to as biases). Additionally, we analyze the effects of BCMs on climate projections by computing the relative change for the period 2065-2099 with respect to the historical period (1980-2014).

3.5 Analysis of Variance

To evaluate the relative contribution of the BCM and TS decisions to the spread of SDBC-biases we perform, for each combination of GCM and grid cell, an analysis of variance (ANOVA). In this case, the ANOVA is simplified as:

$$TV = BCM + AP + Residual \tag{2}$$

where TV stands for the total variance of SDBC-biases, and the residual term is the variance not explained by the BCM nor the TS for a specific GCM-grid cell combination. If the choice of TS had no impact on the biases in climate indices. In that case, the application of Supposey BCM should be able to reduce biases at all temporal scales (e.g., annual, seasonal, or monthly), regardless of the GCM considered. To summarize the information at the grid cell level, we compute the average of BCM/TV, TS/TV, and $Residual/TV$ fractions across GCMs, whereas for the climate groups, we compute the mean relative contribution (estimated by BCM/TV, TS/TV and $Residual/TV$) of TS and BCM to the spread as the average of fractions across the grid cells within that group.

4 Results

We show the climate clustering results, the historical biases after applying the BCMs, and the relative contributions of different methodological choices to historical biases of climate indices at the annual and seasonal scales. Further, we include the TSS performance to examine connections between the raw seasonality of the GCMs and the selection of BCM and TS. For simplicity, we only show the results for precipitation, and the remaining variables can be found in the Supporting Information.

4.1 Clustering

The Bayesian clustering and subsequent spatial aggregation through visual inspection provided ten climate groups for continental Chile (Figure 3). In general, the clusters follow two main climate patterns in Chile: (i) a latitudinal precipitation gradient, from very arid (north) to humid (south), and (ii) a west-east gradient from the coast to the Andes Cordillera. Although northern Chile encloses groups 1, 2, and 3, clusters 2 and 3 are located in the Altiplano region, where larger precipitation and lower temperatures are observed. Groups 5, 6, and 8 span the coast and valley, whereas groups 4 and 7 are located in the Andes. Finally, groups 9 (the rainiest group) and 10 are in southern Chile, characterized by large precipitation amounts in the Andes Cordillera and the coast, with decreasing precipitation and temperature towards the east (Patagonia).

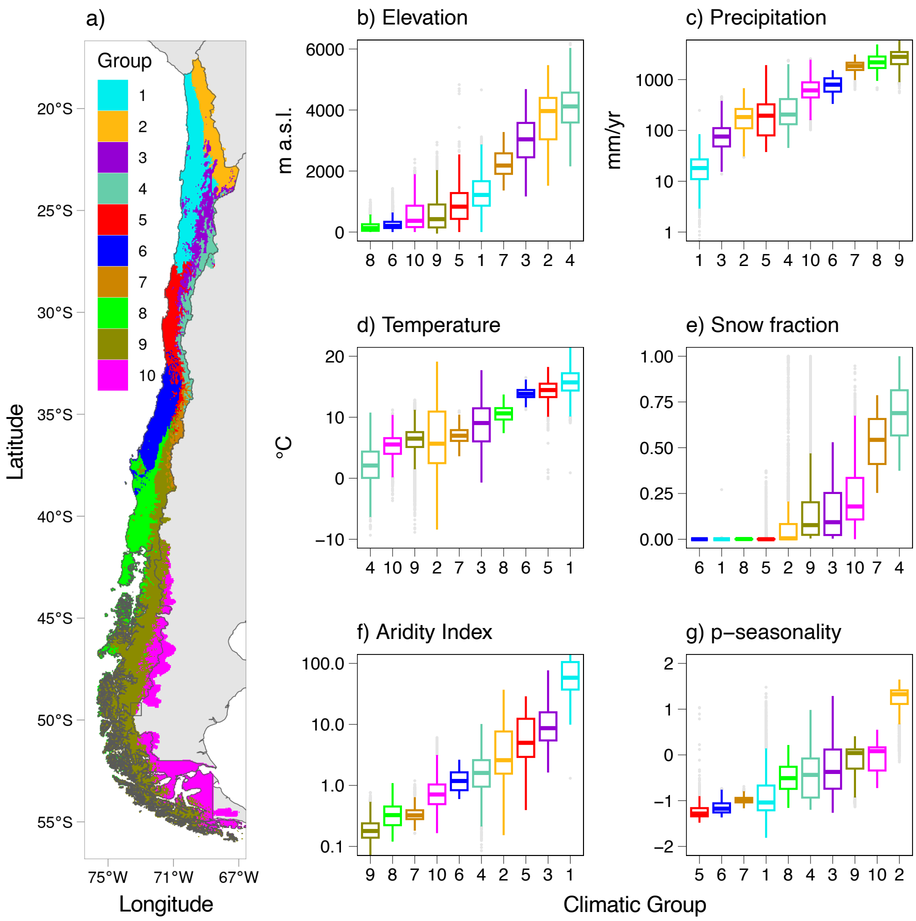

Figure 3. (a) Spatial distribution of climate clusters in continental Chile based on snowfall fraction, aridity index, and p-seasonality. The following attributes are ordered by the median of each group: (b) elevation, (c) precipitation, (d) temperature, (e) snowfall fraction, (f) aridity index, and (g) p-seasonality. All climate indices were computed for the period 1980-2014. Notice that the boxplots in panels b-g are sorted according to the median value, and the group's order on the x-axis differs among variables.

4.2 Performance metrics after bias correction

Figure 4 shows precipitation biases (after bias correction) in three different climate groups (the other variables and climate groups can be found in the Supporting Information). The results show that, regardless of the combination of GCM, BCM, TS and grid cell, biases in annual amounts are close to zero (Figure 4a). When the BCM is applied using all the data in the historical period(Figure 4b, left), biases in monthly precipitation amounts can be large, although the magnitude varies among climate groups. In climate group 2 (Altiplano region), precipitation occurs mostly during the summer (DJF); in this season, the median bias associated with January precipitation is relatively lower - though still considerable (>20%) - compared to the remaining months. In group 6, most precipitation occurs during the winter (JJA), and biases can be found in any month. In group 10, precipitation falls uniformly throughout the year, with slightly larger amounts and larger biases during the summer (DJF). When the BCM is applied seasonally (4b, center), monthly precipitation biases persist. However, these are generally lower compared to the case when the bias correction is applied using the entire dataset, especially in climate group 10. As expected, biases are nearly removed with a monthly TS (Figure 4b, right), regardless of the GCM, bias correction method, grid cell, or climate group.

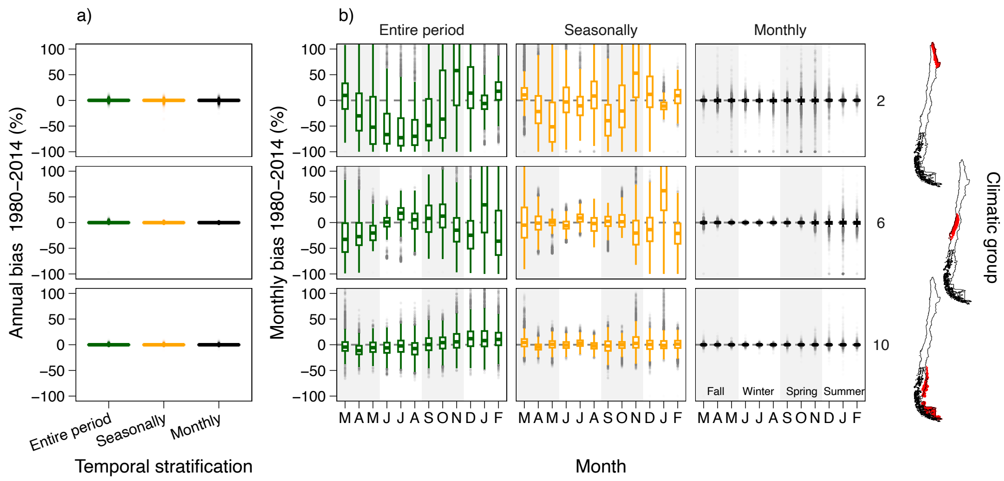

Figure 4. Historical biases in precipitation at the (a) annual and (b) seasonal time scales in three climate groups (rows) after applying the BCMs. The columns in panel b) show results for the three TSs used to apply the BCMs. Each boxplot comprises results from the 100 grid cells within a specific climate group, 29 GCMs, and seven BCMs. The different seasons are highlighted through grey-white areas.

Figure 5 displays the relative contributions of the BCM, TS, and residuals for mean annual, seasonal (summer and winter), and monthly (January and July) precipitation biases averaged across 1,000 grid cells in continental Chile. We show two seasons and months to examine possible differences between the dry and wet seasons. Additionally, the results from different GCMs are stratified according to their historical raw performance, measured by the Taylor Skill Score. As in Figure 4, the ANOVA analysis for historical biases shows differences among temporal stratifications, especially when compared to annual biases (Figure 5a). Because the relative contributions of BCM and TS to precipitation biases do not greatly differ among climate groups, we show results at the national scale. The choice of BCM explains most of the variance for the mean annual precipitation bias, whereas the choice of TS explains almost all the variance for mean seasonal and monthly precipitation biases. It is worth noting that the biases at the annual scale are, in general, very low (Figure 4, <1%), and that the relative importance of the choice of TS for seasonal and monthly biases does not decrease for GCMs with high TSS values. The latter result is counterintuitive since one might expect that GCMs with good raw precipitation seasonality will be effectively bias-corrected, regardless of the TS selected. For variables related to quantiles (highest 1% daily precipitation, dry and wet-spell lengths, and wet-day fraction), the relative importance of BCMs increases for GCMs with higher TSS, being BCM the most important decision, even at seasonally and monthly time scales (Figure S1).

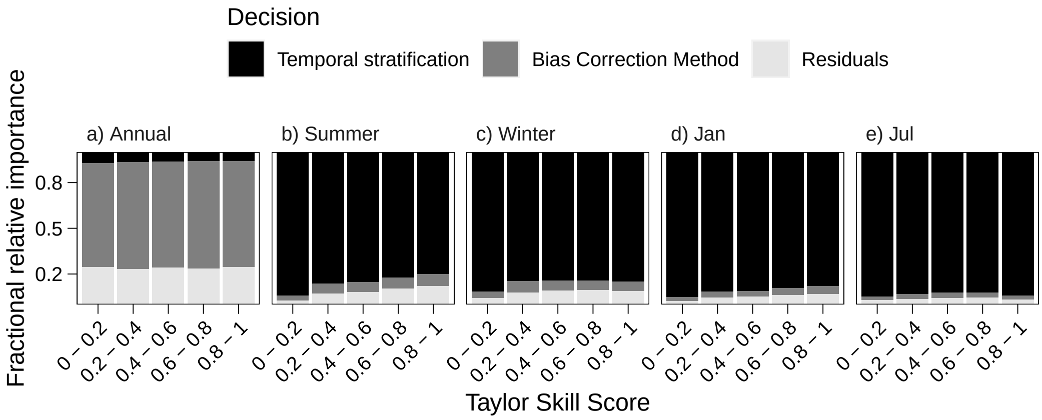

Figure 5. Relative importance (as a fraction averaged from all grid cells and GCMs for continental Chile) of the bias correction method and the temporal stratification to explain the precipitation biases at the annual, seasonal (DJF and JJA), and monthly (January and July) time scales during the historical period (1980-2014), for different levels of historical GCM performance (x-axis). Biases are computed after applying BCMs.

4.3 Projected changes

We now analyze the interplay between the choice of TS, the raw GCM precipitation seasonality, and its effects on projected changes in precipitation for the period 2065-2099 (with respect to 1980-2014) at different time scales. Figure 6 displays projected changes in mean annual, seasonal, and monthly precipitation for one grid cell located in central Chile (red dot in map) and one GCM (INM-CM4-8) with a high R value. For this GCM and grid cell, $TSS = 0.76$ during the period 1980-2014, with a Pearson correlation coefficient between mean monthly raw GCM and reference amounts of 0.98, and a 41% underestimation of the standard deviation. The high value of R indicates a good seasonality of raw GCM outputs. Figure 6 shows that different BCMs yield a high dispersion in projected changes of mean annual precipitation (different lines), with little influence on the selected TS (x-axis of each subplot). Additionally, all BCMs alter the raw GCM projection. For example, if all BCMs are applied using the entire dataset, projected changes in summer precipitation range between -8% to 5%, whereas the raw projection is close to -30%. The application of MBCn using the entire period yields a positive projected change in the mean summer precipitation, while a seasonal and monthly application of the same BCM projects a decrease in summer precipitation. The results for individual months (January and July) reveal more dispersion and interaction among BCMs and the choice of TS. For example, applying the BCM with the entire time series results in positive and negative projections of mean July precipitation (the rainiest month for this grid cell). Similarly, different TSs can also provide different projected signals.

Figure 6 reveals that the choice of TS affects the signal of projected changes in summer precipitation (e.g., for the MBCn method) and, in particular, in January and July precipitation amounts. The TS can be considered relevant for a specific grid cell if it is able to switch the projected signal of a variable for a particular GCM-BCM combination. This is, for example, the case of mean July precipitation (Figure 6), for which the

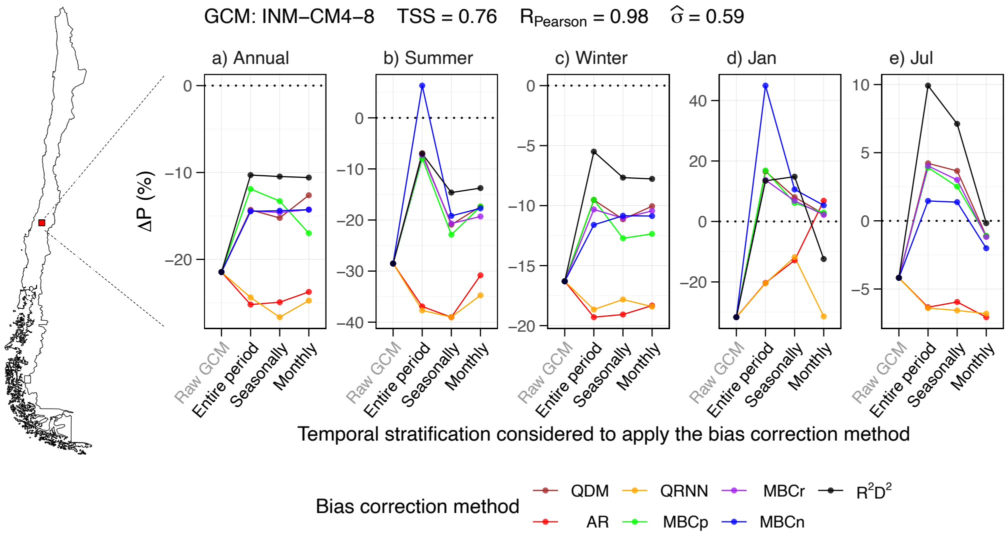

Figure 6. Projected change in annual, seasonal (summer and winter), and monthly (January and July) precipitation for different temporal stratifications (x-axis) and bias correction methods (lines). All combinations of TS and BCM decisions, along with projected changes from the raw (biased) GCMs, are displayed. The results are valid only for the grid cell shown and the GCM INM-CM4-8. The metrics (e.g., TSS) were computed using the raw (biased) GCM data for the period 1980-2014.

signal of projected changes is different among TSs for the MBCn, MBCr, and R^{2}D^2 methods.

Figure 7 shows, for all the grid cells analyzed, the fraction of 'well-behaved' GCMs (i.e., with $TSS \geq 0.7$; e.g., Kwon et al., 2019) for which the selection of TS leads to different signs in projected precipitation changes. Note that the number of GCMs that meet the performance requirement - obtained by spatially averaging the number of GCMs with $TSS \geq 0.7$ at each latitudinal band - varies along the domain. In general, the choice of TS does not alter the signal of projected changes in mean annual precipitation, although a few GCMs are affected by this decision in some areas (e.g., northern Chile). Nevertheless, the effects of TS are more evident in seasonal projections (Figure 7b and 7c). During the summer, >50% of the number of GCMs are affected by the TS in Central Chile (dry season). During winter, the Altiplano region and part of southern Chile are largely influenced by the choice of TS. It should be noted, however, that the summer season in Central Chile and the winter season in the Altiplano region are dry seasons. Therefore, while the signal of projected changes may vary for different TSs, the precipitation amounts involved are small. For mean monthly January and July precipitation, the choice of TS is even more relevant. Indeed, nearly all GCMs are affected by the TS along the coast of northern Chile, while ~50% of the GCMs yield different signals in projected changes for different TSs in Central Chile. The case of July is more interesting since it is the rainiest month in most of continental Chile. In July, ~50% of the GCMs are affected by the TS along the Central Chilean Andes (western border), impacting the accumulation of snow and, therefore, meltwater volume and timing estimates for the spring and summer seasons. In southern Chile, one can find grid cells where GCMs are affected by the TS decision, though that fraction is lower compared to the Central Chilean Andes.

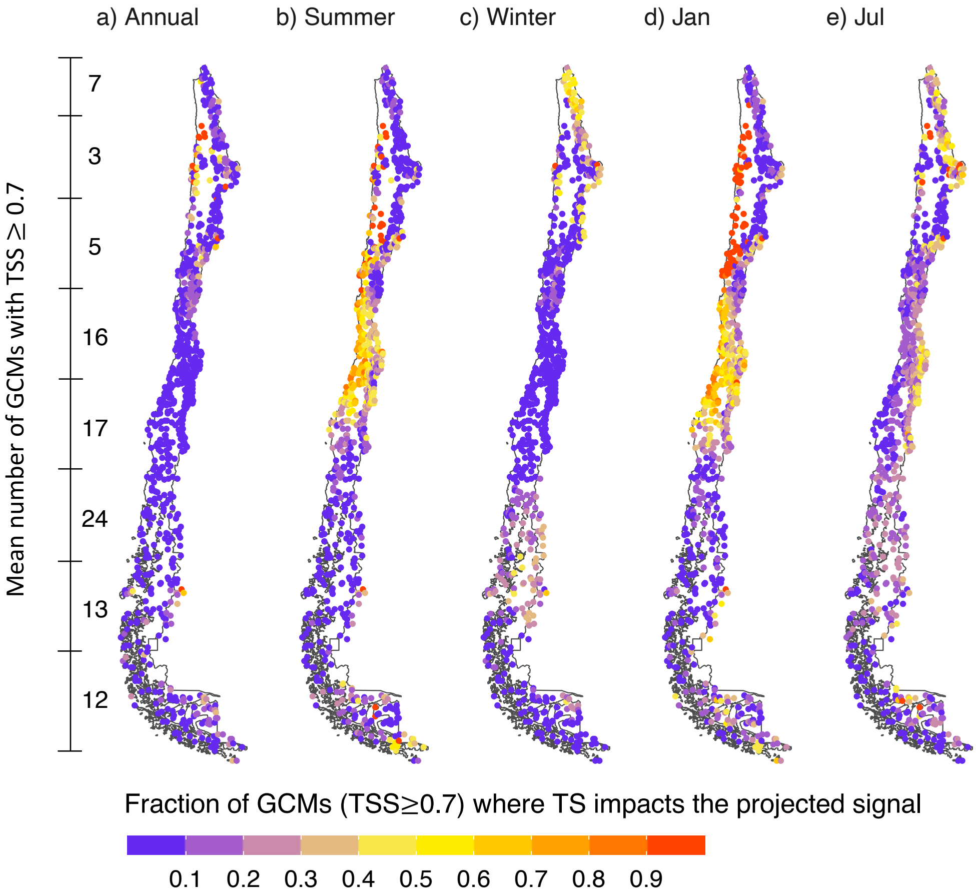

Figure 7. Fraction of GCMs with acceptable performance (i.e., with $TSS \geq 0.7$) for which the TS yields different projected precipitation signals. The number of GCMs that meet the threshold criteria at each $\sim 5°$ latitudinal band is computed as the average of GCMs with $TSS \geq 0.7$ from all grid cells within that band.

Figure 8a compares the raw GCM output (obtained from the GCM ACCESS-CM2) and the reference precipitation seasonality over a historical period at one grid cell located in central-southern Chile (red dot on the map). For this GCM-grid cell combination, $TSS = 0.96$, $R = 0.94$ and $\hat{\sigma} = 1.08$. Note that the GCM simulates the maximum monthly precipitation in July instead of June (when the maximum occurs according to the reference). Figure 8b displays, for the same GCM-grid cell, the projected precipitation seasonality for each BCM-TS combination (thin lighter lines). The results show that applying a BCM using the entire period (green lines) provides the same seasonality as the raw GCM; however, seasonal and monthly TSs distort the raw projected seasonality. Further, when BCMs are applied using a monthly TS (black/gray lines), the projected month of maximum precipitation is June, whereas for seasonal and entire period such month is July. Additionally, seasonal and monthly TSs yield higher precipitation fractions (compared to the raw GCM) during April and May, and smaller values during September and October. Such differences in projected precipitation seasonality may affect any subsequent analyses of simulated hydrological fluxes and states.

To examine the extent to which projected precipitation seasonality is affected by the temporal stratification, we focus on the projected maximum mean monthly precipitation. Hence, we contrast, for each GCM-grid cell combination, three curves obtained with the three temporal stratifications (each obtained by averaging the projections among BCMs for each GCM). We consider that the TS affects the projected seasonality if the month where the maximum mean monthly precipitation amount occurs differs. Conversely, if such a month is the same for the three TSs, we consider that this decision does not impact the seasonality. Figure 8c displays the fraction of the number of GCMs with $TSS \geq 0.7$ for which the TS impacts the projected precipitation seasonality. Interestingly, the number is relatively high (>40%) for most of continental Chile. The fraction of GCMs affected by the TS decision is even higher in northern Chile, the Central Chilean Andes, and the Southernmost part of Chile, where more than 60% of GCMs are affected.

5 Discussion

The results presented here highlight the relevance of the temporal stratification used when applying bias correction techniques, which affects (i) SDBC-biases in seasonal and monthly precipitation amounts over a historical period, and (ii) the signal of projected changes and the seasonality of projections.

5.1 Temporal stratification as a source of uncertainty

Our results show that the temporal stratification can largely affect precipitation biases during a historical period, as well as the signal and seasonality of projected changes. However, this methodological choice has been rarely explored in climate change impact assessments, and the lack of guidance has motivated the use of more than one TS in some studies (e.g., Wootten et al., 2021). Further, model errors may not necessarily be removed in the process. For example, Hakala et al. (2018) obtained that biases in precipitation and streamflow seasonality remained after applying BCMs. Here, we found that only a monthly application of the BCM can replicate the reference precipitation seasonality, even for GCMs with a good raw representation of annual cycles.

5.2 Projected seasonality

Our study reveals that one of the main effects of selecting different TSs is the possibility to distort the precipitation seasonality projected by raw GCM outputs. In hydrologic impact assessments, this artifact may propagate into the timing of simulated variables like snow accumulation and melting, energy fluxes, and streamflow (Meyer et al., 2019). Our results show that when the raw GCM seasonality has timing errors (com-

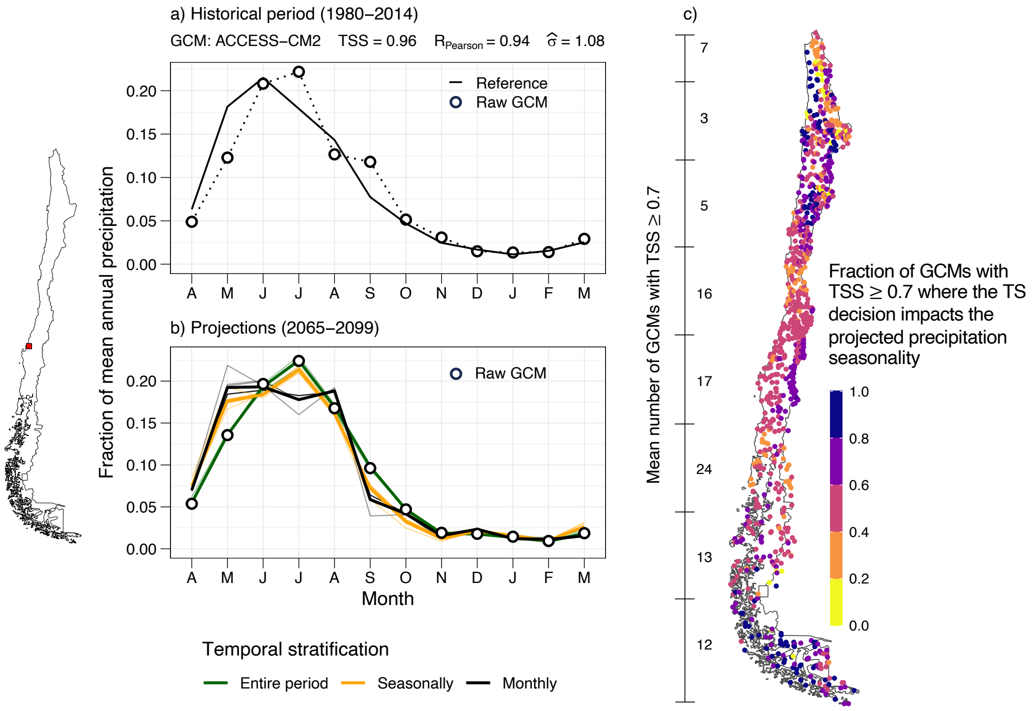

Figure 8. Influence of the temporal stratification used to apply bias correction methods on the projected precipitation seasonality. (a) Dimensionless historical seasonality for one grid cell (red dot on the map) and one GCM (ACCESS-CM2). Note that the sum of monthly fractions is equal to 1. (b) Projected raw (circles) and bias-corrected (colored lines) GCM precipitation seasonality. Lighter and thinner lines represent different BCMs, whereas thick lines represent the average across BCMs. (c) Fraction of the total number GCMs with $TSS \geq 0.7$, for which the temporal stratification yields different projected seasonality, measured as different months for maximum mean monthly precipitation for the 2065-2099 period. In c), the average number of GCMs meeting the TSS criterion is computed for latitudinal bands.

pared to the reference), a pronounced shift in the projected seasonality can be obtained after applying BCMs (compared to the case without bias correction). However, when the raw GCM replicates the historically observed precipitation seasonality reasonably well, one might expect that different TSs yield the same projected seasonality. To test this hypothesis, we compare the precipitation seasonality projected with three TSs (bottom panels) by two GCMs (CanESM5 and NorESM2-MM, Figure 9) that replicate annual cycles (i.e., high Pearson correlation coefficients, with GCM and reference maximum mean monthly precipitation being the same, top panels). For GCM CanESM5 (Figure 9a), the choice of TS has little effect on the projected precipitation seasonality. Conversely, the temporal stratification affects the seasonality projected by NorESM2 (Figure 9b). For example, if the BCM is applied seasonally and monthly, the months of maximum mean monthly precipitation are May and August, respectively. Interestingly, TSS = 0.951 for this GCM, which is higher than the value obtained for CanESM5 (0.694), and both GCMs have similar Pearson correlation coefficients. These results emphasize that even GCMs with a good raw representation of historical seasonality can be affected by the temporal stratification used to apply BCMs.

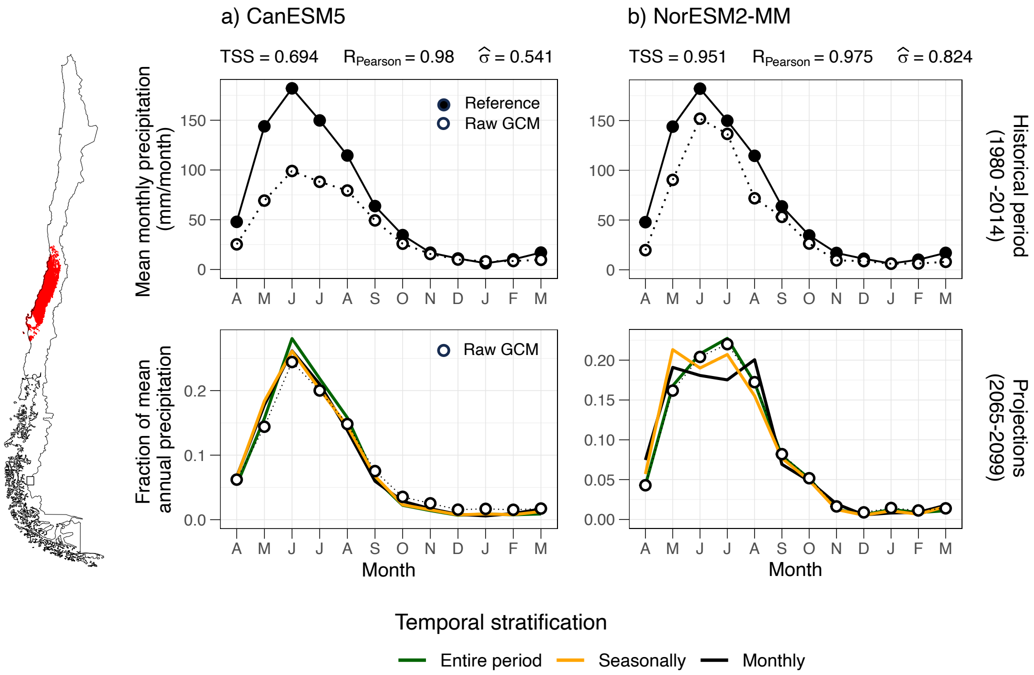

Figure 9. Impact of the temporal stratification used in bias correction for two GCMs. The results presented here are spatially averaged values of the grid cells contained in climate group 6 (highlighted in red on the map). Top row: comparison of the raw GCMs and the reference for the period 1980-2014. Bottom row: projected precipitation seasonality in terms of fraction of mean annual precipitation (average from the seven BCMs).

5.3 A priori evaluation of the TS impact on projected precipitation seasonality

Understanding the potential effects of the TS on the projected signal and seasonality of precipitation from a specific GCM could be helpful for a more detailed assessment of climate change and/or hydrological changes. Here, we propose using the linear scaling method (LSM) (Widmann et al., 2003; Maraun, 2016) - due to its low computational cost and simplicity (Lafon et al., 2013; Chaubey & Mall, 2023) -, as a quick diagnostics tool to inform if the TS may be an influential decision (an example of an LSM application is provided in Appendix B). The LSM removes the bias from the raw GCM time series (f_{bias}) through a multiplicative factor for the case of precipitation and an additive term for temperature, using an observational dataset as a reference. For example, if the reference and raw GCM mean annual precipitation amounts are 500 mm/year and 650 mm/year, respectively, a factor $f_{bias} = 500/650 = 0.77$ is applied to the raw GCM time series to remove the bias. Accordingly, seasonal or monthly applications of LSM require more scaling factors (Maraun et al., 2010). Hence, the raw GCM projected change (f_{Δ}) is preserved (at the TS time scale), since the scaling factors are typically considered to be time-invariant. Additionally, the influence of the temporal stratification and the reference dataset (in case there is more than one available) can be isolated for a specific grid cell-GCM combination.

Figure 10a illustrates the application of the linear scaling method (dashed lines) to the GFDL-CM4 GCM in one grid cell (red dot in map), using the entire period and stratifying the data seasonally and monthly. For this GCM-grid cell combination, $TSS = 0.72$ and $R = 0.7$, and different TSs yield different projected precipitation seasonalities when applying the LSM. Figure 10a shows that the precipitation factors obtained with LSM agree with the averages obtained from all (seven) bias correction methods (solid lines).

Finally, we examine the capability of the LSM to identify the precipitation seasonality projected with different TSs correctly. To this end we obtain, for each grid cell-GCM-TS combination, the precipitation seasonalities from (i) the average between the seven BCMs, and (ii) the application of the LSM. If the months of the projected maximum precipitation agree, we consider that the LSM correctly identifies the seasonality, and if this occurs for the three TSs, we consider that the LSM successfully identifies the projected bias-corrected seasonality for that specific grid cell-GCM combination. Figure 10a illustrates a successful case since, for each TS, the month of maximum precipitation is the same for the average among seven BCMs and from the LSM. Then we compute, for the 1,000 grid cells analyzed here, the fraction of GCMs for which the LSM successfully identifies the projected seasonality (accuracy, Figure 10b). The results show that, in almost all the grid cells, the LSM successfully identifies the projected seasonality of $\sim$70% of the GCMs, whereas for most grid cells ($> 85\%$), the LSM successfully projects the seasonality for more than 85% of the GCMs.

5.4 Limitations and future work

In this study, we selected the SSP5-8.5 scenario and 29 GCMs, although other future scenarios and/or a subset of GCMs could be considered to assess the effects on historical biases (after bias correction) and/or future projections. We did not focus on performance metrics for specific GCMs because evaluating the adequacy of particular bias correction methods is out of the scope of this work; instead, we focus on how these techniques are traditionally applied. Although we selected univariate and multivariate BCMs (e.g., Q. Guo et al., 2020), quantile-based, neural networks, and linear regressions, different approaches could be considered.

Additionally, we did not conduct any hydrological modeling. Instead, we focused on the repercussions of some decisions on the historical biases and the projected seasonality of climate variables required to run hydrological and land surface models. However, previous work has shown that hydrological models tend to amplify biases in the forcings (Teng et al., 2015). We emphasize that any assessment of climate change impacts should ensure that the climatological annual cycles of hydrological simulations forced with (i)

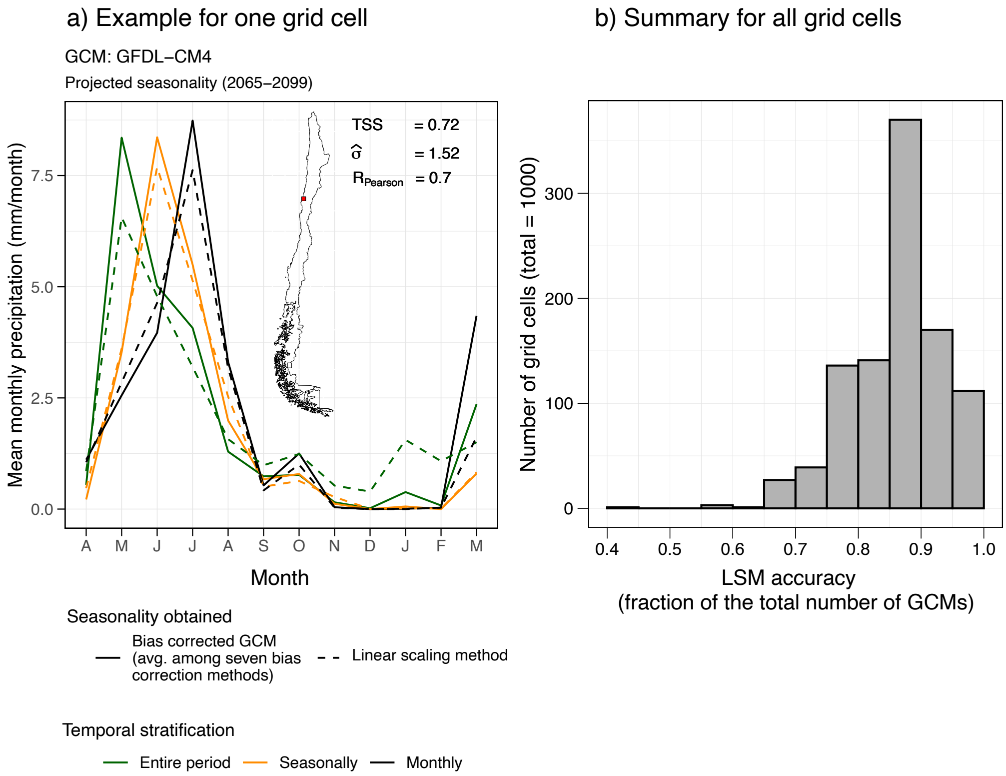

Figure 10. Linear scaling method used as a proxy to estimate the projected precipitation seasonality. (a) Example of projected precipitation seasonalities for one grid cell and one GCM, obtained from applying the LSM and the seven BCMs tested. The metrics summarize the raw (biased) GCM performance for the historical period (1980-2014). (b) LSM accuracy (as a fraction of the total number of GCMs) for all grid cells.

reference data sets and (ii) bias-corrected time series from GCMs/RCMs are similar (Hakala et al., 2018). Hence, verifying the reference and bias-corrected GCM forcing data during a historical period arises as a crucial step (Chen et al., 2013; Clark et al., 2016; Mendoza et al., 2016; Melsen et al., 2019). Future work could consider the impacts of SDBC historical biases and differences in projected seasonality on different aspects of the hydrograph (e.g., mean values, extremes, timing, etc.) and signatures formulated from other variables than streamflow (e.g., SWE, soil moisture; McMillan et al., 2022; Araki et al., 2022).

6 Conclusions

In this paper, we examined how methodological choices involved in GCM bias correction affect historical and future climate portrayals. To this end, we used seven bias correction methods, 29 CMIP6 GCMs, and three temporal stratifications. All the configurations were applied to daily time series of precipitation and maximum and minimum daily temperature derived from the CR2MET gridded observational product, available for continental Chile. Our main findings are as follows:

1. A monthly application of bias correction methods is required to replicate the reference precipitation seasonality, even for GCMs with good raw seasonality.
2. The temporal stratification is the most relevant decision to quantify seasonal and monthly precipitation biases.
3. Different temporal stratifications may yield different projected signals and seasonality, even for GCMs with good raw seasonality.
4. The linear scaling method can be used to estimate the projected seasonality of GCMs and, therefore, to identify the climate models for which the choice of temporal stratification may be critical, before applying more sophisticated and computationally expensive bias correction methods.

Appendix A Selected GCMs

Table A1 shows the GCMs included in this study.

Appendix B Scaling factor example

We illustrate the effects of the temporal stratification by applying the linear scaling method (LSM) (Maraun et al., 2010) for one grid cell-GCM combination. Figure B1a shows monthly precipitation averages from raw GCM outputs, whereas Figure B1b-d shows the bias-corrected GCM values for three different temporal stratifications. Monthly values were obtained from the daily corrected time series.
Note that when the entire period is used to bias-correct the GCM, only one factor is applied. In the grid cell analyzed, the reference annual precipitation is 4371 mm, which is below the historical raw GCM amount for the same period (5020 mm). Hence, the raw GCM precipitation time series is multiplied by the factor $f = 4731/5020 = 0.87$, which removes the annual SDBC bias; nevertheless, monthly SDBC-biases persist (see differences between black and blue lines in Figure B1b). When the LSM is applied seasonally, four factors are used to multiply the raw GCM time series. For example, daily values from March, April, and May are bias-corrected by the seasonal factor obtained from the reference (1134 mm/season) and the raw GCM (1498 mm/season) precipitation amounts. In this case, the factor used to bias-correct daily precipitation from March, April, and May is $f_{MAM} = 1134/1498 = 0.76$. Similarly, if the LSM is applied monthly, daily precipitation amounts from March are bias-corrected using the reference (374 mm/month) and raw GCM (498 mm/month), which yields a factor $f = 374/498 = 0.75$. For the monthly TS, the black and blue lines are the same. Note that the projected maximum

Table A1. GCMs considered in this study

GCM	Δlat	Δlon	Institution
ACCESS-CM2	1.25	1.88	Australian Research Council Centre of Excellence for Climate Science, Australia.
ACCESS-ESM1-5	1.25	1.88	
BCC-CSM2-MR	1.11	1.13	Beijing Climate Center, China.
CanESM5	2.77	2.81	Canadian Centre for Climate Modelling and Analysis, Canada.
CMCC-ESM2	0.94	1.25	Euro-Mediterranean Centre on Climate Change Coupled Climate Model, Italy.
CNRM-CM6-1-HR	0.50	0.50	Centre National de Recherches Météorologiques (CNRM), France.
CNRM-CM6-1	1.40	1.40	
CNRM-ESM2-1	1.40	1.41	
E3SM-1-0	1.00	1.00	Lawrence Livermore National Laboratory, USA.
EC-Earth3-CC	0.70	0.70	EC-Earth Consortium, Europe.
EC-Earth3-Veg-LR	1.12	1.13	
EC-Earth3-Veg	0.70	0.70	
EC-Earth3	0.70	0.70	
FGOALS-g3	2.18	2.00	Chinese Academy of Sciences Flexible Global Ocean-Atmosphere-Land System Model, China.
GFDL-CM4	1.00	1.25	Geophysical Fluid Dynamics Laboratory, USA.
GFDL-ESM4	1.00	1.25	
INM-CM4-8	1.50	2.00	Institute for Numerical Mathematics, Russia.
INM-CM5-0	1.50	2.00	
IPSL-CM6A-LR	1.27	2.50	Institute Pierre Simon Laplace (IPSL), France.
KACE-1-0-G	1.25	1.88	National Institute of Meteorological Sciences (NIMS) and Korea Meteorological Administration (KMA), South Korea.
KIOST-ESM	1.88	1.88	Korea Institute of Ocean Science and Technology Earth System Model and Its Simulation Characteristics, South Korea.
MIROC-ES2L	2.79	2.81	Japan Agency for Marina-Earth Science and Technology (JAMSTEC), Japan.
MIROC6	1.39	1.41	
MPI-ESM1-2-HR	0.93	0.94	Max Planck Institute for Meteorology (MPI-M), Germany.
MPI-ESM1-2-LR	1.87	1.88	
MRI-ESM2-0	1.11	1.13	Meteorological Research Institute, Japan.
NESM3	1.85	1.88	Nanjing University of Information Science and Technology Earth System Model, China.
NorESM2-MM	0.94	1.25	NorESM Climate modeling Consortium, Oslo, Norway.
TaiESM1	0.94	1.25	Research Center for Environmental Changes, Academia Sinica, Nankang, Taipei, Taiwan.

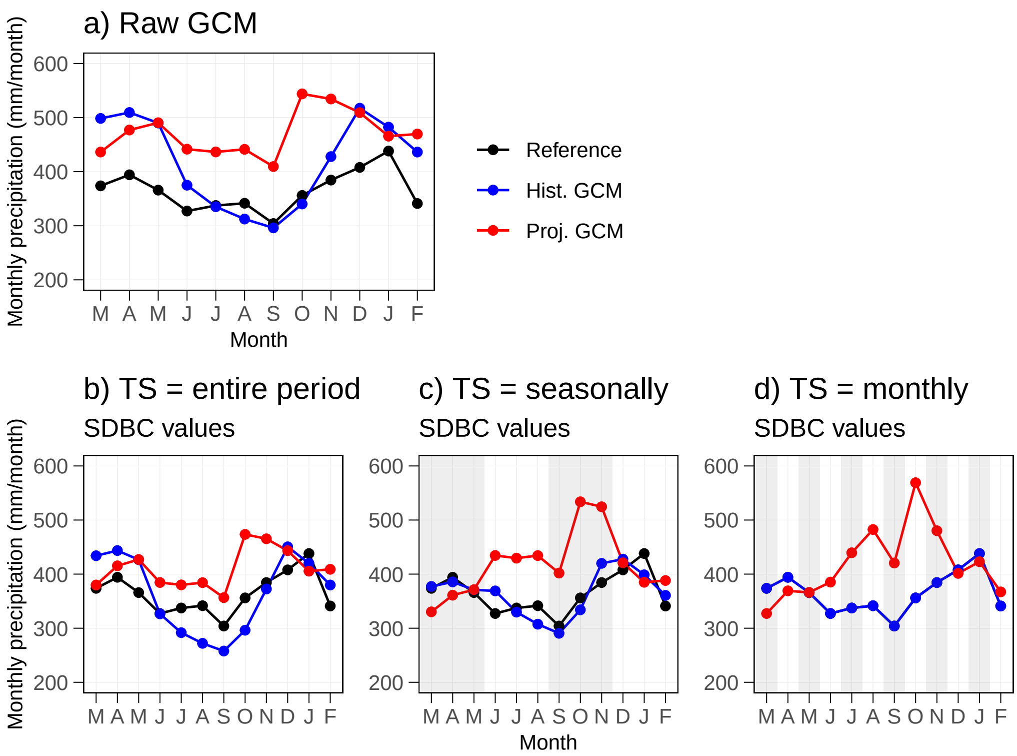

Figure B1. Illustration of the linear scaling method, applied to one grid cell-GCM combination, and its effects on the SDBC-biases and projections. (a) Reference (observational) and raw GCM seasonality during the period 1980-2014 (black and blue lines). The projected raw seasonality is also shown in red (2065-2099). (b), (c) and (d) show the bias-corrected precipitation amounts using the entire period, seasons, and months, respectively, for temporal stratification. The reference value is shown in all panels for completeness, and the shaded areas represent the temporal stratification.

monthly precipitation is October for the three TS, which is the same as the raw GCM projection. However, the projected minimum monthly precipitation is September, March, and March for the entire period, season, and monthly application of the LSM, respectively.

Open Research Section

The CR2MET dataset (Boisier et al., 2018) is available at https://www.cr2.cl/datos-productos-grillados/. The GCMs data was downloaded from the Earth System Grid Federation (https://esgf-node.llnl.gov/search/cmip6/). All the data used in this study is available at https://bhuch.myqnapcloud.com/share.cgi?ssid=43cb3da649cd41ca9bfc42150a855e89.

Acknowledgments

Nicolás Vásquez and Pablo A. Mendoza received support from the Fondecyt project No. 11200142. Nicolás Vásquez also received support from the Emerging Leaders in the Americas Program (ELAP) scholarship (Canada) and the ANID Doctorado Nacional scholarship No. 21230289 (Chile). Pablo A. Mendoza was also supported by ANID/PIA project No. AFB230001.

References

Aceituno, P., Boisier, J. P., Garreaud, R., Rondanelli, R., & Rutllant, J. A. (2021). Climate and Weather in Chile. In *Water resources of chile* (pp. 7–29). doi: 10.1007/978-3-030-56901-3{_}2

Addor, N., Rössler, O., Köplin, N., Huss, M., Weingartner, R., & Seibert, J. (2014). Robust changes and sources of uncertainty in the projected hydrological regimes of Swiss catchments. *Water Resources Research*. doi: 10.1002/2014WR015549

Alder, J. R., & Hostetler, S. W. (2019). The Dependence of Hydroclimate Projections in Snow-Dominated Regions of the Western United States on the Choice of Statistically Downscaled Climate Data. *Water Resources Research*. doi: 10.1029/2018WR023458

Araki, R., Branger, F., Wiekenkamp, I., & McMillan, H. (2022, 4). A signature-based approach to quantify soil moisture dynamics under contrasting land-uses. *Hydrological Processes*, *36*(4). doi: 10.1002/hyp.14553

Boisier, J. P., Álvarez-Garretón, C., Cepeda, J., Osses, A., Vásquez, N., & Rondanelli, R. (2018). CR2MET: A high-resolution precipitation and temperature dataset for hydroclimatic research in Chile. In *Egu general assembly conference abstracts* (p. 19739).

Cannon, A. J. (2011, 9). Quantile regression neural networks: Implementation in R and application to precipitation downscaling. *Computers & Geosciences*, *37*(9), 1277–1284. doi: 10.1016/j.cageo.2010.07.005

Cannon, A. J. (2016, 10). Multivariate Bias Correction of Climate Model Output: Matching Marginal Distributions and Intervariable Dependence Structure. *Journal of Climate*, *29*(19), 7045–7064. doi: 10.1175/JCLI-D-15-0679.1

Cannon, A. J. (2018). Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dynamics*. doi: 10.1007/s00382-017-3580-6

Cannon, A. J., Sobie, S. R., & Murdock, T. Q. (2015). Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate*. doi: 10.1175/JCLI-D-14-00754.1

Chaubey, P. K., & Mall, R. K. (2023, 9). Intensification of Extreme Rainfall in Indian River Basin: Using Bias Corrected CMIP6 Climate Data. *Earth's Future*, *11*(9). doi: 10.1029/2023EF003556

Cheeseman, P., John, R., & Nasa, S. (1996). Bayesian Classification (AutoClass): Theory and Results. *Advances in knowledge discovery and data mining*.

Cheeseman, P., Kelly, J., Self, M., Stutz, J., Taylor, W., & Freeman, D. (1988, 1). AutoClass: A Bayesian Classification System. *Machine Learning Proceedings 1988*, 54–64. doi: 10.1016/B978-0-934613-64-4.50011-6

Chegwidden, O. S., Nijssen, B., Rupp, D. E., Arnold, J. R., Clark, M. P., Hamman, J. J., ... Xiao, M. (2019). How Do Modeling Decisions Affect the Spread Among Hydrologic Climate Change Projections? Exploring a Large Ensemble of Simulations Across a Diversity of Hydroclimates. *Earth's Future*. doi: 10.1029/2018EF001047

Chen, J., Arsenault, R., Brissette, F. P., & Zhang, S. (2021). Climate Change Impact Studies: Should We Bias Correct Climate Model Outputs or Post-Process Impact Model Outputs? *Water Resources Research*. doi: 10.1029/2020WR028638

Chen, J., Brissette, F. P., Chaumont, D., & Braun, M. (2013, 7). Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water Resources Research*, *49*(7), 4187–4205. doi: 10.1002/wrcr.20331

Clark, M. P., Wilby, R. L., Gutmann, E. D., Vano, J. A., Gangopadhyay, S., Wood, A. W., ... Brekke, L. D. (2016, 6). Characterizing Uncertainty of the Hydrologic Impacts of Climate Change. *Current Climate Change Reports*, *2*(2), 55–64. doi: 10.1007/s40641-016-0034-x

Coron, L., Thirel, G., Delaigue, O., Perrin, C., & Andréassian, V. (2017, 8). The suite of lumped GR hydrological models in an R package. *Environmental Modelling & Software*, *94*, 166–171. doi: 10.1016/j.envsoft.2017.05.002

Dettinger, Cayan, D., Meyer, M., & Jeton, A. (2004). Simulated Hydrologic Responses To Climate Variations. *Climatic Change*.

DGA. (2022). *Homologación del cálculo hidrológico para la estimación de la oferta natural del agua histórica y futura en Chile.* (Tech. Rep.). SIT N° 524. Ministerio de Obras Públicas, Dirección General de Aguas, División de Estudios y Planificación, Chile. Elaborado por Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas. Retrieved from https://snia.mop.gob.cl/repositoriodga/handle/20.500.13000/126394

Di Virgilio, G., Ji, F., Tam, E., Nishant, N., Evans, J. P., Thomas, C., ... Delage, F. (2022, 4). Selecting CMIP6 GCMs for CORDEX Dynamical Downscaling: Model Performance, Independence, and Climate Change Signals. *Earth's Future*, *10*(4). doi: 10.1029/2021EF002625

François, B., Vrac, M., Cannon, A. J., Robin, Y., & Allard, D. (2020, 6). Multivariate bias corrections of climate simulations: which benefits for which losses? *Earth System Dynamics*, *11*(2), 537–562. doi: 10.5194/esd-11-537-2020

Ghimire, U., Srinivasan, G., & Agarwal, A. (2019, 3). Assessment of rainfall bias correction techniques for improved hydrological simulation. *International Journal of Climatology*, *39*(4), 2386–2399. doi: 10.1002/joc.5959

Guo, J., Wang, X., Fan, Y., Liang, X., Jia, H., & Liu, L. (2023, 4). How Extreme Events in China Would Be Affected by Global Warming—Insights From a Bias-Corrected CMIP6 Ensemble. *Earth's Future*, *11*(4). doi: 10.1029/2022EF003347

Guo, Q., Chen, J., Zhang, X. J., Xu, C., & Chen, H. (2020, 5). Impacts of Using State-of-the-Art Multivariate Bias Correction Methods on Hydrological Modeling Over North America. *Water Resources Research*, *56*(5). doi: 10.1029/2019WR026659

Gutiérrez, J. M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., ... Pagé, C. (2019, 7). 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*(9), 3750–3785. doi: 10.1002/joc.5462

Gutmann, E., Pruitt, T., Clark, M. P., Brekke, L., Arnold, J. R., Raff, D. A., & Rasmussen, R. M. (2014). An intercomparison of statistical downscaling methods used for water resource assessments in the United States. *Water Resources Research*. doi: 10.1002/2014WR015559

Haerter, J. O., Hagemann, S., Moseley, C., & Piani, C. (2011, 3). Climate model bias correction and the role of timescales. *Hydrology and Earth System Sciences*, *15*(3), 1065–1079. doi: 10.5194/hess-15-1065-2011

Hagemann, S., Chen, C., Haerter, J. O., Heinke, J., Gerten, D., & Piani, C. (2011, 8). Impact of a Statistical Bias Correction on the Projected Hydrological Changes Obtained from Three GCMs and Two Hydrology Models. *Journal of Hydrometeorology*, *12*(4), 556–578. doi: 10.1175/2011JHM1336.1

Hakala, K., Addor, N., & Seibert, J. (2018, 8). Hydrological Modeling to Evaluate Climate Model Simulations and Their Bias Correction. *Journal of Hydrometeorology*, *19*(8), 1321–1337. doi: 10.1175/JHM-D-17-0189.1

Han, P., Long, D., Han, Z., Du, M., Dai, L., & Hao, X. (2019, 4). Improved understanding of snowmelt runoff from the headwaters of China's Yangtze River using remotely sensed snow products and hydrological modeling. *Remote Sensing of Environment*, *224*, 44–59. doi: 10.1016/j.rse.2019.01.041

Hanus, S., Hrachowitz, M., Zekollari, H., Schoups, G., Vizcaino, M., & Kaitna, R. (2021). Future changes in annual, seasonal and monthly runoff signatures in contrasting Alpine catchments in Austria. *Hydrology and Earth System Sciences*. doi: 10.5194/hess-25-3429-2021

Hattermann, F. F., Vetter, T., Breuer, L., Su, B., Daggupati, P., Donnelly, C., ... Krysnaova, V. (2018). Sources of uncertainty in hydrological climate impact assessment: A cross-scale study. *Environmental Research Letters*. doi: 10.1088/1748-9326/aa9938

Her, Y., Yoo, S. H., Cho, J., Hwang, S., Jeong, J., & Seong, C. (2019). Uncertainty in hydrological analysis of climate change: multi-parameter vs. multi-GCM ensemble predictions. *Scientific Reports*. doi: 10.1038/s41598-019-41334-7

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... Thépaut, J. (2020, 7). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. doi: 10.1002/qj.3803

Hess, P., Lange, S., Schötz, C., & Boers, N. (2023, 10). Deep Learning for Bias-Correcting CMIP6-Class Earth System Models. *Earth's Future*, *11*(10). doi: 10.1029/2023EF004002

Jennings, K. S., Winchell, T. S., Livneh, B., & Molotch, N. P. (2018, 3). Spatial variation of the rain–snow temperature threshold across the Northern Hemisphere. *Nature Communications*, *9*(1), 1148. doi: 10.1038/s41467-018-03629-7

Knoben, W. J., Woods, R. A., & Freer, J. E. (2018). A Quantitative Hydrological Climate Classification Evaluated With Independent Streamflow Data. *Water Resources Research*. doi: 10.1029/2018WR022913

Kwon, S., Kim, J., Boo, K., Shim, S., Kim, Y., & Byun, Y. (2019, 3). Performance-based projection of the climate-change effects on precipitation extremes in East Asia using two metrics. *International Journal of Climatology*, *39*(4), 2324–2335. doi: 10.1002/joc.5954

Lafon, T., Dadson, S., Buys, G., & Prudhomme, C. (2013, 5). Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. *International Journal of Climatology*, *33*(6), 1367–1381. doi: 10.1002/joc.3518

Maraun, D. (2016, 12). Bias Correcting Climate Change Simulations - a Critical Review. *Current Climate Change Reports*, *2*(4), 211–220. doi: 10.1007/s40641-016-0050-x

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., ... Thiele-Eich, I. (2010, 9). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, *48*(3), RG3003. doi: 10.1029/2009RG000314

Matiu, M., & Hanzer, F. (2022, 6). Bias adjustment and downscaling of snow cover fraction projections from regional climate models using remote sensing for the European Alps. *Hydrology and Earth System Sciences*, *26*(12), 3037–3054. doi: 10.5194/hess-26-3037-2022

Maurer, E. P., & Pierce, D. W. (2014, 3). Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean. *Hydrology and Earth System Sciences*, *18*(3), 915–925. doi: 10.5194/hess-18-915-2014

McMillan, H. K., Gnann, S. J., & Araki, R. (2022, 6). Large Scale Evaluation of Relationships Between Hydrologic Signatures and Processes. *Water Resources Research*, *58*(6). doi: 10.1029/2021WR031751

Melsen, L. A., Addor, N., Mizukami, N., Newman, A. J., Torfs, P. J., Clark, M. P., ... Teuling, A. J. (2018). Mapping (dis)agreement in hydrologic projections.

Hydrology and Earth System Sciences. doi: 10.5194/hess-22-1775-2018

Melsen, L. A., Teuling, A. J., Torfs, P. J., Zappa, M., Mizukami, N., Mendoza, P. A., ... Uijlenhoet, R. (2019, 1). Subjective modeling decisions can significantly impact the simulation of flood and drought events. *Journal of Hydrology*, *568*, 1093–1104. doi: 10.1016/J.JHYDROL.2018.11.046

Mendoza, P. A., Clark, M. P., Mizukami, N., Gutmann, E. D., Arnold, J. R., Brekke, L. D., & Rajagopalan, B. (2016). How do hydrologic modeling decisions affect the portrayal of climate change impacts? *Hydrological Processes*. doi: 10.1002/hyp.10684

Meyer, J., Kohn, I., Stahl, K., Hakala, K., Seibert, J., & Cannon, A. J. (2019, 3). Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments. *Hydrology and Earth System Sciences*, *23*(3), 1339–1354. doi: 10.5194/hess-23-1339-2019

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., ... Sanderson, B. M. (2016, 9). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482. doi: 10.5194/gmd-9-3461-2016

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005, 3). Which potential evapotranspiration input for a lumped rainfall–runoff model? *Journal of Hydrology*, *303*(1-4), 290–306. doi: 10.1016/j.jhydrol.2004.08.026

Pierce, D. W., Cayan, D. R., Maurer, E. P., Abatzoglou, J. T., & Hegewisch, K. C. (2015, 12). Improved Bias Correction Techniques for Hydrological Simulations of Climate Change*. *Journal of Hydrometeorology*, *16*(6), 2421–2442. doi: 10.1175/JHM-D-14-0236.1

Rastogi, D., Kao, S., & Ashfaq, M. (2022, 8). How May the Choice of Downscaling Techniques and Meteorological Reference Observations Affect Future Hydroclimate Projections? *Earth's Future*, *10*(8). doi: 10.1029/2022EF002734

Reiter, P., Gutjahr, O., Schefczyk, L., Heinemann, G., & Casper, M. (2018, 3). Does applying quantile mapping to subsamples improve the bias correction of daily precipitation? *International Journal of Climatology*, *38*(4), 1623–1633. doi: 10.1002/joc.5283

Ruffault, J., Martin-StPaul, N. K., Duffet, C., Goge, F., & Mouillot, F. (2014, 7). Projecting future drought in Mediterranean forests: bias correction of climate models matters! *Theoretical and Applied Climatology*, *117*(1-2), 113–122. doi: 10.1007/s00704-013-0992-z

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., & Carrillo, G. (2011). Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrology and Earth System Sciences Discussions*. doi: 10.5194/hessd-8-4495-2011

Sepúlveda, U. M., Mendoza, P. A., Mizukami, N., & Newman, A. J. (2022, 7). Revisiting parameter sensitivities in the variable infiltration capacity model across a hydroclimatic gradient. *Hydrology and Earth System Sciences*, *26*(13), 3419–3445. doi: 10.5194/hess-26-3419-2022

Stoner, A. M., Hayhoe, K., Yang, X., & Wuebbles, D. J. (2013). An asynchronous regional regression model for statistical downscaling of daily climate variables. *International Journal of Climatology*. doi: 10.1002/joc.3603

Switanek, M. B., Troch, P. A., Castro, C. L., Leuprecht, A., Chang, H.-I., Mukherjee, R., & Demaria, E. M. C. (2017, 6). Scaled distribution mapping: a bias correction method that preserves raw climate model projected changes. *Hydrology and Earth System Sciences*, *21*(6), 2649–2666. doi: 10.5194/hess-21-2649-2017

Taylor, K. E. (2001, 4). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, *106*(D7), 7183–7192. doi: 10.1029/2000JD900719

Teng, J., Potter, N. J., Chiew, F. H. S., Zhang, L., Wang, B., Vaze, J., & Evans, J. P. (2015, 2). How does bias correction of regional climate model precipitation affect modelled runoff? *Hydrology and Earth System Sciences*, *19*(2), 711–728. doi: 10.5194/hess-19-711-2015

Teutschbein, C., & Seibert, J. (2010, 7). Regional Climate Models for Hydrological Impact Studies at the Catchment Scale: A Review of Recent Modeling Strategies. *Geography Compass*, *4*(7), 834–860. doi: 10.1111/j.1749-8198.2010.00357.x

Vano, J. A., Kim, J. B., Rupp, D. E., & Mote, P. W. (2015). Selecting climate change scenarios using impact-relevant sensitivities. *Geophysical Research Letters*. doi: 10.1002/2015GL063208

Vásquez, N., Cepeda, J., Gómez, T., Mendoza, P. A., Lagos, M., Boisier, J. P., ... Vargas, X. (2021). Catchment-Scale Natural Water Balance in Chile. In (pp. 189–208). Retrieved from `http://link.springer.com/10.1007/978-3-030-56901-3_9` doi: 10.1007/978-3-030-56901-3{_}9

Vicuña, S., Vargas, X., Boisier, J. P., Mendoza, P. A., Gómez, T., Vásquez, N., & Cepeda, J. (2021). Impacts of Climate Change on Water Resources in Chile. In B. Fernández & J. Gironás (Eds.), *Water resources of chile* (pp. 347–363). Cham: Springer International Publishing. Retrieved from `https://doi.org/10.1007/978-3-030-56901-3_19` doi: 10.1007/978-3-030-56901-3{_}19

Vogel, E., Johnson, F., Marshall, L., Bende-Michl, U., Wilson, L., Peter, J. R., ... Duong, V. C. (2023, 7). An evaluation framework for downscaling and bias correction in climate change impact studies. *Journal of Hydrology*, *622*, 129693. doi: 10.1016/j.jhydrol.2023.129693

Vrac, M., & Thao, S. (2020, 11). R2D2; v2.0: accounting for temporal dependences in multivariate bias correction via analogue rank resampling. *Geoscientific Model Development*, *13*(11), 5367–5387. doi: 10.5194/gmd-13-5367-2020

Wan, Z. (2014, 1). New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote Sensing of Environment*, *140*, 36–45. doi: 10.1016/j.rse.2013.08.027

Werner, A. T., & Cannon, A. J. (2016, 4). Hydrologic extremes – an intercomparison of multiple gridded statistical downscaling methods. *Hydrology and Earth System Sciences*, *20*(4), 1483–1508. doi: 10.5194/hess-20-1483-2016

Widmann, M., Bretherton, C. S., & Salathé, E. P. (2003, 3). Statistical Precipitation Downscaling over the Northwestern United States Using Numerically Simulated Precipitation as a Predictor*. *Journal of Climate*, *16*(5), 799–816. doi: 10.1175/1520-0442(2003)016⟨0799:SPDOTN⟩2.0.CO;2

Wilby, R. L., & Dessai, S. (2010). Robust adaptation to climate change. *Weather*. doi: 10.1002/wea.543

Woods, R. A. (2009, 10). Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks. *Advances in Water Resources*, *32*(10), 1465–1481. doi: 10.1016/j.advwatres.2009.06.011

Wootten, A. M., Dixon, K. W., Adams-Smith, D. J., & McPherson, R. A. (2021, 2). Statistically downscaled precipitation sensitivity to gridded observation data and downscaling technique. *International Journal of Climatology*, *41*(2), 980–1001. doi: 10.1002/joc.6716

Wu, Y., Miao, C., Fan, X., Gou, J., Zhang, Q., & Zheng, H. (2022, 11). Quantifying the Uncertainty Sources of Future Climate Projections and Narrowing Uncertainties With Bias Correction Techniques. *Earth's Future*, *10*(11). doi: 10.1029/2022EF002963

Xavier, A. C. F., Martins, L. L., Rudke, A. P., de Morais, M. V. B., Martins, J. A., & Blain, G. C. (2022, 1). Evaluation of Quantile Delta Mapping as a bias-correction method in maximum rainfall dataset from downscaled models in São Paulo state (Brazil). *International Journal of Climatology*, *42*(1), 175–190. doi: 10.1002/joc.7238

Pitfalls in using statistical bias-correction methods to characterize climate change impacts

Nicolás A. Vásquez[1], Pablo A. Mendoza[1,2], Wouter J. M. Knoben[3*], Louise Arnal[3†], Miguel Lagos-Zúñiga[4], Martyn Clark[3‡], Ximena Vargas[1]

[1]Department of Civil Engineering, Universidad de Chile
[2]Advanced Mining Technology Center, Universidad de Chile
[3]Centre for Hydrology, University of Saskatchewan, Canmore, Alberta, Canada
[4]Center for Climate and Resilience Research, Universidad de Chile

Key Points:

- The choice of temporal stratification for GCM bias correction is crucial for removing biases, even for GCMs with good raw seasonality.
- Different temporal stratifications used for GCM bias correction may yield different future seasonalities and signals in projected changes.
- The scaling factor method is effective to assess if the temporal stratification affects the precipitation seasonality projected by a GCM.

[*]now at the Department of Civil Engineering, Schulich School of Engineering, University of Calgary, Calgary, Alberta, Canada
[†]now at Ouranos, Montreal, Quebec, Canada
[‡]now at the Department of Civil Engineering, Schulich School of Engineering, University of Calgary, Calgary, Alberta, Canada

Corresponding author: Nicolás A. Vásquez, nicolas.vasquez.pl@uchile.cl

Abstract

Characterizing climate change impacts on water resources typically relies on Global Climate Model (GCM) outputs that are bias-corrected using observational datasets. In this process, two pivotal decisions are (i) the Bias Correction Method (BCM) and (ii) how to handle the historically observed time series, which can be used as a continuous whole (i.e., without dividing it into sub-periods), or partitioned into monthly, seasonal (e.g., three months), or any other temporal stratification (TS). Here, we examine how the interplay between the choice of BCM, TS, and the raw GCM seasonality may affect historical portrayals and projected changes. To this end, we use outputs from 29 GCMs belonging to the CMIP6 under the Shared Socioeconomic Pathway 5–8.5 scenario, using seven BCMs and three TSs (entire period, seasonal, and monthly). The results show that the effectiveness of BCMs in removing biases can vary depending on the TS and climate indices analyzed. Further, the choice of BCM and TS may yield different projected change signals and seasonality (especially for precipitation), even for climate models with low bias and a reasonable representation of precipitation seasonality during a reference period. Because some BCMs may be computationally expensive, we recommend using the linear scaling method as a diagnostics tool to assess how the choice of TS may affect the projected precipitation seasonality of a specific GCM. More generally, the results presented here unveil trade-offs in the way BCMs are applied, regardless of the climate regime, urging the hydroclimate community for a careful implementation of these techniques.

Plain Language Summary

Global Climate Models (GCMs) are useful tools to characterize the historical and future evolution of the Earth's climate and its impacts on water resources. Because these models contain errors and their horizontal resolution is too coarse for local impact assessments, spatial downscaling and bias correction are required steps. In particular, bias correction methods can be trained and applied using all the available historical data or by splitting the time series (e.g., by season or months). Since there is no guideline on selecting a temporal stratification, we analyze bias-corrected GCM outputs obtained with three types of strategy (entire period, seasons, and months) and seven bias-correction techniques over continental Chile. We show that the choice of bias correction method and the temporal stratification applied can modify the projected precipitation signal and seasonality. We also propose a simple statistical technique to identify if, for a given climate model, the temporal stratification may be a relevant decision for climate impact assessments.

1 Introduction

Understanding and quantifying climate change impacts is crucial for long-term water resources planning and management. Such characterization typically involves hydrologic model simulations forced by an ensemble of scenario-driven meteorological time series obtained from Statistically Downscaled Bias-Corrected (SDBC) Global Climate Model (GCM) outputs (e.g., Addor et al., 2014; Hattermann et al., 2018; Her et al., 2019; Chen et al., 2021; Hanus et al., 2021; Vicuña et al., 2021). This approach usually requires the choice of emission scenario (e.g., Vano et al., 2015; Chegwidden et al., 2019), the choice of GCM (e.g., Hakala et al., 2018; Di Virgilio et al., 2022), the selection of Bias Correction Method (BCM) (e.g., Werner & Cannon, 2016; Gutiérrez et al., 2019; Hess et al., 2023), and the choice of observational (or reference) dataset (e.g., Wootten et al., 2021; Rastogi et al., 2022).

Among the above decisions, the selection and configuration of BCMs is a critical step given the risk of introducing artificial perturbations in GCM outputs (Hagemann et al., 2011; Maurer & Pierce, 2014; Wootten et al., 2021), generating a mismatch between simulated (i.e., obtained from bias-corrected GCMs) and observed (i.e., obtained from a reference dataset) annual cycles of climate variables (e.g., precipitation; Teutschbein & Seibert, 2010; Alder & Hostetler, 2019; Chen et al., 2021), with potential effects on projected climate change impacts and subsequent interpretations and adaptation strategies. A somewhat overlooked step is the strategy for handling the time series when applying BCMs, hereafter referred to as temporal stratification (TS). For example, the bias correction of simulated daily time series can be performed using all the historical period (i.e., a single application of the BCM; e.g., Ghimire et al., 2019) or sub-periods of the historical time series, such as seasons (e.g., four applications of the BCM; e.g., Ruffault et al., 2014; Teng et al., 2015), months (i.e., twelve applications of the BCM; e.g., Pierce et al., 2015; Switanek et al., 2017; Matiu & Hanzer, 2022; Wu et al., 2022; J. Guo et al., 2023), or any other temporal window (e.g., Haerter et al., 2011; Reiter et al., 2018).

Despite the large body of work exploring modeling decisions at the top of the 'cascade of uncertainty' (Wilby & Dessai, 2010), climate impact studies have typically relied on subjectively selected TSs. For example, Teng et al. (2015) compared four BCMs (applied with a seasonal TS) for hydrological projections in southeastern Australia, concluding that the hydrological model amplifies biases in precipitation after applying the BCMs, and that the large spread in the projected signal of changes in precipitation extremes yields different impacts on runoff. Hakala et al. (2018) applied the quantile mapping (QM) method (using a seasonal TS) to assess whether a hydrological model, forced by SDBC GCMs, can replicate the hydrological climatology observed during a historical reference period, obtaining that, even after bias correction, biases in precipitation and streamflow seasonality persist. To analyze the effects of different observational datasets and BCMs on climate projections, Wootten et al. (2021) used three observational datasets to apply two BCMs: (i) the 'Delta' approach with a 3-month moving window, and (ii) the quantile delta mapping (QDM) method over four periods consisting of three non-overlapping months. They concluded that the selection of BCMs and observational datasets have different impacts on historical and projected time series for different variables, although they did not isolate the effect of the TS.

Other studies have focused on the ability of different BCMs to reproduce historically observed climate indices (e.g., Gutmann et al., 2014; François et al., 2020; Xavier et al., 2022), or the effects on climate projections (e.g., Maurer & Pierce, 2014; Melsen et al., 2018), without emphasizing the role of the TS and the evaluation timescale. More recently, Vogel et al. (2023) proposed a framework to evaluate downscaling and BCMs for climate change studies and demonstrated it over Australia using four GCMs, three BCMs and two downscaling methods, considering different TS (monthly, 3-month, and multi-time scales) for the BCMs. They suggested that the TS may influence the analysis (after bias correction) and should be adequately chosen after a careful bias assessment.

Although the preceding studies have covered domains with specific climate types, the trade-offs in selecting TS, BCMs, and GCMs for estimating historical biases (after applying BCMs) and projections across contrasting climates remain unclear. Hence, this paper seeks to disentangle the relative contribution of these decisions (especially TS) to the spread of bias-corrected time series at the annual, seasonal, and monthly timescales during historical and future periods rather than finding the 'best' configuration for the assessment of climate change impacts. Specifically, we address the following research questions:

1. To what extent does the choice of bias correction method and temporal stratification alter historical GCM simulations across different climate regions?
2. What are the effects of bias correction methods and temporal stratification on the projected signal and seasonality of different climate variables?

3. Are there any connections between the effects of TS (on historical biases and projections) and the capability of raw GCM output to replicate historically observed climatology?

To seek answers, we evaluate the performance of 29 SDBC GCMs from the sixth phase of the Coupled Models Intercomparison Project (CMIP6; O'Neill et al., 2016) over different climate groups in continental Chile. We use seven methods (three univariate and four multivariate) to correct biases in precipitation and maximum and minimum temperature. All BCMs are applied at three different TSs: (i) using the entire period (i.e., all daily data simultaneously used for one application of the BCM), (ii) seasonally (i.e., four applications of the BCM using four seasonally stratified time series), and (iii) monthly (i.e., twelve applications of the BCM for twelve monthly stratified time series).

2 Study area and datasets

2.1 Study area

Our study domain is continental Chile, which is suitable for a comprehensive assessment of the TS-BCM-GCM interplay in very different climate types. Figure 1 shows the spatial distribution of mean annual precipitation, mean annual temperature, and three climate indices. The snowfall fraction $SF = Sn/P$ (Figure 1d) is the fraction of mean annual precipitation (P, Figure 1b) falling as snow (Sn). The aridity index (Figure 1e) is the ratio between mean annual potential evapotranspiration (PET) and mean annual precipitation. Finally, the precipitation seasonality (p-seasonality, Figure 1f) indicates whether most precipitation falls during winter (negative values) or summer (positive values). In this paper, we use the season names within the context of the Southern Hemisphere (i.e., winter refers to months JJA, while summer to DJF).

In the northern area (17°S-25°S), two main climate zones can be identified: (i) the super-arid coastal area, with very low annual precipitation amounts (<50 mm/yr), and (ii) the Altiplano region, with lower temperatures due to increasing altitude and larger annual precipitation (∼200 mm/yr). The mean annual precipitation increases towards the south, although the Andes Cordillera generates a west-east gradient, with larger precipitation amounts and lower temperatures on the western slopes of the Andes Cordillera compared to the valleys. Moving south from ∼37°S, the altitude of Andean mountains progressively decreases, as well as the contribution of snowmelt to runoff, whereas precipitation increases. South from 45°S, a west-to-east precipitation gradient produces high precipitation amounts on the coast (>2500 mm/yr), whereas a dry climate develops in Patagonia a few kilometers to the east, with decreasing precipitation amounts. In summary: (i) most snowfall occurs in the Andes Cordillera, though snowfall events can also occur in the valleys of Austral Chile (<45°S); (ii) the hydroclimate is water-limited ($PET/P > 1$) in approximately half of the Chilean territory, especially from ∼35°S to the north, whereas the hydroclimate of the south is energy limited ($PET/P < 1$); and (iii) most precipitation in Chile falls during the winter (red color in panel f), being the Altiplano (northern Chile) and Patagonia (∼50-55°S) two notable exceptions. For a more comprehensive review of the climate and weather of Chile, readers are referred to Aceituno et al. (2021) and Vásquez et al. (2021).

2.2 Datasets

We use the gridded meteorological product CR2MET v2.5 (Boisier et al., 2018; DGA, 2022) as the observational baseline (hereafter reference dataset). CR2MET precipitation estimates (pr) are obtained through a combination of (i) logistic regression models and (ii) multiple linear regression models that use ERA5 reanalysis outputs (Hersbach et al., 2020) and geomorphological attributes as predictors and daily precipitation from meteorological stations as predictands. For daily extreme temperatures (tmax and tmin),

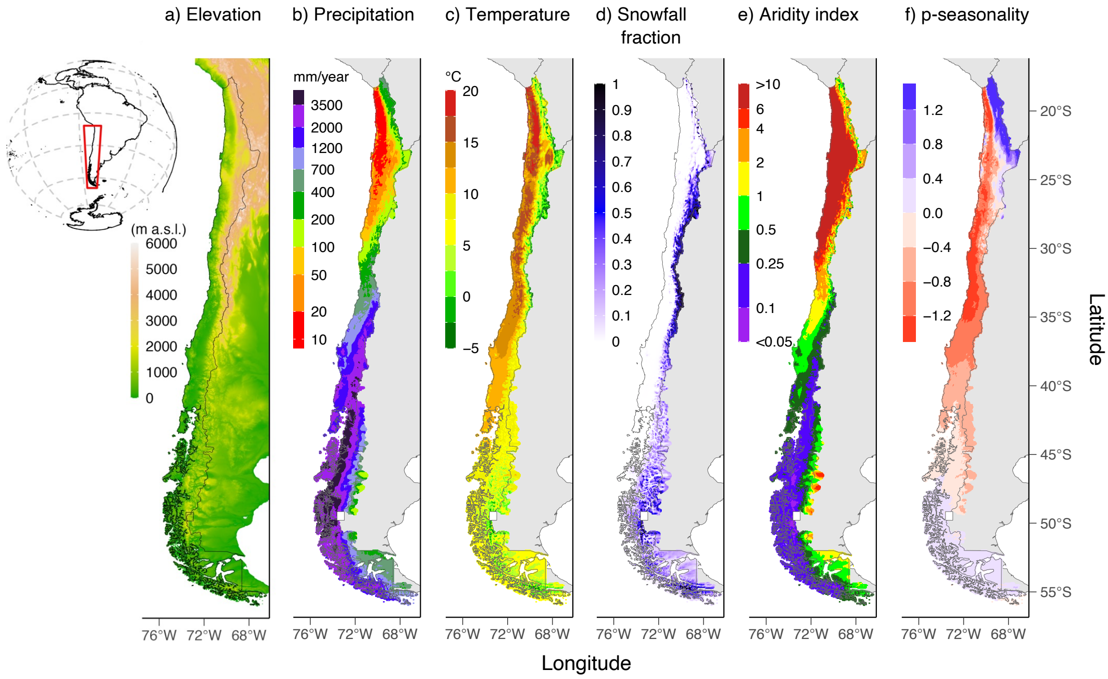

Figure 1. Main physiographic and climate attributes of continental Chile for the period 1980-2014 (34 water years): (a) elevation, (b) mean annual precipitation, (c) mean annual temperature, (d) snowfall fraction, (e) aridity index, and (f) p-seasonality.

land surface temperature from MODIS AQUA and TERRA (Wan, 2014) are also included as predictors. All variables (pr, tmax, and tmin) are available at a daily time step for the period January/1979-March/2020, covering continental Chile at a horizontal resolution of 0.05° x 0.05°. The mean daily temperature is computed as the average between tmax and tmin. It should be noted that CR2MET is, arguably, the most accurate meteorological dataset for continental Chile since its development incorporated local meteorological stations.

We use outputs from 29 GCMs from the CMIP6 (O'Neill et al., 2016), based on the data availability for pr, tmax and tmin during the historical and projected periods, and the SSP5-8.5 scenario for being the worst in terms of greenhouse emissions and the 'business as usual' development case. The name and horizontal resolution of each GCM are included in Table A1.

3 Methodology

Figure 2 shows the main steps of our approach. First, we delineate climate zones across Chile using cluster analysis (step 1), with the aim to examine possible relationships between climate types and the BCM-TS-GCM interplay. Step 2 considers different strategies for correcting biases in GCM outputs (i.e., seven bias-correction methods are applied using three different stratification periods). In step 3, we compute several climate indices derived from precipitation and temperature at different time scales (e.g., annual, seasonal, and monthly mean values), for a historical and a future period. Finally, we conduct an Analysis of Variance (ANOVA) to quantify the relative contribution of different decisions to the spread of historical estimates. More details can be found in the following sections.

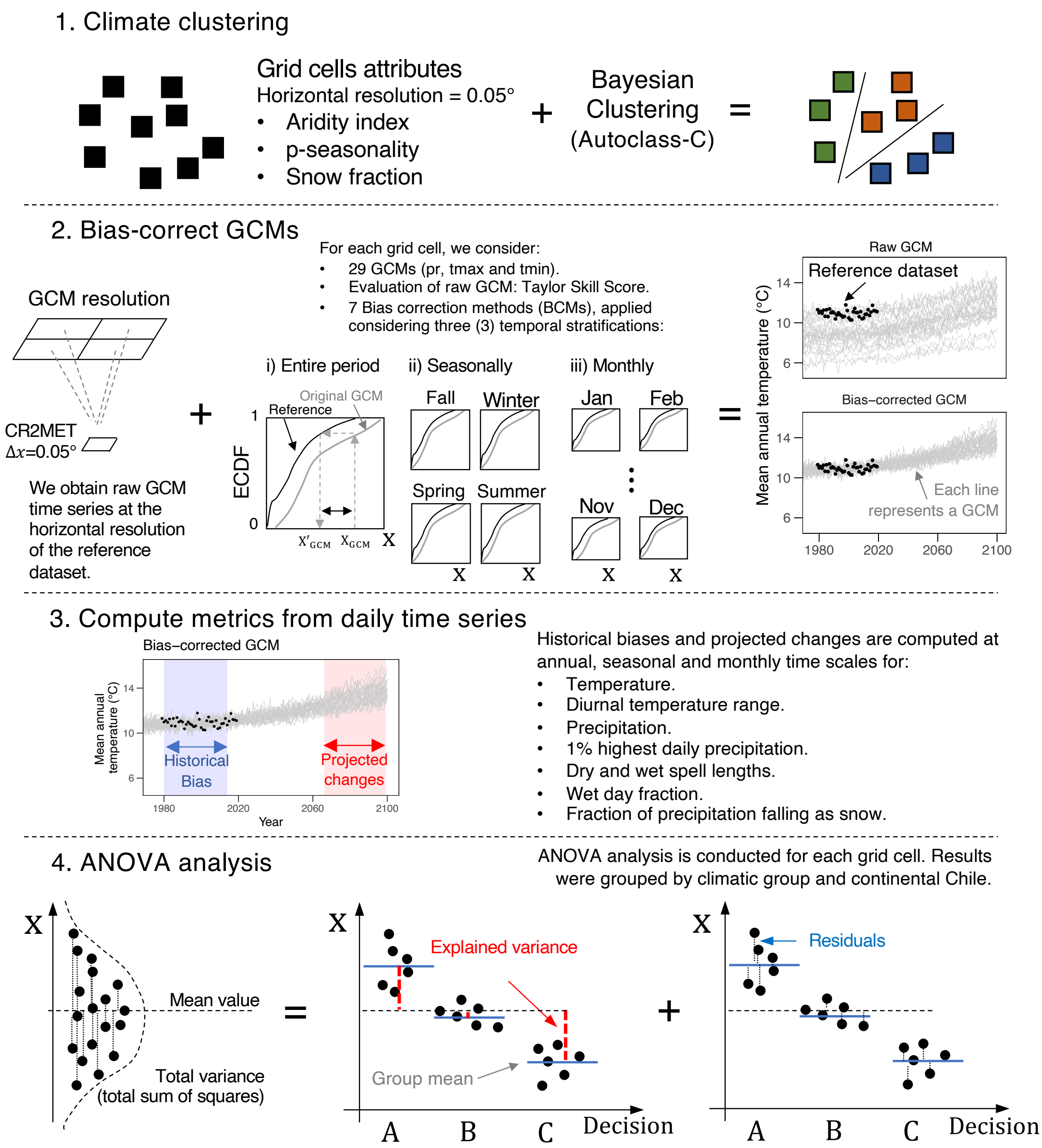

Figure 2. Diagram of the methodology used in this study

3.1 Climate clustering

We perform a Bayesian clustering to identify climate zones across Chile. To this end, we use the aridity index (PET/P), the p-seasonality, and the fraction of precipitation falling as snow as explanatory variables, since they reflect observed hydrological behaviors (Knoben et al., 2018). PET is computed using the Oudin et al. (2005) formula - available in the R Package airGR (Coron et al., 2017) - which requires air temperature (provided at daily time steps here) and latitude as inputs. To estimate Sn, we consider that snowfall occurs when the mean daily temperature is below 2°C (Jennings et al., 2018; Han et al., 2019; Sepúlveda et al., 2022), and p-seasonality is computed with the formula proposed by Woods (2009).
Prior climate groups are defined with the Autoclass-C software (Cheeseman et al., 1988, 1996), which has been previously used in hydrological applications (e.g., Sawicz et al., 2011). We subsequently refined the clustering results through visual inspection, grouping small clusters based on spatial proximity and climate similarity.

3.2 Raw GCM performance

We use the Taylor Skill Score (TSS; Taylor, 2001) to evaluate the role of the raw GCM performance and its interplay with BCM and TS for SDBC-biases and projections at different time scales. The TSS is computed at the grid cell level (0.05° x 0.05°) for the period 1980-2014, contrasting downscaled GCM outputs against the reference dataset, as is commonly done for local climate impact assessments (e.g., Lafon et al., 2013). In this study, TSS is computed for precipitation, as shown in Eq. 1.

$$TSS = \frac{4(1+R)}{\left(\hat{\sigma} + \frac{1}{\hat{\sigma}}\right)^2 (1+R_o)} \tag{1}$$

where R is the Pearson correlation coefficient between the raw GCM and the reference mean seasonality, and $\hat{\sigma} = \sigma_{GCM}/\sigma_{REF}$ is the ratio between the standard deviation of raw monthly values (σ_{GCM}) and the reference (σ_{REF}). R, and $\hat{\sigma}$ are computed using simulated and observed mean monthly values of each variable (i.e., 12 values of GCMs vs. 12 reference values). R_o is the maximum achievable Pearson correlation coefficient for a specific GCM, which is assumed to be $R_o \cong 1$ to simplify the analysis. When $R \to R_o$ and $\hat{\sigma} \to 1$, the $TSS \to 1$. Alternatively, $TSS \to 0$ when R decreases or $\hat{\sigma}$ approaches zero or infinity. Hence, TSS ranges between 0 and 1. Further, we compute the TSS for each climate group, estimating the mean group climatology through spatial averages.

3.3 Bias correction of GCMs

3.3.1 Bias correction methods

We downscale the raw GCM outputs to the CR2MET grid using inverse distance weighting, considering the four closest GCM grid cells. We use seven bias correction methods, including three univariate and four multivariate techniques, listed in Table 1 and briefly reviewed here. The quantile delta mapping (QDM) preserves the projected change for each quantile while correcting the bias. Empirical cumulative density functions are estimated for the historical reference ($F_{h,ref}$), the raw historical GCM ($F_{h,GCM}$), and the raw projected GCM ($F_{p,GCM}$) to relate (X) with the cumulative probability (τ). For a specific value during the historical period $X_{h,GCM}$, the correction (for pr) is given by $X'_{h,GCM} = F^{-1}_{h,ref}\left(F_{h,GCM}\left(X_{h,GCM}\right)\right)$, while for a projected raw GCM value $X_{p,GCM}$, the corrected value is $X'_{p,GCM} = \Delta \cdot F^{-1}_{h,GCM}\left(F_{p,GCM}\left(X_{p,GCM}\right)\right)$, where Δ is computed as $\Delta = X_{p,GCM}/F^{-1}_{h,GCM}\left(F_{p,GCM}\left(X_{p,GCM}\right)\right)$ for precipitation.
The asynchronous regression (AR) relies on a piecewise linear regression calibrated with sorted raw GCM and reference data during a historical period (i.e., $F_{h,ref}$ is a function of $F_{h,GCM}$). Although a simple linear regression could be used, the error in the tails of

the regression can be large and, therefore, the data is split by including different knots (up to six) to reduce errors in low and high values. To bias-correct projected values, the calibrated piecewise linear regression is applied. The quantile regressions neural network (QRNN) uses neural networks to bias correct the sorted data (i.e., quantiles) from simulations and the reference. QRNN is a flexible model since it does not assume a specific relationship between the raw GCM and the reference data.
The rank resampling for distributions and dependences (R^2D^2) corrects the covariance among sites and/or variables through four steps: (i) the univariate bias correction of each variable/site separately, (ii) the selection of one variable/site and the computation of the ranking for all variables/sites, (iii) for a specific date, select the same ranking in the reference period for the dimension selected, and (iv) the shuffling of the other variables/sites to maintain rank structure.
The 'multivariate bias correction' family (MBC) includes three different methods using the Pearson correlation coefficient (MBCp), the Spearman rank correlation coefficient (MBCr), and an N-dimensional probability density function (MBCn) to transform the raw correlated GCM data (i.e., the intervariable dependence structure) through consecutive iterations. For MBCp and MBCr, the transformation relies on the Cholesky matrix decomposition and the correction of the covariance matrix. Conversely, MBCn relies on an orthogonal rotation, the application of QDM to these orthogonal variables, and, finally, the application of an inverse matrix (the one used to compute the orthogonal variables) to obtain the resulting data. The reader is referred to the studies listed in Table 1 for more details on the methods.

Table 1. Methods considered in this study to bias-correct GCMs outputs (pr, tmax, and tmin).

Acronym	Name	Type	Reference
QDM	Quantile Delta Mapping	Univariate	Cannon et al. (2015)
AR	Asynchronous Regression		Dettinger et al. (2004); Stoner et al. (2013)
QRNN	Quantile Regression Neural Network		Cannon (2011)
R^2D^2	Rank Resampling for Distributions and Dependences	Multivariate	Vrac and Thao (2020)
MBCp	Multivariate Bias Correction method - Pearson		Cannon (2016)
MBCr	Multivariate Bias Correction method - Rank		
MBCn	Multivariate Bias Correction method – QDM		Cannon (2018)

We stress that it is not our aim to perform detailed comparisons among different bias correction techniques but to quantify the impact of this and other methodological choices on historical biases and projected changes in climate indices. All bias correction methods were applied using the statistical software 'R' (http://www.r-project.org/). The QDM, MBCp, MBCr, MBCn, and R^2D^2 methods were applied using the library 'MBC' (Cannon, 2018). QRNN was implemented using the 'qrnn' library (also available in R), while the AR method was implemented following Stoner et al. (2013). To reduce the computational effort, we randomly select 100 grid cells within each climate group, and all subsequent analyses are conducted at these grid cells ($100 \cdot N_{clusters}$).

3.3.2 Choice of the temporal stratification

Bias correction methods can be applied using different stratification strategies. For example, a BCM can be applied at daily time steps using all the data in the historical period (usually 30 years), which means that all ∼10,950 days (∼365 days · 30 years) are simultaneously bias-corrected. For a seasonal TS, BCMs are applied four times, each one considering ∼2730 days (∼91 days · 30 years), whereas for a monthly TS, the BCM is applied 12 times considering ∼900 days (∼30 days · 30 years). Note that other temporal stratifications could be considered. Here, we applied BCMs to daily time series of pr, tmax, and tmin (e.g., Rastogi et al., 2022) using the entire time series in the historical period (1980-2014), and stratifying the data seasonally and monthly, since these TSs are typically considered for climate change impact assessments. For all combinations of BCM and TS, we obtained daily time series from 1980 to 2100.

3.4 Climate indices

We consider several climate indices that are relevant to reproduce historically observed hydrological responses (e.g., Gutmann et al., 2014), including (i) mean annual, seasonal, and monthly total precipitation, (ii) highest 1% daily precipitation, (iii), wet-day fraction, (iv) wet and dry-spell lengths, (v) fraction of precipitation falling as snow, and (vi) annual, seasonal and monthly averages of mean daily temperature and diurnal temperature ranges. To estimate the mean annual snowfall, we add all precipitation amounts for days with a mean daily temperature below 2°C. Wet-spell and dry-spell lengths (mean consecutive rainy and non-rainy days, respectively), as well as the wet-day fraction (mean fraction of rainy days) are computed as in Gutmann et al. (2014), considering 0.1 mm/d as a threshold. To examine the capability of BCMs to replicate historically observed climate indices, we computed the difference between SDBC-GCM outputs and the reference dataset during the historical period 1980-2014 as a percent bias (hereafter referred to as biases). Additionally, we analyze the effects of BCMs on climate projections by computing the relative change for the period 2065-2099 with respect to the historical period (1980-2014).

3.5 Analysis of Variance

To evaluate the relative contribution of the BCM and TS decisions to the spread of SDBC-biases we perform, for each combination of GCM and grid cell, an analysis of variance (ANOVA). In this case, the ANOVA is simplified as:

$$TV = BCM + AP + Residual \tag{2}$$

where TV stands for the total variance of SDBC-biases, and the residual term is the variance not explained by the BCM nor the TS for a specific GCM-grid cell combination. If the choice of TS had no impact on the biases in climate indices. In that case, the application of Supposey BCM should be able to reduce biases at all temporal scales (e.g., annual, seasonal, or monthly), regardless of the GCM considered. To summarize the information at the grid cell level, we compute the average of BCM/TV, TS/TV, and $Residual/TV$ fractions across GCMs, whereas for the climate groups, we compute the mean relative contribution (estimated by BCM/TV, TS/TV and $Residual/TV$) of TS and BCM to the spread as the average of fractions across the grid cells within that group.

4 Results

We show the climate clustering results, the historical biases after applying the BCMs, and the relative contributions of different methodological choices to historical biases of climate indices at the annual and seasonal scales. Further, we include the TSS performance to examine connections between the raw seasonality of the GCMs and the selection of BCM and TS. For simplicity, we only show the results for precipitation, and the remaining variables can be found in the Supporting Information.

4.1 Clustering

The Bayesian clustering and subsequent spatial aggregation through visual inspection provided ten climate groups for continental Chile (Figure 3). In general, the clusters follow two main climate patterns in Chile: (i) a latitudinal precipitation gradient, from very arid (north) to humid (south), and (ii) a west-east gradient from the coast to the Andes Cordillera. Although northern Chile encloses groups 1, 2, and 3, clusters 2 and 3 are located in the Altiplano region, where larger precipitation and lower temperatures are observed. Groups 5, 6, and 8 span the coast and valley, whereas groups 4 and 7 are located in the Andes. Finally, groups 9 (the rainiest group) and 10 are in southern Chile, characterized by large precipitation amounts in the Andes Cordillera and the coast, with decreasing precipitation and temperature towards the east (Patagonia).

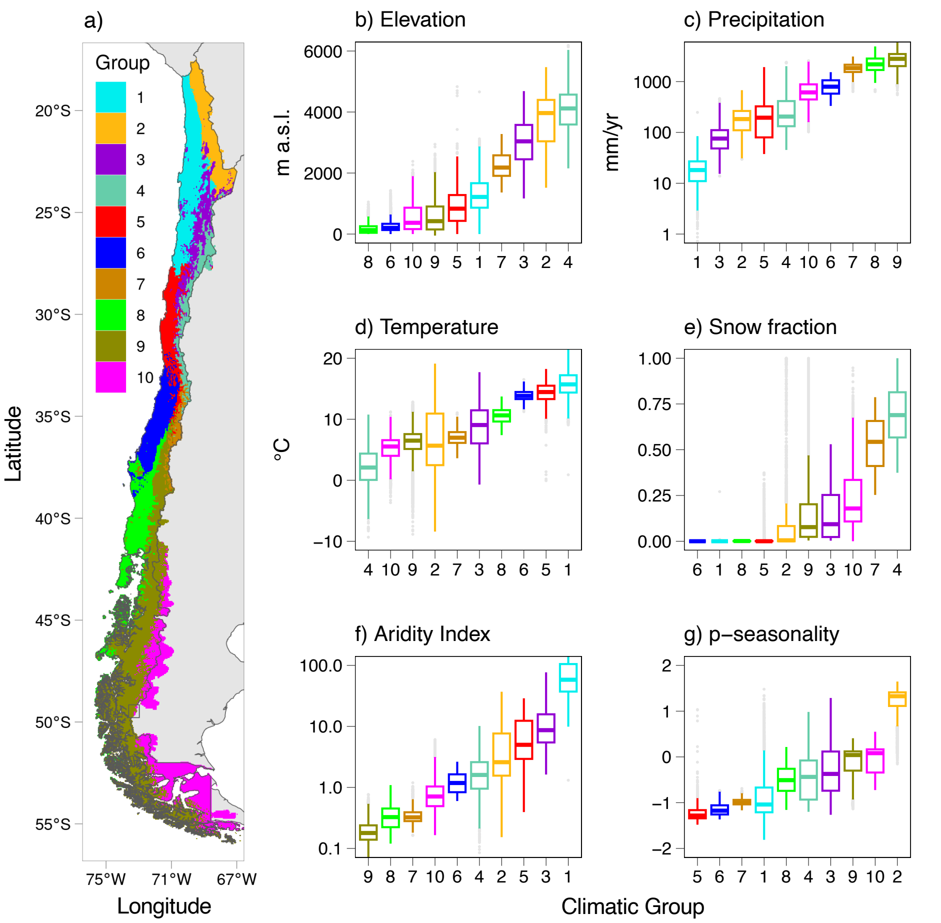

Figure 3. (a) Spatial distribution of climate clusters in continental Chile based on snowfall fraction, aridity index, and p-seasonality. The following attributes are ordered by the median of each group: (b) elevation, (c) precipitation, (d) temperature, (e) snowfall fraction, (f) aridity index, and (g) p-seasonality. All climate indices were computed for the period 1980-2014. Notice that the boxplots in panels b-g are sorted according to the median value, and the group's order on the x-axis differs among variables.

4.2 Performance metrics after bias correction

Figure 4 shows precipitation biases (after bias correction) in three different climate groups (the other variables and climate groups can be found in the Supporting Information). The results show that, regardless of the combination of GCM, BCM, TS and grid cell, biases in annual amounts are close to zero (Figure 4a). When the BCM is applied using all the data in the historical period(Figure 4b, left), biases in monthly precipitation amounts can be large, although the magnitude varies among climate groups. In climate group 2 (Altiplano region), precipitation occurs mostly during the summer (DJF); in this season, the median bias associated with January precipitation is relatively lower - though still considerable (>20%) - compared to the remaining months. In group 6, most precipitation occurs during the winter (JJA), and biases can be found in any month. In group 10, precipitation falls uniformly throughout the year, with slightly larger amounts and larger biases during the summer (DJF). When the BCM is applied seasonally (4b, center), monthly precipitation biases persist. However, these are generally lower compared to the case when the bias correction is applied using the entire dataset, especially in climate group 10. As expected, biases are nearly removed with a monthly TS (Figure 4b, right), regardless of the GCM, bias correction method, grid cell, or climate group.

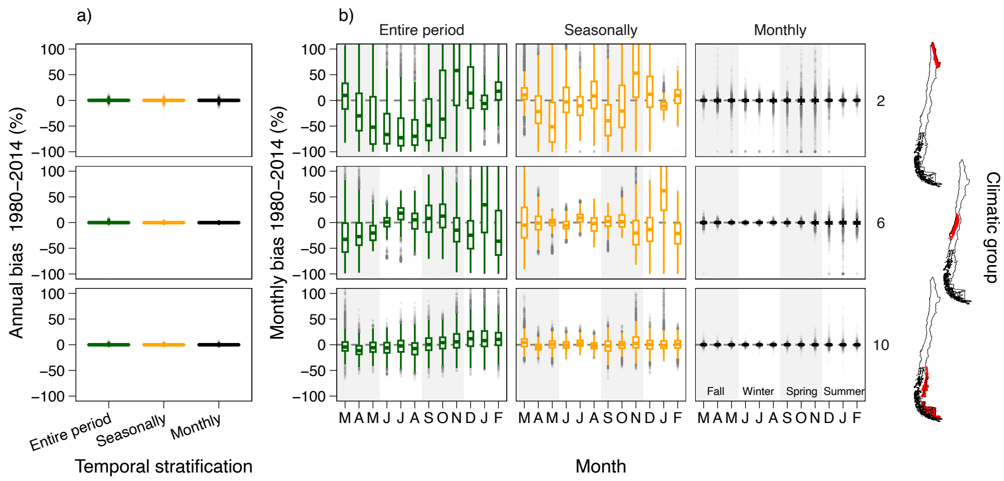

Figure 4. Historical biases in precipitation at the (a) annual and (b) seasonal time scales in three climate groups (rows) after applying the BCMs. The columns in panel b) show results for the three TSs used to apply the BCMs. Each boxplot comprises results from the 100 grid cells within a specific climate group, 29 GCMs, and seven BCMs. The different seasons are highlighted through grey-white areas.

Figure 5 displays the relative contributions of the BCM, TS, and residuals for mean annual, seasonal (summer and winter), and monthly (January and July) precipitation biases averaged across 1,000 grid cells in continental Chile. We show two seasons and months to examine possible differences between the dry and wet seasons. Additionally, the results from different GCMs are stratified according to their historical raw performance, measured by the Taylor Skill Score. As in Figure 4, the ANOVA analysis for historical biases shows differences among temporal stratifications, especially when compared to annual biases (Figure 5a). Because the relative contributions of BCM and TS to precipitation biases do not greatly differ among climate groups, we show results at the national scale. The choice of BCM explains most of the variance for the mean annual precipitation bias, whereas the choice of TS explains almost all the variance for mean seasonal and monthly precipitation biases. It is worth noting that the biases at the annual scale are, in general, very low (Figure 4, <1%), and that the relative importance of the choice of TS for seasonal and monthly biases does not decrease for GCMs with high TSS values. The latter result is counterintuitive since one might expect that GCMs with good raw precipitation seasonality will be effectively bias-corrected, regardless of the TS selected. For variables related to quantiles (highest 1% daily precipitation, dry and wet-spell lengths, and wet-day fraction), the relative importance of BCMs increases for GCMs with higher TSS, being BCM the most important decision, even at seasonally and monthly time scales (Figure S1).

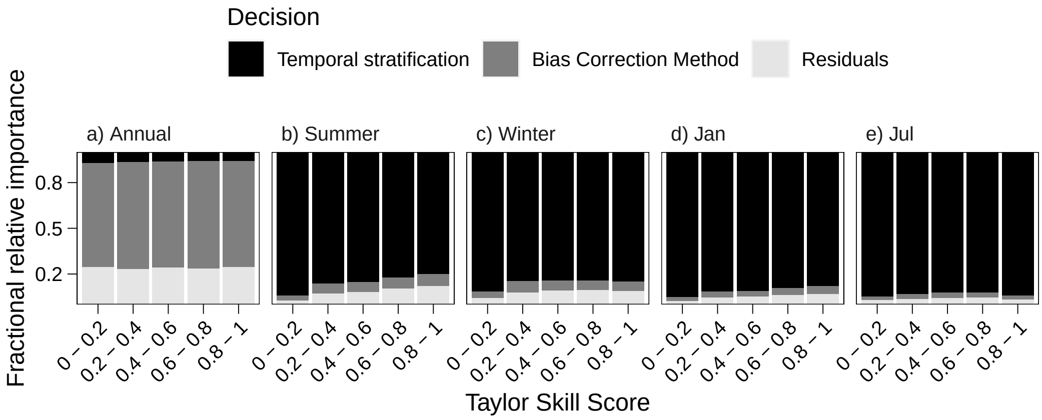

Figure 5. Relative importance (as a fraction averaged from all grid cells and GCMs for continental Chile) of the bias correction method and the temporal stratification to explain the precipitation biases at the annual, seasonal (DJF and JJA), and monthly (January and July) time scales during the historical period (1980-2014), for different levels of historical GCM performance (x-axis). Biases are computed after applying BCMs.

4.3 Projected changes

We now analyze the interplay between the choice of TS, the raw GCM precipitation seasonality, and its effects on projected changes in precipitation for the period 2065-2099 (with respect to 1980-2014) at different time scales. Figure 6 displays projected changes in mean annual, seasonal, and monthly precipitation for one grid cell located in central Chile (red dot in map) and one GCM (INM-CM4-8) with a high R value. For this GCM and grid cell, $TSS = 0.76$ during the period 1980-2014, with a Pearson correlation coefficient between mean monthly raw GCM and reference amounts of 0.98, and a 41% underestimation of the standard deviation. The high value of R indicates a good seasonality of raw GCM outputs. Figure 6 shows that different BCMs yield a high dispersion in projected changes of mean annual precipitation (different lines), with little influence on the selected TS (x-axis of each subplot). Additionally, all BCMs alter the raw GCM projection. For example, if all BCMs are applied using the entire dataset, projected changes in summer precipitation range between -8% to 5%, whereas the raw projection is close to -30%. The application of MBCn using the entire period yields a positive projected change in the mean summer precipitation, while a seasonal and monthly application of the same BCM projects a decrease in summer precipitation. The results for individual months (January and July) reveal more dispersion and interaction among BCMs and the choice of TS. For example, applying the BCM with the entire time series results in positive and negative projections of mean July precipitation (the rainiest month for this grid cell). Similarly, different TSs can also provide different projected signals.

Figure 6 reveals that the choice of TS affects the signal of projected changes in summer precipitation (e.g., for the MBCn method) and, in particular, in January and July precipitation amounts. The TS can be considered relevant for a specific grid cell if it is able to switch the projected signal of a variable for a particular GCM-BCM combination. This is, for example, the case of mean July precipitation (Figure 6), for which the

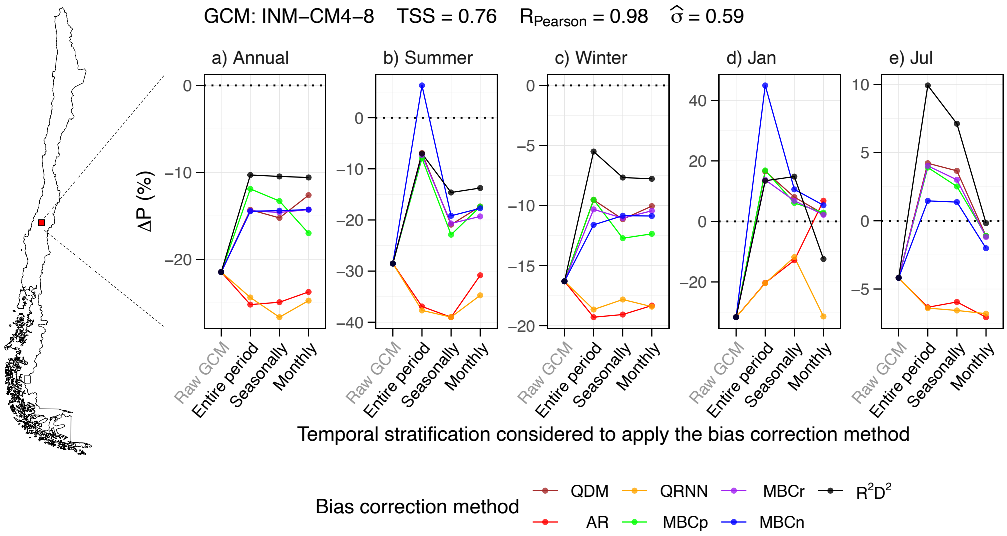

Figure 6. Projected change in annual, seasonal (summer and winter), and monthly (January and July) precipitation for different temporal stratifications (x-axis) and bias correction methods (lines). All combinations of TS and BCM decisions, along with projected changes from the raw (biased) GCMs, are displayed. The results are valid only for the grid cell shown and the GCM INM-CM4-8. The metrics (e.g., TSS) were computed using the raw (biased) GCM data for the period 1980-2014.

signal of projected changes is different among TSs for the MBCn, MBCr, and R^2D^2 methods.

Figure 7 shows, for all the grid cells analyzed, the fraction of 'well-behaved' GCMs (i.e., with $TSS \geq 0.7$; e.g., Kwon et al., 2019) for which the selection of TS leads to different signs in projected precipitation changes. Note that the number of GCMs that meet the performance requirement - obtained by spatially averaging the number of GCMs with $TSS \geq 0.7$ at each latitudinal band - varies along the domain. In general, the choice of TS does not alter the signal of projected changes in mean annual precipitation, although a few GCMs are affected by this decision in some areas (e.g., northern Chile). Nevertheless, the effects of TS are more evident in seasonal projections (Figure 7b and 7c). During the summer, >50% of the number of GCMs are affected by the TS in Central Chile (dry season). During winter, the Altiplano region and part of southern Chile are largely influenced by the choice of TS. It should be noted, however, that the summer season in Central Chile and the winter season in the Altiplano region are dry seasons. Therefore, while the signal of projected changes may vary for different TSs, the precipitation amounts involved are small. For mean monthly January and July precipitation, the choice of TS is even more relevant. Indeed, nearly all GCMs are affected by the TS along the coast of northern Chile, while ∼50% of the GCMs yield different signals in projected changes for different TSs in Central Chile. The case of July is more interesting since it is the rainiest month in most of continental Chile. In July, ∼50% of the GCMs are affected by the TS along the Central Chilean Andes (western border), impacting the accumulation of snow and, therefore, meltwater volume and timing estimates for the spring and summer seasons. In southern Chile, one can find grid cells where GCMs are affected by the TS decision, though that fraction is lower compared to the Central Chilean Andes.

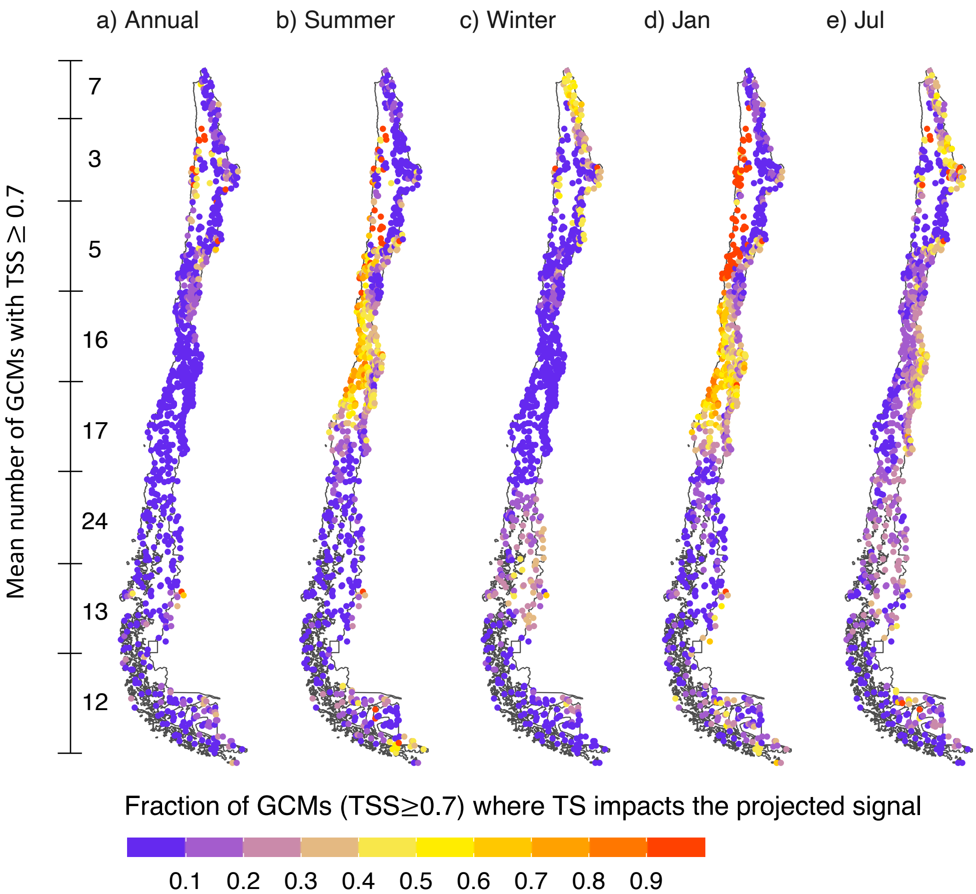

Figure 7. Fraction of GCMs with acceptable performance (i.e., with $TSS \geq 0.7$) for which the TS yields different projected precipitation signals. The number of GCMs that meet the threshold criteria at each ∼5° latitudinal band is computed as the average of GCMs with $TSS \geq 0.7$ from all grid cells within that band.

Figure 8a compares the raw GCM output (obtained from the GCM ACCESS-CM2) and the reference precipitation seasonality over a historical period at one grid cell located in central-southern Chile (red dot on the map). For this GCM-grid cell combination, $TSS = 0.96$, $R = 0.94$ and $\hat{\sigma} = 1.08$. Note that the GCM simulates the maximum monthly precipitation in July instead of June (when the maximum occurs according to the reference). Figure 8b displays, for the same GCM-grid cell, the projected precipitation seasonality for each BCM-TS combination (thin lighter lines). The results show that applying a BCM using the entire period (green lines) provides the same seasonality as the raw GCM; however, seasonal and monthly TSs distort the raw projected seasonality. Further, when BCMs are applied using a monthly TS (black/gray lines), the projected month of maximum precipitation is June, whereas for seasonal and entire period such month is July. Additionally, seasonal and monthly TSs yield higher precipitation fractions (compared to the raw GCM) during April and May, and smaller values during September and October. Such differences in projected precipitation seasonality may affect any subsequent analyses of simulated hydrological fluxes and states.

To examine the extent to which projected precipitation seasonality is affected by the temporal stratification, we focus on the projected maximum mean monthly precipitation. Hence, we contrast, for each GCM-grid cell combination, three curves obtained with the three temporal stratifications (each obtained by averaging the projections among BCMs for each GCM). We consider that the TS affects the projected seasonality if the month where the maximum mean monthly precipitation amount occurs differs. Conversely, if such a month is the same for the three TSs, we consider that this decision does not impact the seasonality. Figure 8c displays the fraction of the number of GCMs with $TSS \geq 0.7$ for which the TS impacts the projected precipitation seasonality. Interestingly, the number is relatively high (>40%) for most of continental Chile. The fraction of GCMs affected by the TS decision is even higher in northern Chile, the Central Chilean Andes, and the Southernmost part of Chile, where more than 60% of GCMs are affected.

5 Discussion

The results presented here highlight the relevance of the temporal stratification used when applying bias correction techniques, which affects (i) SDBC-biases in seasonal and monthly precipitation amounts over a historical period, and (ii) the signal of projected changes and the seasonality of projections.

5.1 Temporal stratification as a source of uncertainty

Our results show that the temporal stratification can largely affect precipitation biases during a historical period, as well as the signal and seasonality of projected changes. However, this methodological choice has been rarely explored in climate change impact assessments, and the lack of guidance has motivated the use of more than one TS in some studies (e.g., Wootten et al., 2021). Further, model errors may not necessarily be removed in the process. For example, Hakala et al. (2018) obtained that biases in precipitation and streamflow seasonality remained after applying BCMs. Here, we found that only a monthly application of the BCM can replicate the reference precipitation seasonality, even for GCMs with a good raw representation of annual cycles.

5.2 Projected seasonality

Our study reveals that one of the main effects of selecting different TSs is the possibility to distort the precipitation seasonality projected by raw GCM outputs. In hydrologic impact assessments, this artifact may propagate into the timing of simulated variables like snow accumulation and melting, energy fluxes, and streamflow (Meyer et al., 2019). Our results show that when the raw GCM seasonality has timing errors (com-

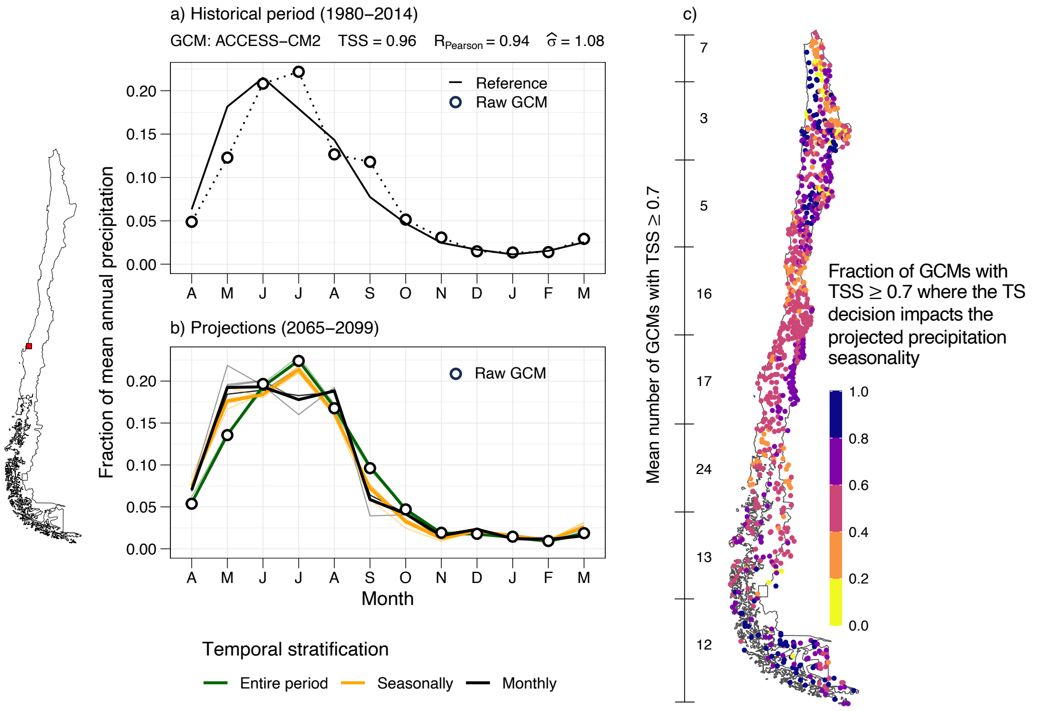

Figure 8. Influence of the temporal stratification used to apply bias correction methods on the projected precipitation seasonality. (a) Dimensionless historical seasonality for one grid cell (red dot on the map) and one GCM (ACCESS-CM2). Note that the sum of monthly fractions is equal to 1. (b) Projected raw (circles) and bias-corrected (colored lines) GCM precipitation seasonality. Lighter and thinner lines represent different BCMs, whereas thick lines represent the average across BCMs. (c) Fraction of the total number GCMs with $TSS \geq 0.7$, for which the temporal stratification yields different projected seasonality, measured as different months for maximum mean monthly precipitation for the 2065-2099 period. In c), the average number of GCMs meeting the TSS criterion is computed for latitudinal bands.

pared to the reference), a pronounced shift in the projected seasonality can be obtained after applying BCMs (compared to the case without bias correction). However, when the raw GCM replicates the historically observed precipitation seasonality reasonably well, one might expect that different TSs yield the same projected seasonality. To test this hypothesis, we compare the precipitation seasonality projected with three TSs (bottom panels) by two GCMs (CanESM5 and NorESM2-MM, Figure 9) that replicate annual cycles (i.e., high Pearson correlation coefficients, with GCM and reference maximum mean monthly precipitation being the same, top panels). For GCM CanESM5 (Figure 9a), the choice of TS has little effect on the projected precipitation seasonality. Conversely, the temporal stratification affects the seasonality projected by NorESM2 (Figure 9b). For example, if the BCM is applied seasonally and monthly, the months of maximum mean monthly precipitation are May and August, respectively. Interestingly, TSS = 0.951 for this GCM, which is higher than the value obtained for CanESM5 (0.694), and both GCMs have similar Pearson correlation coefficients. These results emphasize that even GCMs with a good raw representation of historical seasonality can be affected by the temporal stratification used to apply BCMs.

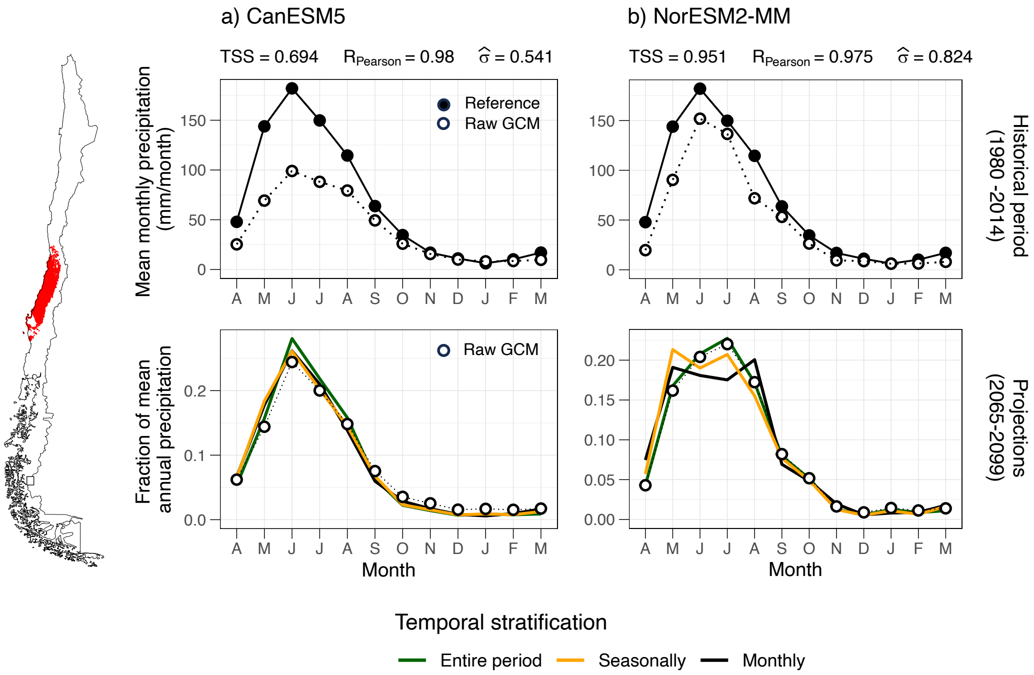

Figure 9. Impact of the temporal stratification used in bias correction for two GCMs. The results presented here are spatially averaged values of the grid cells contained in climate group 6 (highlighted in red on the map). Top row: comparison of the raw GCMs and the reference for the period 1980-2014. Bottom row: projected precipitation seasonality in terms of fraction of mean annual precipitation (average from the seven BCMs).

5.3 A priori evaluation of the TS impact on projected precipitation seasonality

Understanding the potential effects of the TS on the projected signal and seasonality of precipitation from a specific GCM could be helpful for a more detailed assessment of climate change and/or hydrological changes. Here, we propose using the linear scaling method (LSM) (Widmann et al., 2003; Maraun, 2016) - due to its low computational cost and simplicity (Lafon et al., 2013; Chaubey & Mall, 2023) -, as a quick diagnostics tool to inform if the TS may be an influential decision (an example of an LSM application is provided in Appendix B). The LSM removes the bias from the raw GCM time series (f_{bias}) through a multiplicative factor for the case of precipitation and an additive term for temperature, using an observational dataset as a reference. For example, if the reference and raw GCM mean annual precipitation amounts are 500 mm/year and 650 mm/year, respectively, a factor $f_{bias} = 500/650 = 0.77$ is applied to the raw GCM time series to remove the bias. Accordingly, seasonal or monthly applications of LSM require more scaling factors (Maraun et al., 2010). Hence, the raw GCM projected change (f_Δ) is preserved (at the TS time scale), since the scaling factors are typically considered to be time-invariant. Additionally, the influence of the temporal stratification and the reference dataset (in case there is more than one available) can be isolated for a specific grid cell-GCM combination.

Figure 10a illustrates the application of the linear scaling method (dashed lines) to the GFDL-CM4 GCM in one grid cell (red dot in map), using the entire period and stratifying the data seasonally and monthly. For this GCM-grid cell combination, $TSS = 0.72$ and $R = 0.7$, and different TSs yield different projected precipitation seasonalities when applying the LSM. Figure 10a shows that the precipitation factors obtained with LSM agree with the averages obtained from all (seven) bias correction methods (solid lines).

Finally, we examine the capability of the LSM to identify the precipitation seasonality projected with different TSs correctly. To this end we obtain, for each grid cell-GCM-TS combination, the precipitation seasonalities from (i) the average between the seven BCMs, and (ii) the application of the LSM. If the months of the projected maximum precipitation agree, we consider that the LSM correctly identifies the seasonality, and if this occurs for the three TSs, we consider that the LSM successfully identifies the projected bias-corrected seasonality for that specific grid cell-GCM combination. Figure 10a illustrates a successful case since, for each TS, the month of maximum precipitation is the same for the average among seven BCMs and from the LSM. Then we compute, for the 1,000 grid cells analyzed here, the fraction of GCMs for which the LSM successfully identifies the projected seasonality (accuracy, Figure 10b). The results show that, in almost all the grid cells, the LSM successfully identifies the projected seasonality of $\sim$70% of the GCMs, whereas for most grid cells ($> 85\%$), the LSM successfully projects the seasonality for more than 85% of the GCMs.

5.4 Limitations and future work

In this study, we selected the SSP5-8.5 scenario and 29 GCMs, although other future scenarios and/or a subset of GCMs could be considered to assess the effects on historical biases (after bias correction) and/or future projections. We did not focus on performance metrics for specific GCMs because evaluating the adequacy of particular bias correction methods is out of the scope of this work; instead, we focus on how these techniques are traditionally applied. Although we selected univariate and multivariate BCMs (e.g., Q. Guo et al., 2020), quantile-based, neural networks, and linear regressions, different approaches could be considered.

Additionally, we did not conduct any hydrological modeling. Instead, we focused on the repercussions of some decisions on the historical biases and the projected seasonality of climate variables required to run hydrological and land surface models. However, previous work has shown that hydrological models tend to amplify biases in the forcings (Teng et al., 2015). We emphasize that any assessment of climate change impacts should ensure that the climatological annual cycles of hydrological simulations forced with (i)

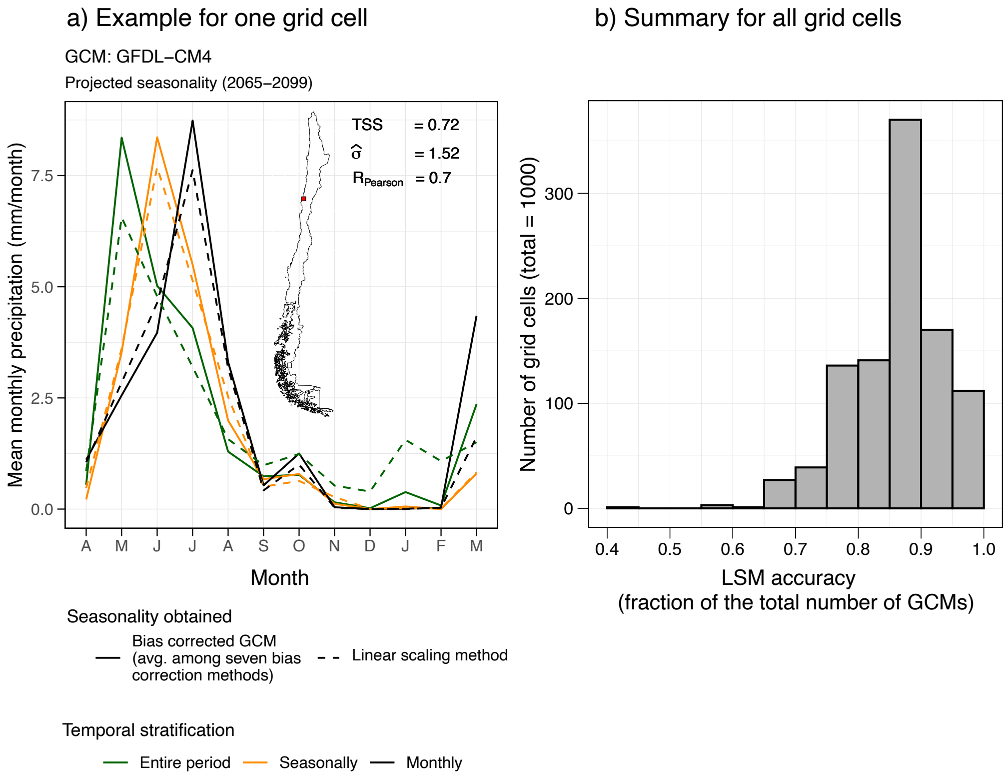

Figure 10. Linear scaling method used as a proxy to estimate the projected precipitation seasonality. (a) Example of projected precipitation seasonalities for one grid cell and one GCM, obtained from applying the LSM and the seven BCMs tested. The metrics summarize the raw (biased) GCM performance for the historical period (1980-2014). (b) LSM accuracy (as a fraction of the total number of GCMs) for all grid cells.

reference data sets and (ii) bias-corrected time series from GCMs/RCMs are similar (Hakala et al., 2018). Hence, verifying the reference and bias-corrected GCM forcing data during a historical period arises as a crucial step (Chen et al., 2013; Clark et al., 2016; Mendoza et al., 2016; Melsen et al., 2019). Future work could consider the impacts of SDBC historical biases and differences in projected seasonality on different aspects of the hydrograph (e.g., mean values, extremes, timing, etc.) and signatures formulated from other variables than streamflow (e.g., SWE, soil moisture; McMillan et al., 2022; Araki et al., 2022).

6 Conclusions

In this paper, we examined how methodological choices involved in GCM bias correction affect historical and future climate portrayals. To this end, we used seven bias correction methods, 29 CMIP6 GCMs, and three temporal stratifications. All the configurations were applied to daily time series of precipitation and maximum and minimum daily temperature derived from the CR2MET gridded observational product, available for continental Chile. Our main findings are as follows:

1. A monthly application of bias correction methods is required to replicate the reference precipitation seasonality, even for GCMs with good raw seasonality.
2. The temporal stratification is the most relevant decision to quantify seasonal and monthly precipitation biases.
3. Different temporal stratifications may yield different projected signals and seasonality, even for GCMs with good raw seasonality.
4. The linear scaling method can be used to estimate the projected seasonality of GCMs and, therefore, to identify the climate models for which the choice of temporal stratification may be critical, before applying more sophisticated and computationally expensive bias correction methods.

Appendix A Selected GCMs

Table A1 shows the GCMs included in this study.

Appendix B Scaling factor example

We illustrate the effects of the temporal stratification by applying the linear scaling method (LSM) (Maraun et al., 2010) for one grid cell-GCM combination. Figure B1a shows monthly precipitation averages from raw GCM outputs, whereas Figure B1b-d shows the bias-corrected GCM values for three different temporal stratifications. Monthly values were obtained from the daily corrected time series.
Note that when the entire period is used to bias-correct the GCM, only one factor is applied. In the grid cell analyzed, the reference annual precipitation is 4371 mm, which is below the historical raw GCM amount for the same period (5020 mm). Hence, the raw GCM precipitation time series is multiplied by the factor $f = 4731/5020 = 0.87$, which removes the annual SDBC bias; nevertheless, monthly SDBC-biases persist (see differences between black and blue lines in Figure B1b). When the LSM is applied seasonally, four factors are used to multiply the raw GCM time series. For example, daily values from March, April, and May are bias-corrected by the seasonal factor obtained from the reference (1134 mm/season) and the raw GCM (1498 mm/season) precipitation amounts. In this case, the factor used to bias-correct daily precipitation from March, April, and May is $f_{MAM} = 1134/1498 = 0.76$. Similarly, if the LSM is applied monthly, daily precipitation amounts from March are bias-corrected using the reference (374 mm/month) and raw GCM (498 mm/month), which yields a factor $f = 374/498 = 0.75$. For the monthly TS, the black and blue lines are the same. Note that the projected maximum

Table A1. GCMs considered in this study

GCM	Δlat	Δlon	Institution
ACCESS-CM2	1.25	1.88	Australian Research Council Centre of Excellence for Climate Science, Australia.
ACCESS-ESM1-5	1.25	1.88	
BCC-CSM2-MR	1.11	1.13	Beijing Climate Center, China.
CanESM5	2.77	2.81	Canadian Centre for Climate Modelling and Analysis, Canada.
CMCC-ESM2	0.94	1.25	Euro-Mediterranean Centre on Climate Change Coupled Climate Model, Italy.
CNRM-CM6-1-HR	0.50	0.50	Centre National de Recherches Météorologiques (CNRM), France.
CNRM-CM6-1	1.40	1.40	
CNRM-ESM2-1	1.40	1.41	
E3SM-1-0	1.00	1.00	Lawrence Livermore National Laboratory, USA.
EC-Earth3-CC	0.70	0.70	EC-Earth Consortium, Europe.
EC-Earth3-Veg-LR	1.12	1.13	
EC-Earth3-Veg	0.70	0.70	
EC-Earth3	0.70	0.70	
FGOALS-g3	2.18	2.00	Chinese Academy of Sciences Flexible Global Ocean-Atmosphere-Land System Model, China.
GFDL-CM4	1.00	1.25	Geophysical Fluid Dynamics Laboratory, USA.
GFDL-ESM4	1.00	1.25	
INM-CM4-8	1.50	2.00	Institute for Numerical Mathematics, Russia.
INM-CM5-0	1.50	2.00	
IPSL-CM6A-LR	1.27	2.50	Institute Pierre Simon Laplace (IPSL), France.
KACE-1-0-G	1.25	1.88	National Institute of Meteorological Sciences (NIMS) and Korea Meteorological Administration (KMA), South Korea.
KIOST-ESM	1.88	1.88	Korea Institute of Ocean Science and Technology Earth System Model and Its Simulation Characteristics, South Korea.
MIROC-ES2L	2.79	2.81	Japan Agency for Marina-Earth Science and Technology (JAMSTEC), Japan.
MIROC6	1.39	1.41	
MPI-ESM1-2-HR	0.93	0.94	Max Planck Institute for Meteorology (MPI-M), Germany.
MPI-ESM1-2-LR	1.87	1.88	
MRI-ESM2-0	1.11	1.13	Meteorological Research Institute, Japan.
NESM3	1.85	1.88	Nanjing University of Information Science and Technology Earth System Model, China.
NorESM2-MM	0.94	1.25	NorESM Climate modeling Consortium, Oslo, Norway.
TaiESM1	0.94	1.25	Research Center for Environmental Changes, Academia Sinica, Nankang, Taipei, Taiwan.

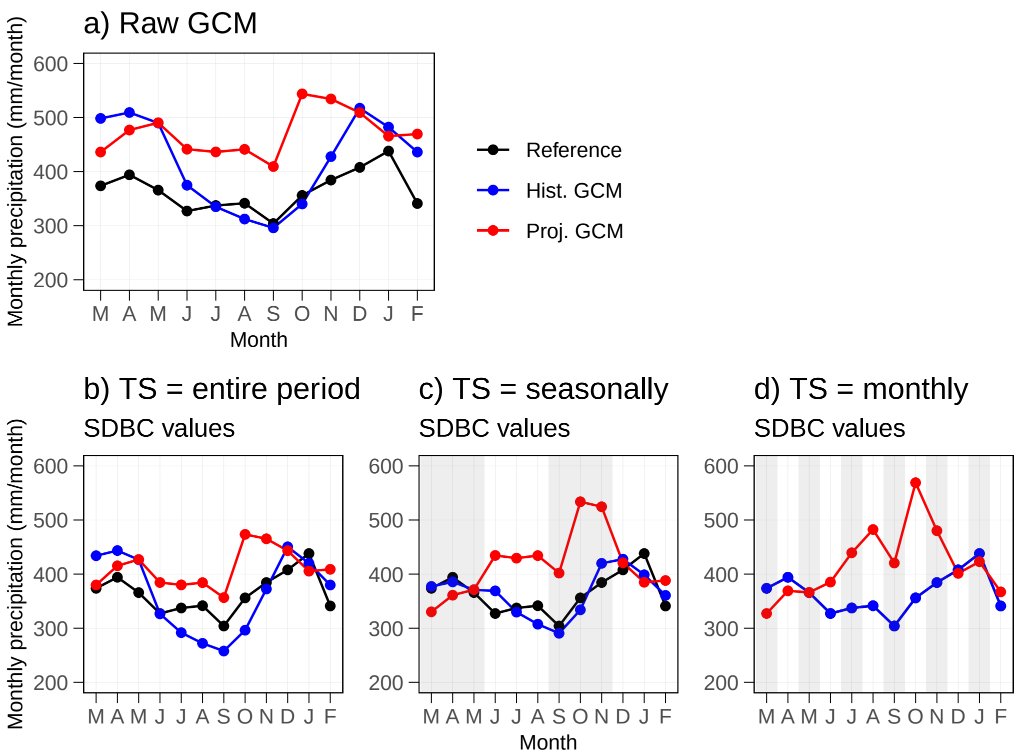

Figure B1. Illustration of the linear scaling method, applied to one grid cell-GCM combination, and its effects on the SDBC-biases and projections. (a) Reference (observational) and raw GCM seasonality during the period 1980-2014 (black and blue lines). The projected raw seasonality is also shown in red (2065-2099). (b), (c) and (d) show the bias-corrected precipitation amounts using the entire period, seasons, and months, respectively, for temporal stratification. The reference value is shown in all panels for completeness, and the shaded areas represent the temporal stratification.

monthly precipitation is October for the three TS, which is the same as the raw GCM projection. However, the projected minimum monthly precipitation is September, March, and March for the entire period, season, and monthly application of the LSM, respectively.

Open Research Section

The CR2MET dataset (Boisier et al., 2018) is available at https://www.cr2.cl/datos-productos-grillados/. The GCMs data was downloaded from the Earth System Grid Federation (https://esgf-node.llnl.gov/search/cmip6/). All the data used in this study is available at https://bhuch.myqnapcloud.com/share.cgi?ssid=43cb3da649cd41ca9bfc42150a855e89.

Acknowledgments

Nicolás Vásquez and Pablo A. Mendoza received support from the Fondecyt project No. 11200142. Nicolás Vásquez also received support from the Emerging Leaders in the Americas Program (ELAP) scholarship (Canada) and the ANID Doctorado Nacional scholarship No. 21230289 (Chile). Pablo A. Mendoza was also supported by ANID/PIA project No. AFB230001.

References

Aceituno, P., Boisier, J. P., Garreaud, R., Rondanelli, R., & Rutllant, J. A. (2021). Climate and Weather in Chile. In *Water resources of chile* (pp. 7–29). doi: 10.1007/978-3-030-56901-3{_}2

Addor, N., Rössler, O., Köplin, N., Huss, M., Weingartner, R., & Seibert, J. (2014). Robust changes and sources of uncertainty in the projected hydrological regimes of Swiss catchments. *Water Resources Research*. doi: 10.1002/2014WR015549

Alder, J. R., & Hostetler, S. W. (2019). The Dependence of Hydroclimate Projections in Snow-Dominated Regions of the Western United States on the Choice of Statistically Downscaled Climate Data. *Water Resources Research*. doi: 10.1029/2018WR023458

Araki, R., Branger, F., Wiekenkamp, I., & McMillan, H. (2022, 4). A signature-based approach to quantify soil moisture dynamics under contrasting land-uses. *Hydrological Processes*, *36*(4). doi: 10.1002/hyp.14553

Boisier, J. P., Álvarez-Garretón, C., Cepeda, J., Osses, A., Vásquez, N., & Rondanelli, R. (2018). CR2MET: A high-resolution precipitation and temperature dataset for hydroclimatic research in Chile. In *Egu general assembly conference abstracts* (p. 19739).

Cannon, A. J. (2011, 9). Quantile regression neural networks: Implementation in R and application to precipitation downscaling. *Computers & Geosciences*, *37*(9), 1277–1284. doi: 10.1016/j.cageo.2010.07.005

Cannon, A. J. (2016, 10). Multivariate Bias Correction of Climate Model Output: Matching Marginal Distributions and Intervariable Dependence Structure. *Journal of Climate*, *29*(19), 7045–7064. doi: 10.1175/JCLI-D-15-0679.1

Cannon, A. J. (2018). Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dynamics*. doi: 10.1007/s00382-017-3580-6

Cannon, A. J., Sobie, S. R., & Murdock, T. Q. (2015). Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate*. doi: 10.1175/JCLI-D-14-00754.1

Chaubey, P. K., & Mall, R. K. (2023, 9). Intensification of Extreme Rainfall in Indian River Basin: Using Bias Corrected CMIP6 Climate Data. *Earth's Future*, *11*(9). doi: 10.1029/2023EF003556

Cheeseman, P., John, R., & Nasa, S. (1996). Bayesian Classification (AutoClass): Theory and Results. *Advances in knowledge discovery and data mining*.

Cheeseman, P., Kelly, J., Self, M., Stutz, J., Taylor, W., & Freeman, D. (1988, 1). AutoClass: A Bayesian Classification System. *Machine Learning Proceedings 1988*, 54–64. doi: 10.1016/B978-0-934613-64-4.50011-6

Chegwidden, O. S., Nijssen, B., Rupp, D. E., Arnold, J. R., Clark, M. P., Hamman, J. J., ... Xiao, M. (2019). How Do Modeling Decisions Affect the Spread Among Hydrologic Climate Change Projections? Exploring a Large Ensemble of Simulations Across a Diversity of Hydroclimates. *Earth's Future*. doi: 10.1029/2018EF001047

Chen, J., Arsenault, R., Brissette, F. P., & Zhang, S. (2021). Climate Change Impact Studies: Should We Bias Correct Climate Model Outputs or Post-Process Impact Model Outputs? *Water Resources Research*. doi: 10.1029/2020WR028638

Chen, J., Brissette, F. P., Chaumont, D., & Braun, M. (2013, 7). Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water Resources Research*, *49*(7), 4187–4205. doi: 10.1002/wrcr.20331

Clark, M. P., Wilby, R. L., Gutmann, E. D., Vano, J. A., Gangopadhyay, S., Wood, A. W., ... Brekke, L. D. (2016, 6). Characterizing Uncertainty of the Hydrologic Impacts of Climate Change. *Current Climate Change Reports*, *2*(2), 55–64. doi: 10.1007/s40641-016-0034-x

Coron, L., Thirel, G., Delaigue, O., Perrin, C., & Andréassian, V. (2017, 8). The suite of lumped GR hydrological models in an R package. *Environmental Modelling & Software*, *94*, 166–171. doi: 10.1016/j.envsoft.2017.05.002

Dettinger, Cayan, D., Meyer, M., & Jeton, A. (2004). Simulated Hydrologic Responses To Climate Variations. *Climatic Change*.

DGA. (2022). *Homologación del cálculo hidrológico para la estimación de la oferta natural del agua histórica y futura en Chile.* (Tech. Rep.). SIT N° 524. Ministerio de Obras Públicas, Dirección General de Aguas, División de Estudios y Planificación, Chile. Elaborado por Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas. Retrieved from https://snia.mop.gob.cl/repositoriodga/handle/20.500.13000/126394

Di Virgilio, G., Ji, F., Tam, E., Nishant, N., Evans, J. P., Thomas, C., ... Delage, F. (2022, 4). Selecting CMIP6 GCMs for CORDEX Dynamical Downscaling: Model Performance, Independence, and Climate Change Signals. *Earth's Future*, *10*(4). doi: 10.1029/2021EF002625

François, B., Vrac, M., Cannon, A. J., Robin, Y., & Allard, D. (2020, 6). Multivariate bias corrections of climate simulations: which benefits for which losses? *Earth System Dynamics*, *11*(2), 537–562. doi: 10.5194/esd-11-537-2020

Ghimire, U., Srinivasan, G., & Agarwal, A. (2019, 3). Assessment of rainfall bias correction techniques for improved hydrological simulation. *International Journal of Climatology*, *39*(4), 2386–2399. doi: 10.1002/joc.5959

Guo, J., Wang, X., Fan, Y., Liang, X., Jia, H., & Liu, L. (2023, 4). How Extreme Events in China Would Be Affected by Global Warming—Insights From a Bias-Corrected CMIP6 Ensemble. *Earth's Future*, *11*(4). doi: 10.1029/2022EF003347

Guo, Q., Chen, J., Zhang, X. J., Xu, C., & Chen, H. (2020, 5). Impacts of Using State-of-the-Art Multivariate Bias Correction Methods on Hydrological Modeling Over North America. *Water Resources Research*, *56*(5). doi: 10.1029/2019WR026659

Gutiérrez, J. M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., ... Pagé, C. (2019, 7). 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*(9), 3750–3785. doi: 10.1002/joc.5462

Gutmann, E., Pruitt, T., Clark, M. P., Brekke, L., Arnold, J. R., Raff, D. A., & Rasmussen, R. M. (2014). An intercomparison of statistical downscaling methods used for water resource assessments in the United States. *Water Resources Research*. doi: 10.1002/2014WR015559

Haerter, J. O., Hagemann, S., Moseley, C., & Piani, C. (2011, 3). Climate model bias correction and the role of timescales. *Hydrology and Earth System Sciences*, *15*(3), 1065–1079. doi: 10.5194/hess-15-1065-2011

Hagemann, S., Chen, C., Haerter, J. O., Heinke, J., Gerten, D., & Piani, C. (2011, 8). Impact of a Statistical Bias Correction on the Projected Hydrological Changes Obtained from Three GCMs and Two Hydrology Models. *Journal of Hydrometeorology*, *12*(4), 556–578. doi: 10.1175/2011JHM1336.1

Hakala, K., Addor, N., & Seibert, J. (2018, 8). Hydrological Modeling to Evaluate Climate Model Simulations and Their Bias Correction. *Journal of Hydrometeorology*, *19*(8), 1321–1337. doi: 10.1175/JHM-D-17-0189.1

Han, P., Long, D., Han, Z., Du, M., Dai, L., & Hao, X. (2019, 4). Improved understanding of snowmelt runoff from the headwaters of China's Yangtze River using remotely sensed snow products and hydrological modeling. *Remote Sensing of Environment*, *224*, 44–59. doi: 10.1016/j.rse.2019.01.041

Hanus, S., Hrachowitz, M., Zekollari, H., Schoups, G., Vizcaino, M., & Kaitna, R. (2021). Future changes in annual, seasonal and monthly runoff signatures in contrasting Alpine catchments in Austria. *Hydrology and Earth System Sciences*. doi: 10.5194/hess-25-3429-2021

Hattermann, F. F., Vetter, T., Breuer, L., Su, B., Daggupati, P., Donnelly, C., ... Krysnaova, V. (2018). Sources of uncertainty in hydrological climate impact assessment: A cross-scale study. *Environmental Research Letters*. doi: 10.1088/1748-9326/aa9938

Her, Y., Yoo, S. H., Cho, J., Hwang, S., Jeong, J., & Seong, C. (2019). Uncertainty in hydrological analysis of climate change: multi-parameter vs. multi-GCM ensemble predictions. *Scientific Reports*. doi: 10.1038/s41598-019-41334-7

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... Thépaut, J. (2020, 7). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. doi: 10.1002/qj.3803

Hess, P., Lange, S., Schötz, C., & Boers, N. (2023, 10). Deep Learning for Bias-Correcting CMIP6-Class Earth System Models. *Earth's Future*, *11*(10). doi: 10.1029/2023EF004002

Jennings, K. S., Winchell, T. S., Livneh, B., & Molotch, N. P. (2018, 3). Spatial variation of the rain–snow temperature threshold across the Northern Hemisphere. *Nature Communications*, *9*(1), 1148. doi: 10.1038/s41467-018-03629-7

Knoben, W. J., Woods, R. A., & Freer, J. E. (2018). A Quantitative Hydrological Climate Classification Evaluated With Independent Streamflow Data. *Water Resources Research*. doi: 10.1029/2018WR022913

Kwon, S., Kim, J., Boo, K., Shim, S., Kim, Y., & Byun, Y. (2019, 3). Performance-based projection of the climate-change effects on precipitation extremes in East Asia using two metrics. *International Journal of Climatology*, *39*(4), 2324–2335. doi: 10.1002/joc.5954

Lafon, T., Dadson, S., Buys, G., & Prudhomme, C. (2013, 5). Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. *International Journal of Climatology*, *33*(6), 1367–1381. doi: 10.1002/joc.3518

Maraun, D. (2016, 12). Bias Correcting Climate Change Simulations - a Critical Review. *Current Climate Change Reports*, *2*(4), 211–220. doi: 10.1007/s40641-016-0050-x

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., ... Thiele-Eich, I. (2010, 9). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, *48*(3), RG3003. doi: 10.1029/2009RG000314

Matiu, M., & Hanzer, F. (2022, 6). Bias adjustment and downscaling of snow cover fraction projections from regional climate models using remote sensing for the European Alps. *Hydrology and Earth System Sciences*, *26*(12), 3037–3054. doi: 10.5194/hess-26-3037-2022

Maurer, E. P., & Pierce, D. W. (2014, 3). Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean. *Hydrology and Earth System Sciences*, *18*(3), 915–925. doi: 10.5194/hess-18-915-2014

McMillan, H. K., Gnann, S. J., & Araki, R. (2022, 6). Large Scale Evaluation of Relationships Between Hydrologic Signatures and Processes. *Water Resources Research*, *58*(6). doi: 10.1029/2021WR031751

Melsen, L. A., Addor, N., Mizukami, N., Newman, A. J., Torfs, P. J., Clark, M. P., ... Teuling, A. J. (2018). Mapping (dis)agreement in hydrologic projections.

Hydrology and Earth System Sciences. doi: 10.5194/hess-22-1775-2018

Melsen, L. A., Teuling, A. J., Torfs, P. J., Zappa, M., Mizukami, N., Mendoza, P. A., . . . Uijlenhoet, R. (2019, 1). Subjective modeling decisions can significantly impact the simulation of flood and drought events. *Journal of Hydrology*, *568*, 1093–1104. doi: 10.1016/J.JHYDROL.2018.11.046

Mendoza, P. A., Clark, M. P., Mizukami, N., Gutmann, E. D., Arnold, J. R., Brekke, L. D., & Rajagopalan, B. (2016). How do hydrologic modeling decisions affect the portrayal of climate change impacts? *Hydrological Processes*. doi: 10.1002/hyp.10684

Meyer, J., Kohn, I., Stahl, K., Hakala, K., Seibert, J., & Cannon, A. J. (2019, 3). Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments. *Hydrology and Earth System Sciences*, *23*(3), 1339–1354. doi: 10.5194/hess-23-1339-2019

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., . . . Sanderson, B. M. (2016, 9). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482. doi: 10.5194/gmd-9-3461-2016

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005, 3). Which potential evapotranspiration input for a lumped rainfall–runoff model? *Journal of Hydrology*, *303*(1-4), 290–306. doi: 10.1016/j.jhydrol.2004.08.026

Pierce, D. W., Cayan, D. R., Maurer, E. P., Abatzoglou, J. T., & Hegewisch, K. C. (2015, 12). Improved Bias Correction Techniques for Hydrological Simulations of Climate Change*. *Journal of Hydrometeorology*, *16*(6), 2421–2442. doi: 10.1175/JHM-D-14-0236.1

Rastogi, D., Kao, S., & Ashfaq, M. (2022, 8). How May the Choice of Downscaling Techniques and Meteorological Reference Observations Affect Future Hydroclimate Projections? *Earth's Future*, *10*(8). doi: 10.1029/2022EF002734

Reiter, P., Gutjahr, O., Schefczyk, L., Heinemann, G., & Casper, M. (2018, 3). Does applying quantile mapping to subsamples improve the bias correction of daily precipitation? *International Journal of Climatology*, *38*(4), 1623–1633. doi: 10.1002/joc.5283

Ruffault, J., Martin-StPaul, N. K., Duffet, C., Goge, F., & Mouillot, F. (2014, 7). Projecting future drought in Mediterranean forests: bias correction of climate models matters! *Theoretical and Applied Climatology*, *117*(1-2), 113–122. doi: 10.1007/s00704-013-0992-z

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., & Carrillo, G. (2011). Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrology and Earth System Sciences Discussions*. doi: 10.5194/hessd-8-4495-2011

Sepúlveda, U. M., Mendoza, P. A., Mizukami, N., & Newman, A. J. (2022, 7). Revisiting parameter sensitivities in the variable infiltration capacity model across a hydroclimatic gradient. *Hydrology and Earth System Sciences*, *26*(13), 3419–3445. doi: 10.5194/hess-26-3419-2022

Stoner, A. M., Hayhoe, K., Yang, X., & Wuebbles, D. J. (2013). An asynchronous regional regression model for statistical downscaling of daily climate variables. *International Journal of Climatology*. doi: 10.1002/joc.3603

Switanek, M. B., Troch, P. A., Castro, C. L., Leuprecht, A., Chang, H.-I., Mukherjee, R., & Demaria, E. M. C. (2017, 6). Scaled distribution mapping: a bias correction method that preserves raw climate model projected changes. *Hydrology and Earth System Sciences*, *21*(6), 2649–2666. doi: 10.5194/hess-21-2649-2017

Taylor, K. E. (2001, 4). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, *106*(D7), 7183–7192. doi: 10.1029/2000JD900719

Teng, J., Potter, N. J., Chiew, F. H. S., Zhang, L., Wang, B., Vaze, J., & Evans, J. P. (2015, 2). How does bias correction of regional climate model precipitation affect modelled runoff? *Hydrology and Earth System Sciences*, *19*(2), 711–728. doi: 10.5194/hess-19-711-2015

Teutschbein, C., & Seibert, J. (2010, 7). Regional Climate Models for Hydrological Impact Studies at the Catchment Scale: A Review of Recent Modeling Strategies. *Geography Compass*, *4*(7), 834–860. doi: 10.1111/j.1749-8198.2010.00357.x

Vano, J. A., Kim, J. B., Rupp, D. E., & Mote, P. W. (2015). Selecting climate change scenarios using impact-relevant sensitivities. *Geophysical Research Letters*. doi: 10.1002/2015GL063208

Vásquez, N., Cepeda, J., Gómez, T., Mendoza, P. A., Lagos, M., Boisier, J. P., ... Vargas, X. (2021). Catchment-Scale Natural Water Balance in Chile. In (pp. 189–208). Retrieved from `http://link.springer.com/10.1007/978-3-030-56901-3_9` doi: 10.1007/978-3-030-56901-3{_}9

Vicuña, S., Vargas, X., Boisier, J. P., Mendoza, P. A., Gómez, T., Vásquez, N., & Cepeda, J. (2021). Impacts of Climate Change on Water Resources in Chile. In B. Fernández & J. Gironás (Eds.), *Water resources of chile* (pp. 347–363). Cham: Springer International Publishing. Retrieved from `https://doi.org/10.1007/978-3-030-56901-3_19` doi: 10.1007/978-3-030-56901-3{_}19

Vogel, E., Johnson, F., Marshall, L., Bende-Michl, U., Wilson, L., Peter, J. R., ... Duong, V. C. (2023, 7). An evaluation framework for downscaling and bias correction in climate change impact studies. *Journal of Hydrology*, *622*, 129693. doi: 10.1016/j.jhydrol.2023.129693

Vrac, M., & Thao, S. (2020, 11). R2D2; v2.0: accounting for temporal dependences in multivariate bias correction via analogue rank resampling. *Geoscientific Model Development*, *13*(11), 5367–5387. doi: 10.5194/gmd-13-5367-2020

Wan, Z. (2014, 1). New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote Sensing of Environment*, *140*, 36–45. doi: 10.1016/j.rse.2013.08.027

Werner, A. T., & Cannon, A. J. (2016, 4). Hydrologic extremes – an intercomparison of multiple gridded statistical downscaling methods. *Hydrology and Earth System Sciences*, *20*(4), 1483–1508. doi: 10.5194/hess-20-1483-2016

Widmann, M., Bretherton, C. S., & Salathé, E. P. (2003, 3). Statistical Precipitation Downscaling over the Northwestern United States Using Numerically Simulated Precipitation as a Predictor*. *Journal of Climate*, *16*(5), 799–816. doi: 10.1175/1520-0442(2003)016⟨0799:SPDOTN⟩2.0.CO;2

Wilby, R. L., & Dessai, S. (2010). Robust adaptation to climate change. *Weather*. doi: 10.1002/wea.543

Woods, R. A. (2009, 10). Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks. *Advances in Water Resources*, *32*(10), 1465–1481. doi: 10.1016/j.advwatres.2009.06.011

Wootten, A. M., Dixon, K. W., Adams-Smith, D. J., & McPherson, R. A. (2021, 2). Statistically downscaled precipitation sensitivity to gridded observation data and downscaling technique. *International Journal of Climatology*, *41*(2), 980–1001. doi: 10.1002/joc.6716

Wu, Y., Miao, C., Fan, X., Gou, J., Zhang, Q., & Zheng, H. (2022, 11). Quantifying the Uncertainty Sources of Future Climate Projections and Narrowing Uncertainties With Bias Correction Techniques. *Earth's Future*, *10*(11). doi: 10.1029/2022EF002963

Xavier, A. C. F., Martins, L. L., Rudke, A. P., de Morais, M. V. B., Martins, J. A., & Blain, G. C. (2022, 1). Evaluation of Quantile Delta Mapping as a bias-correction method in maximum rainfall dataset from downscaled models in São Paulo state (Brazil). *International Journal of Climatology*, *42*(1), 175–190. doi: 10.1002/joc.7238

Pitfalls in using statistical bias-correction methods to characterize climate change impacts

Nicolás A. Vásquez[1], Pablo A. Mendoza[1,2], Wouter J. M. Knoben[3 *], Louise Arnal[3 †], Miguel Lagos-Zúñiga[4], Martyn Clark[3 ‡], Ximena Vargas[1]

[1]Department of Civil Engineering, Universidad de Chile

[2]Advanced Mining Technology Center, Universidad de Chile

[3]Centre for Hydrology, University of Saskatchewan, Canmore, Alberta, Canada

[4]Center for Climate and Resilience Research, Universidad de Chile

---

Corresponding author: N. A. Vásquez, Department of Civil Engineering, Universidad de Chile, 2002 Almirante Blanco Encalada Av., Santiago, 8370449, Chile. (nicolas.vasquez.pl@uchile.cl)

[*]Department of Civil Engineering, Schulich School of Engineering, University of Calgary, Calgary, Alberta, Canada

[†]now at Ouranos, Montreal, Quebec, Canada

[‡]Department of Civil Engineering, Schulich School of Engineering, University of Calgary, Calgary, Alberta, Canada

October 31, 2023, 3:10pm

Contents of this file

1. Figures S1 to S19 expand the results to other climate indices since the main manuscript only contains results for precipitation at different time scales.

Additional Supporting Information (Files uploaded separately)

1. Taylor Skill Score (Taylor, 2001) for (i) each climate region (ds01) and (ii) grid cell (ds02), uploaded as excel (.xlsx) files.

2. Grid cells' coordinates and attributes used for clustering (ds03; as .csv).

Introduction. The material included here expands the results presented for precipitation (during the historical period) to the rest of the climate indices: (i) air temperature (T), (ii) diurnal temperature range (DTR), (iii) precipitation (P), (iv) coefficient of variation of inter-annual precipitation (c.o.v. P), (v) 1% highest daily precipitation (P-1%), (vi) wet spell length (WSL), (vii) dry spell length (DSL), (viii) wet fraction (WF) and (ix) snowfall fraction (SF). Figures S1 to S9 display the biases at the annual, seasonal, and monthly time scales. Figures S10 to S19 show the results of the Analysis of Variance (ANOVA). Two additional files were uploaded separately. They contain the Taylor Skill score values (TSS) for each grid cell and also the TSS computed at the climate cluster scale. TSS cluster values were calculated from the mean cluster precipitation seasonality for the period 1980-2014 (as the grid cell average within each cluster).

Biases After Applying Correction Methods Figures S1 to S9 display the bias of each climate index at the annual and monthly time scales, disaggregated by the temporal stratification (TS) considered to bias correct the raw GCM outputs (entire period, seasons, and months). When not shown, the unit of the bias corresponds to the difference between the bias-corrected GCM and the reference ($X_{GCM} - X_{ref}$).

Relative Importance to Remove Biases. Figure S10 shows the relative importance of the bias correction method (BCM) and the TS to explain the variance of errors in bias-corrected climate indices during the historical period for Continental Chile, based on Analysis of Variance (ANOVA). The Total Variance (TV) is formulated as $TV = BCM + TS + Residuals$. Results from the ANOVA analysis (BCM/TV, TS/TV, and $Residuals/TV$) are computed for each grid cell and GCM and subsequently averaged for continental Chile. Figures S11 to S19 show the same results, disaggregated by climate clusters.

References

Taylor, K. E. (2001, 4). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, *106*(D7), 7183–7192. doi: 10.1029/2000JD900719

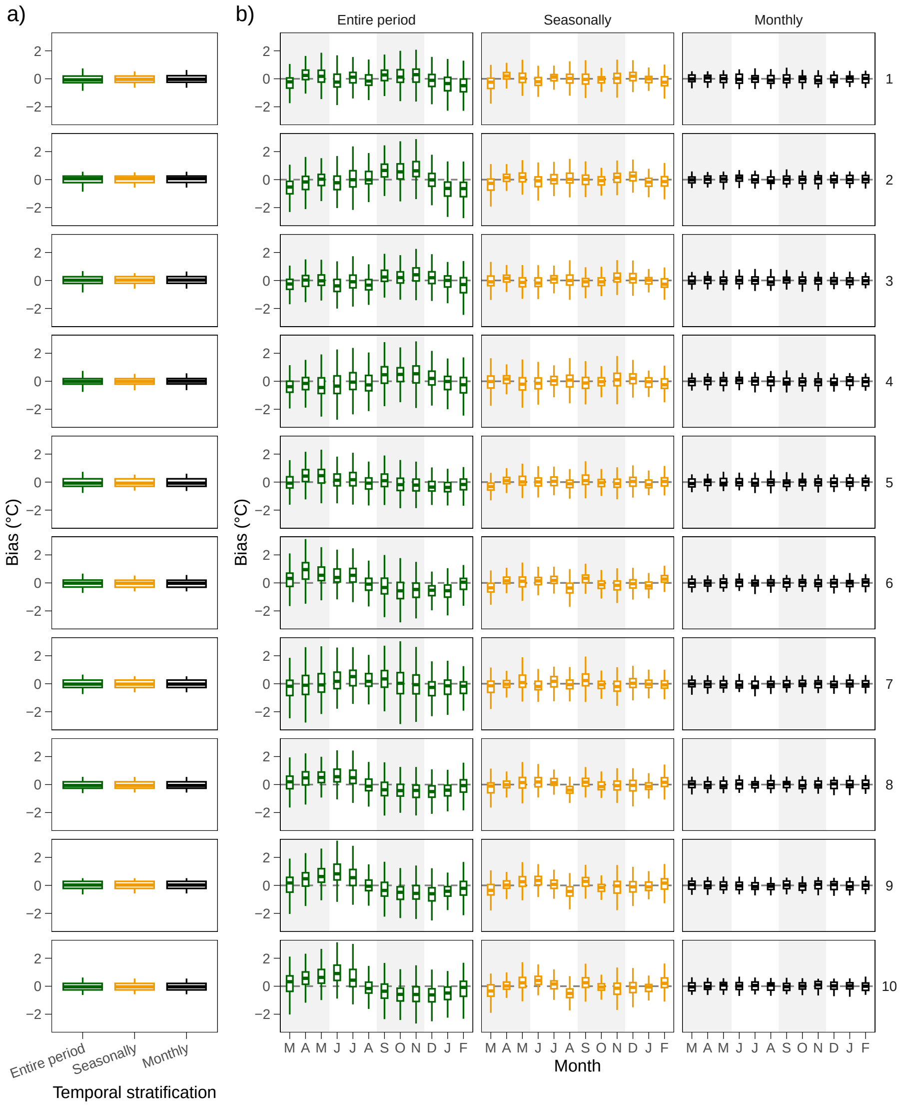

Figure S1. Temperature bias at the (a) annual and (b) monthly time scales after bias correction, separated for each climatic group (rows). The colors indicate the temporal stratification used to apply the BCM. Biases are computed for the 1980-2014 period.

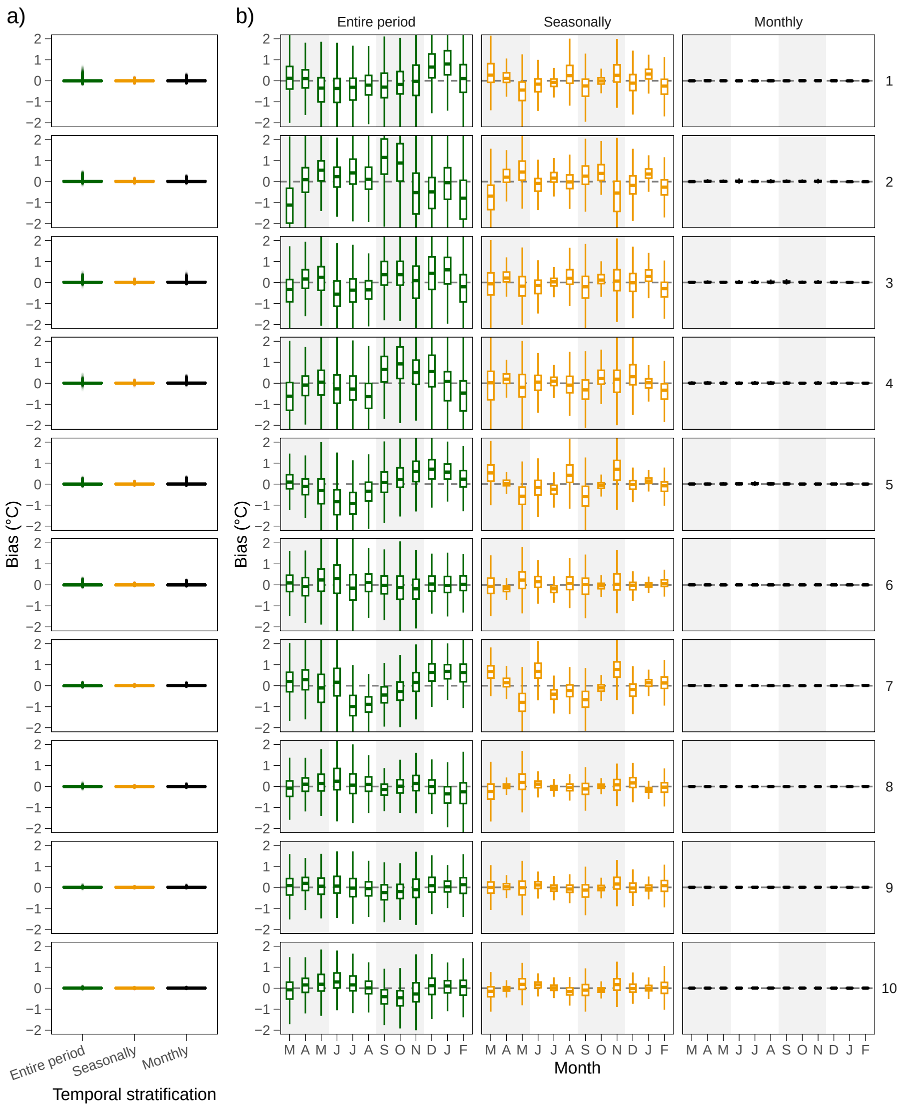

Figure S2. Same as Figure S1, but for diurnal temperature range.

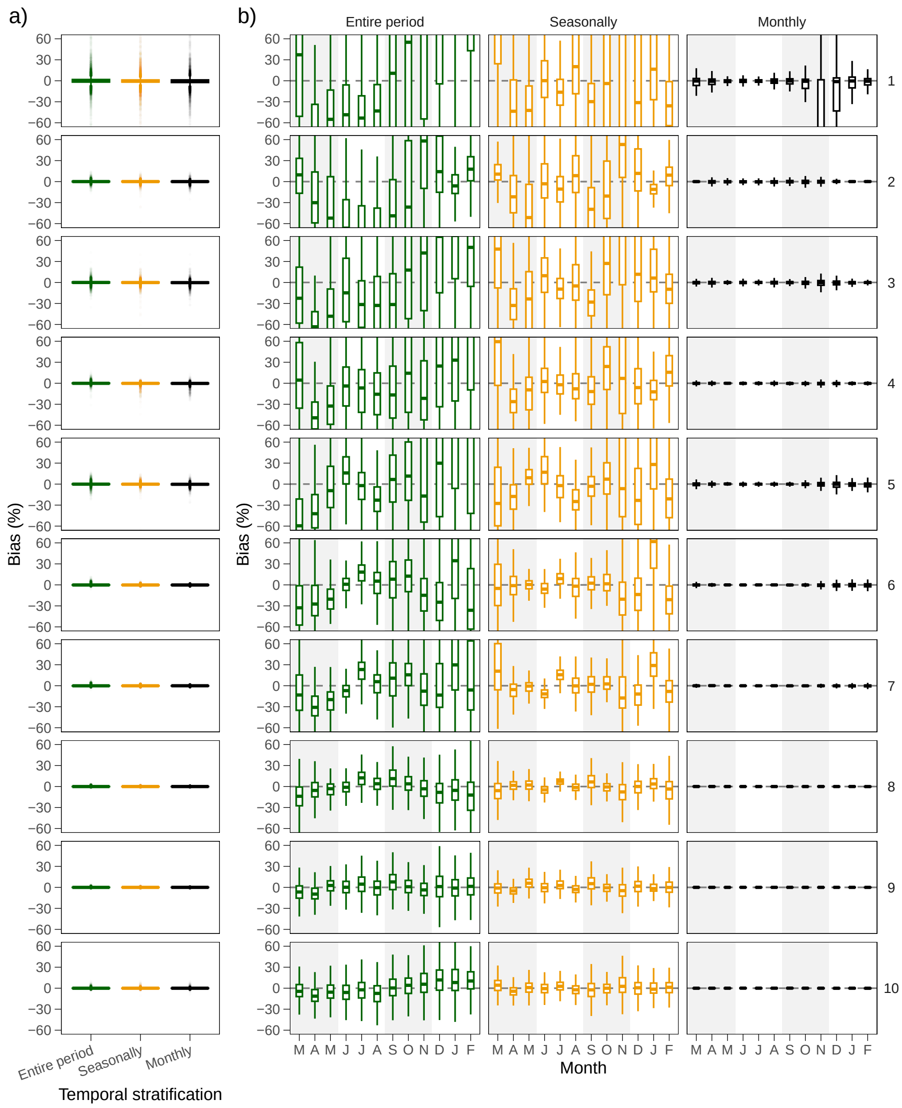

Figure S3. Same as in Figure S1, but for precipitation

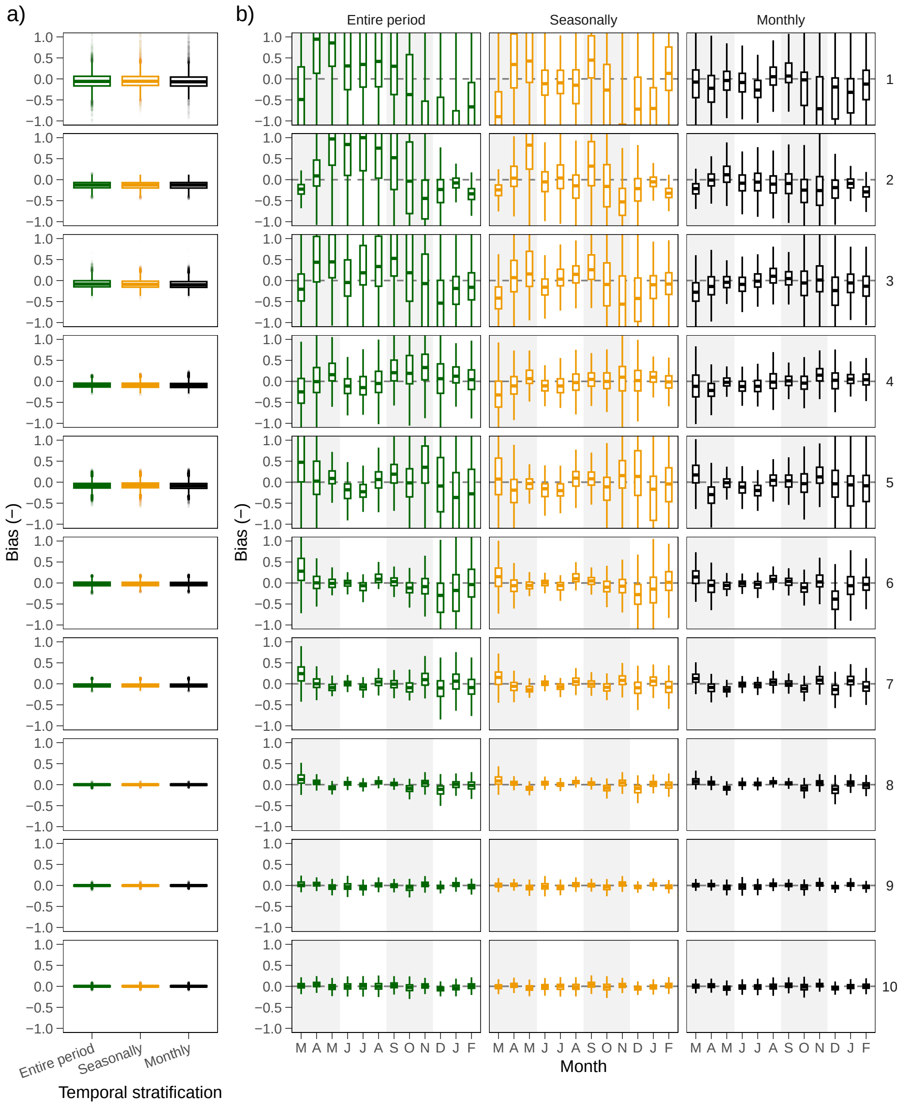

Figure S4. Same as Figure S1, but for the coefficient of variation for inter-annual precipitation

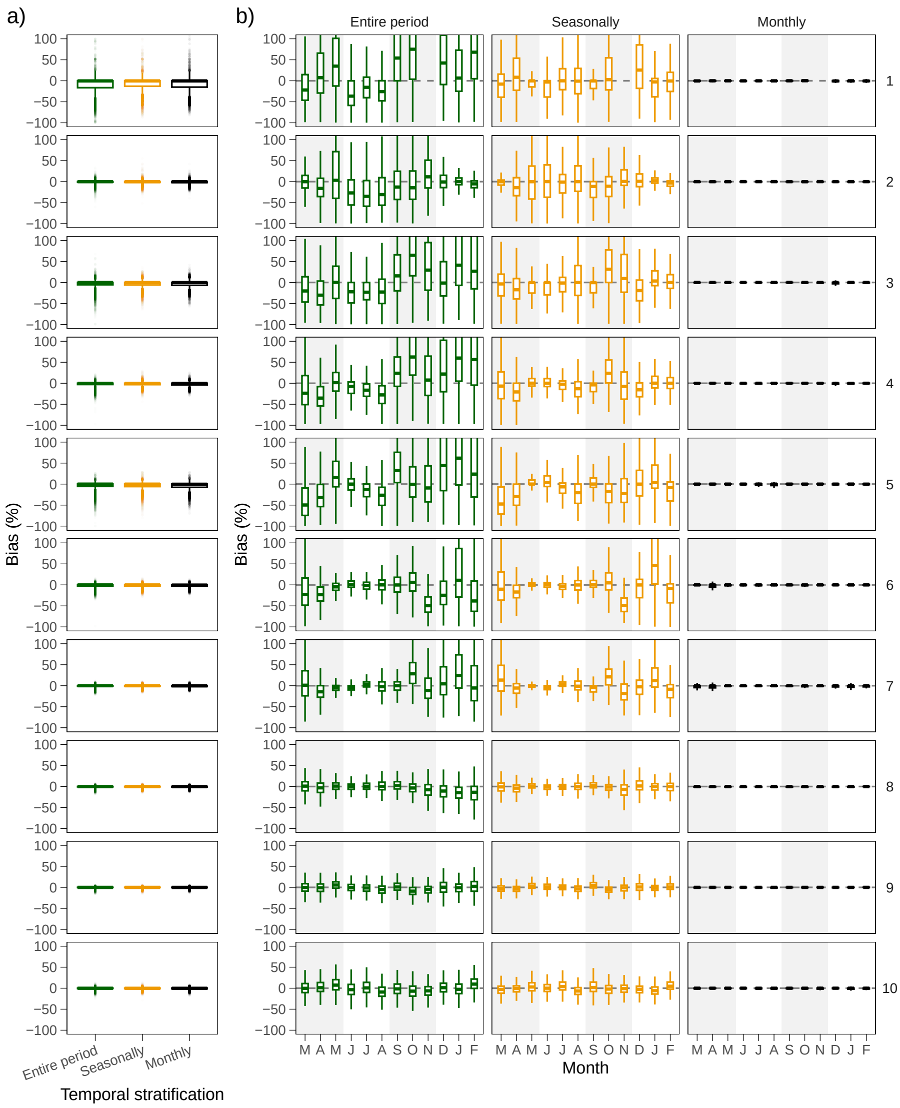

Figure S5. Same as in Figure S1, but for the highest 1% daily precipitation

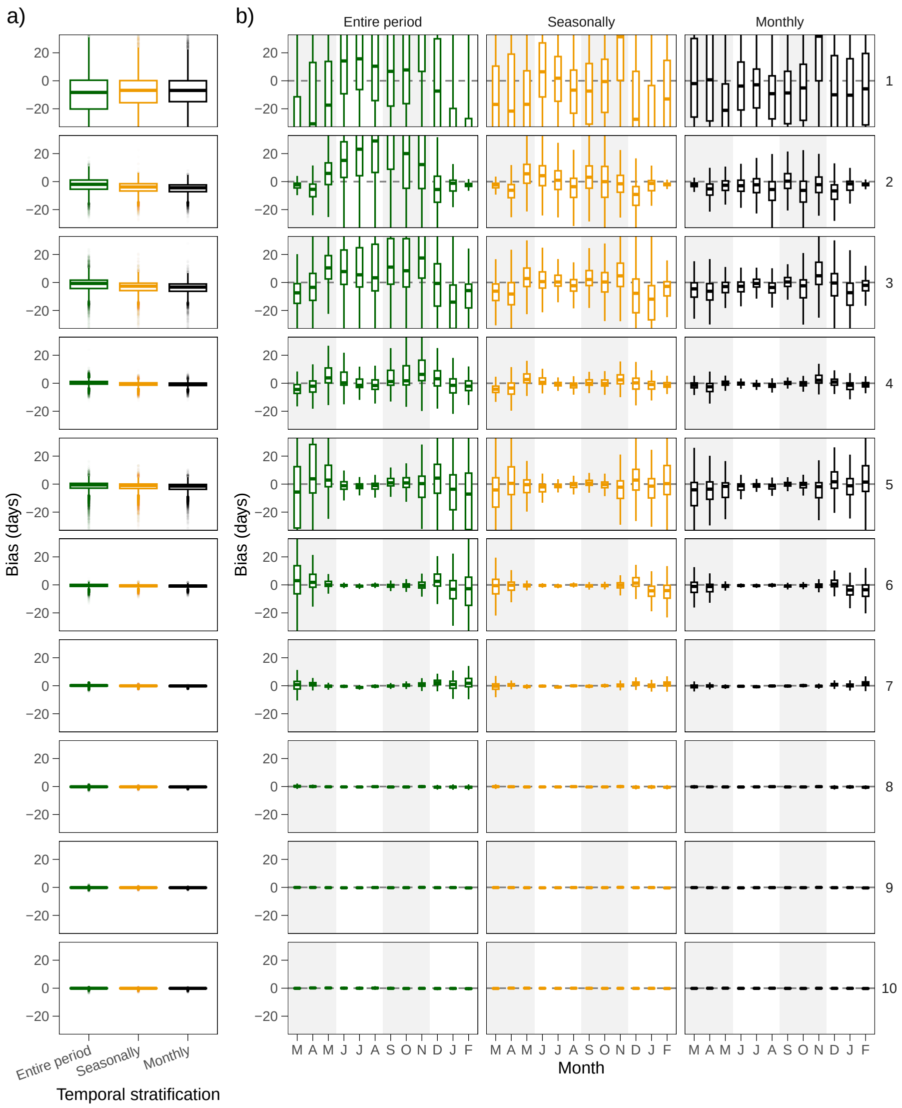

Figure S6. Same as in Figure S1, but for the dry spell length

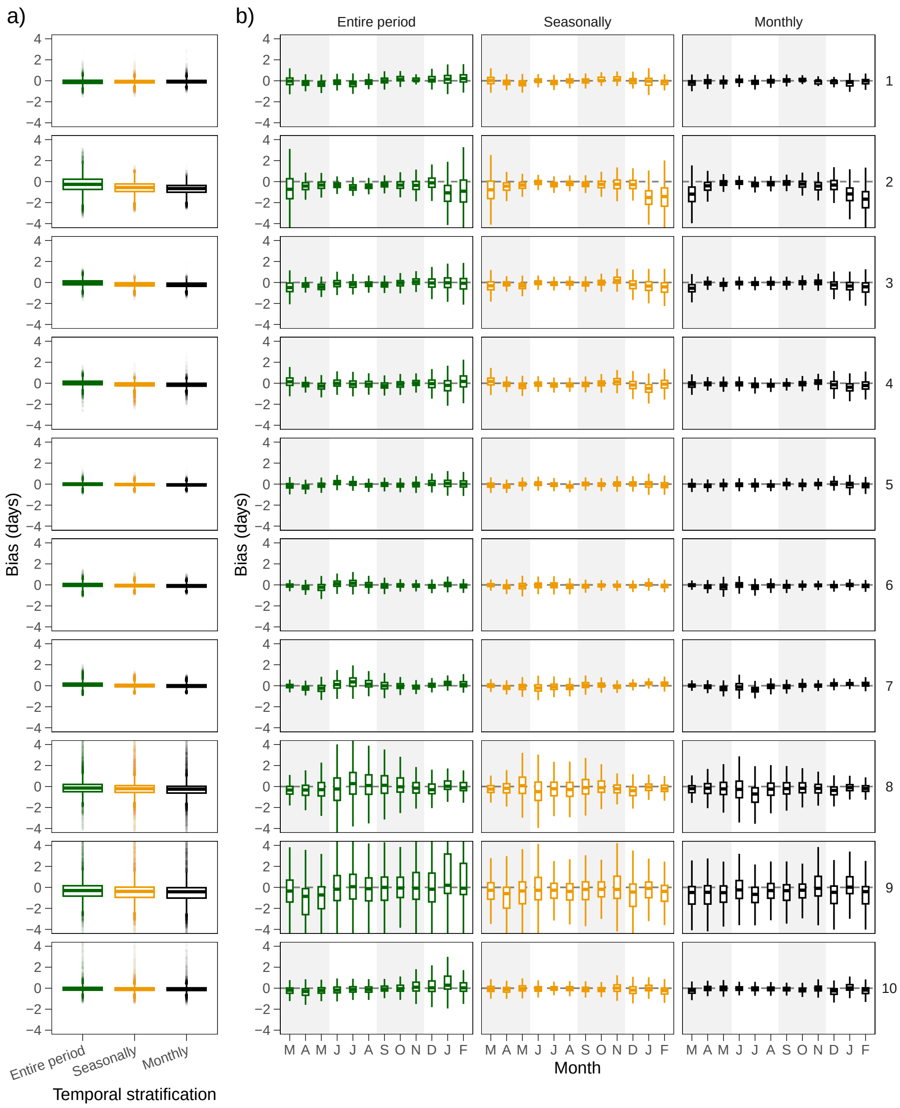

Figure S7. Same as in Figure S1, but for the wet spell length

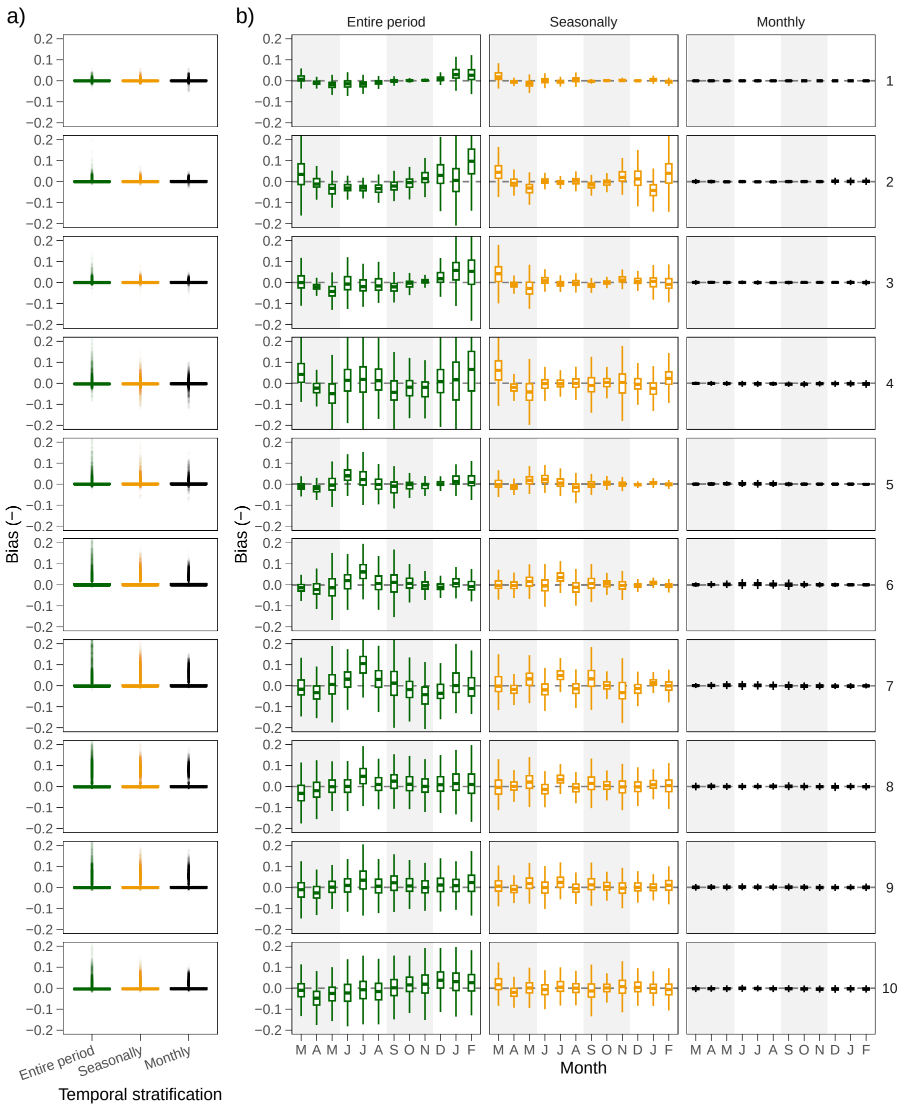

Figure S8. Same as in Figure S1, but for the wet day fraction

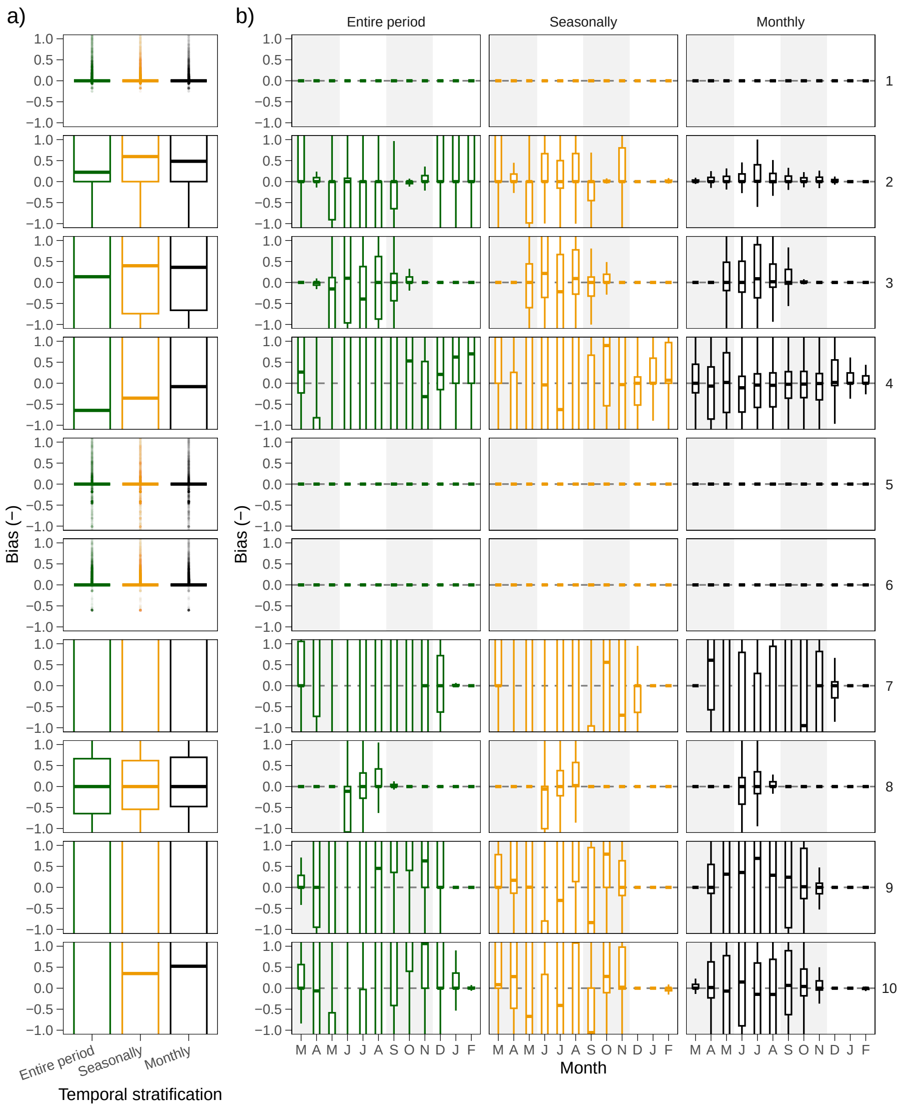

Figure S9. Same as in Figure S1, but for the snowfall fraction

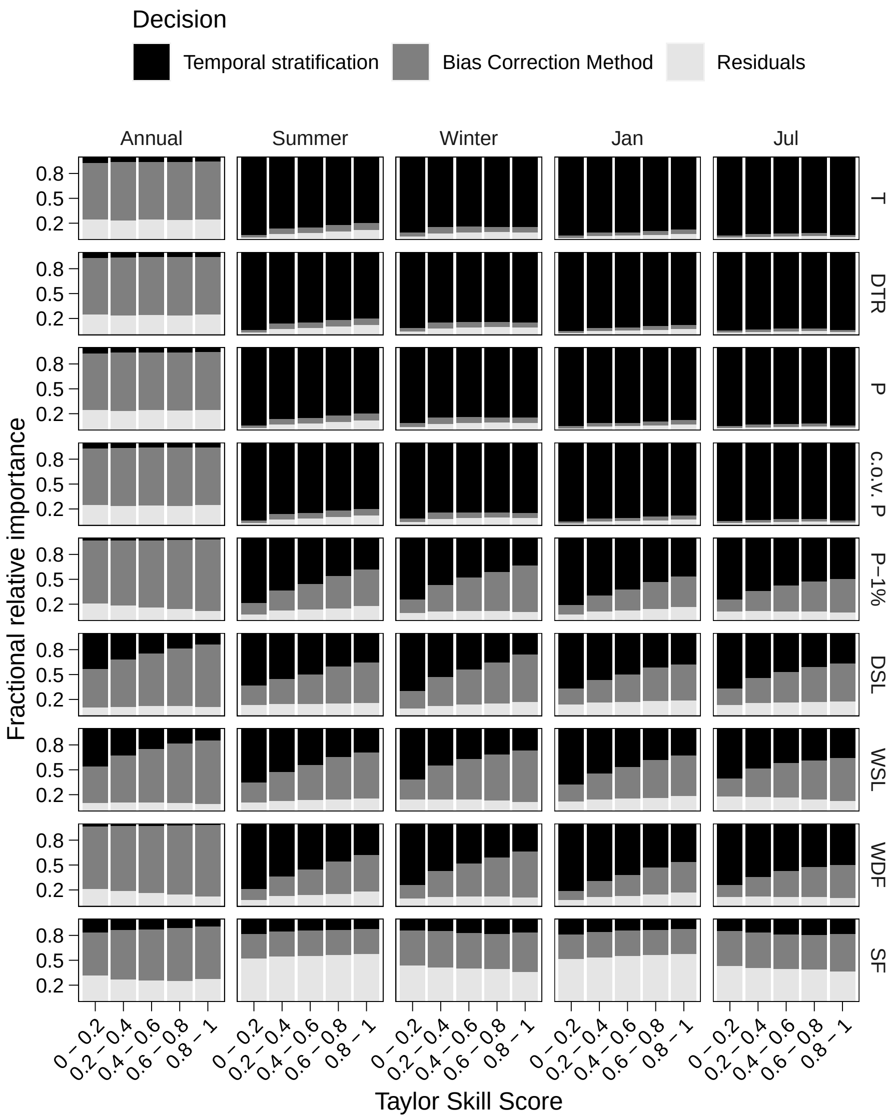

Figure S10. Relative importance (averaged across all grid cells and GCMs) of the bias correction method and the temporal stratification to explain the dispersion of biases with respect to the reference dataset at the annual, seasonal (DJF and JJA), and monthly (January and July) time scales during the historical period (1980-2014). Results are stratified according to the historical raw GCM performance (measured by the TSS; x-axis). Biases are computed after applying the BCMs, and results are displayed for temperature (T), diurnal temperature range (DTR), precipitation (P), coefficient of variation of inter-annual precipitation (c.o.v. P), highest 1% daily precipitation amount (P-1%), dry spell length (DSL), wet spell length (WSL) and snowfall fraction (SF).

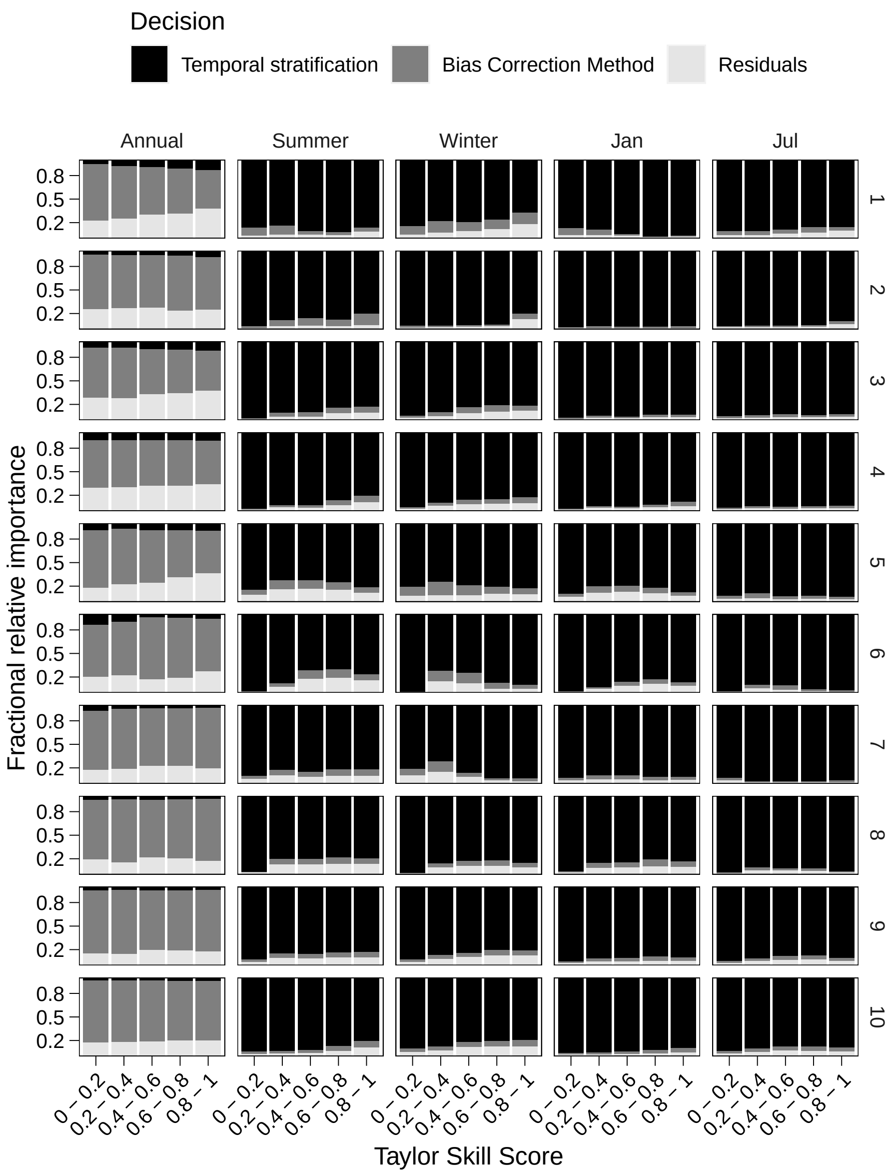

Figure S11. Relative importance (averaged across all GCMs and grid cells within each climate group) of the bias correction method and the temporal stratification to explain the dispersion of temperature biases (with respect to the reference dataset) at the annual, seasonal (DJF and JJA), and monthly (January and July) time scales during the historical period (1980-2014). Results are stratified according to the historical raw GCM performance (measured by the TSS; x-axis) and climate group (rows). Biases are computed after applying the BCMs.

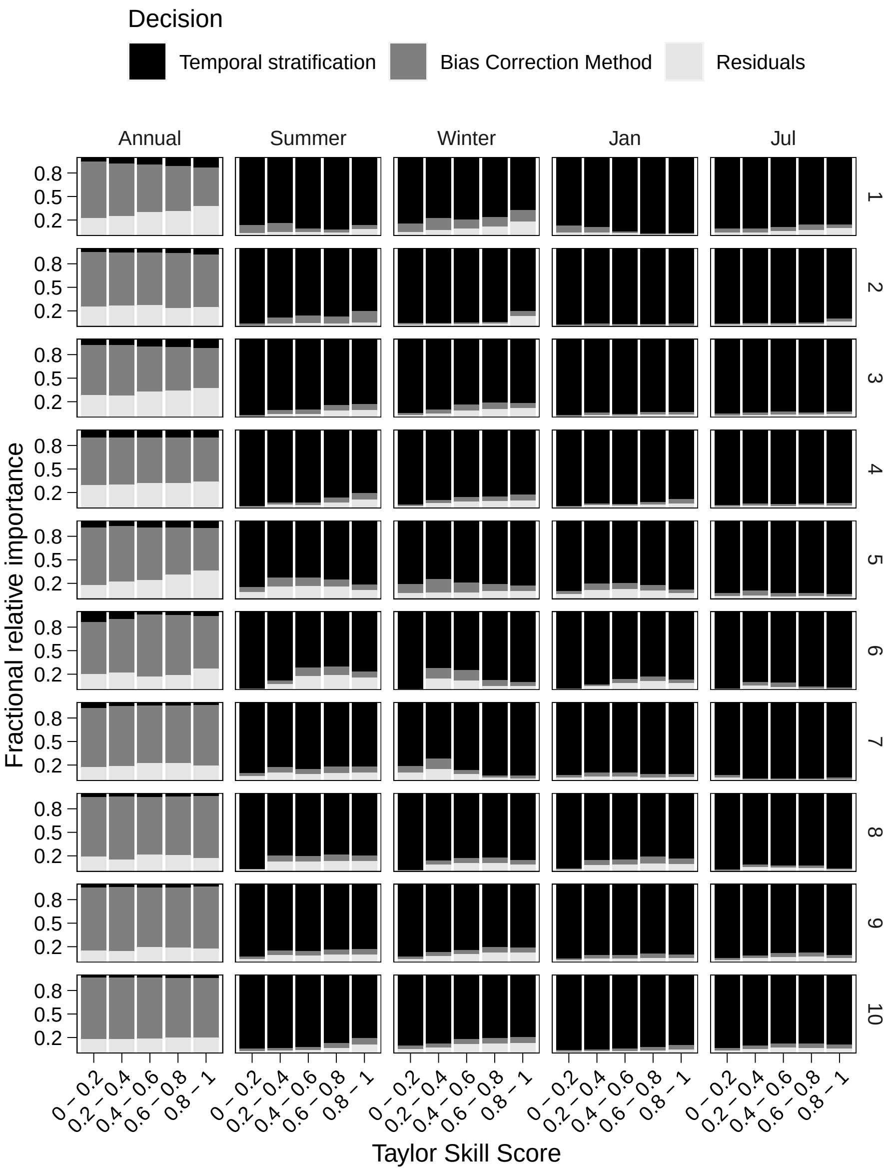

Figure S12. Same as in Figure S11, but for diurnal temperature range.

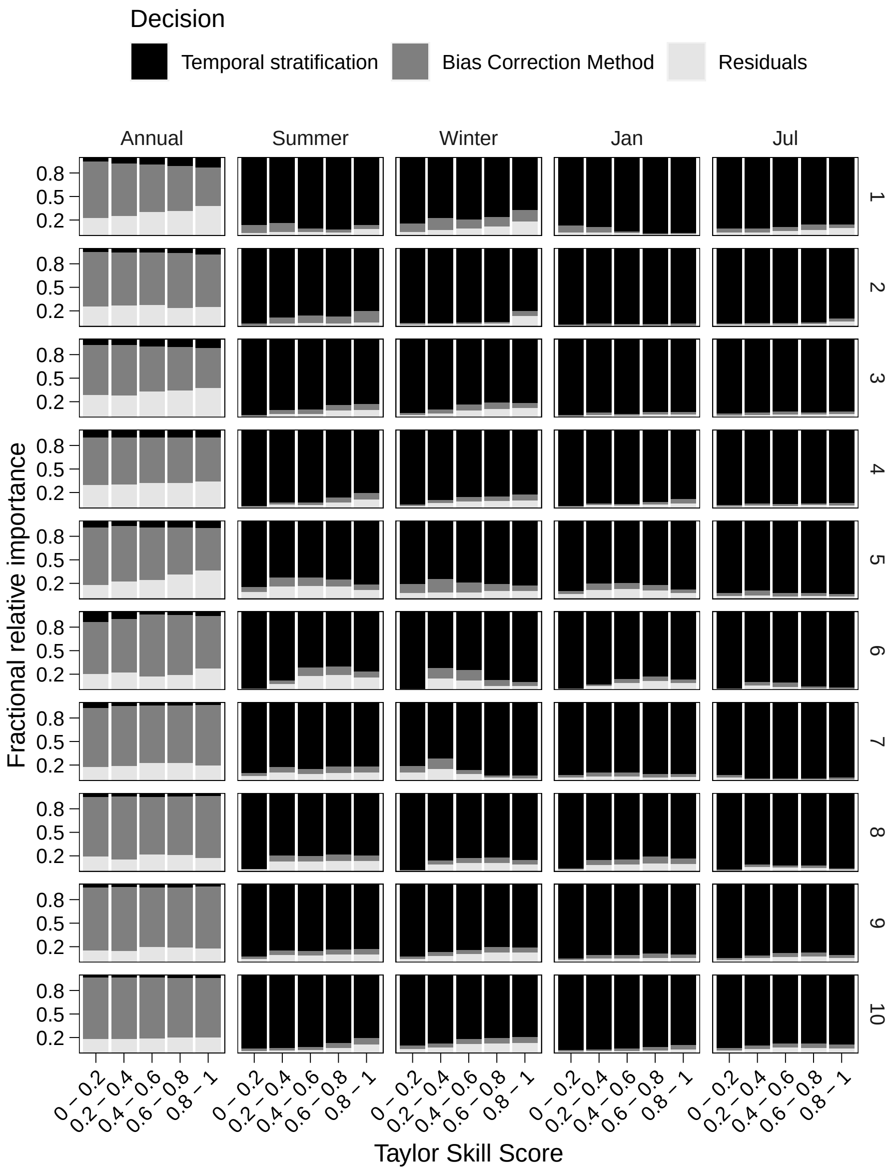

Figure S13. Same as in Figure S11, but for precipitation.

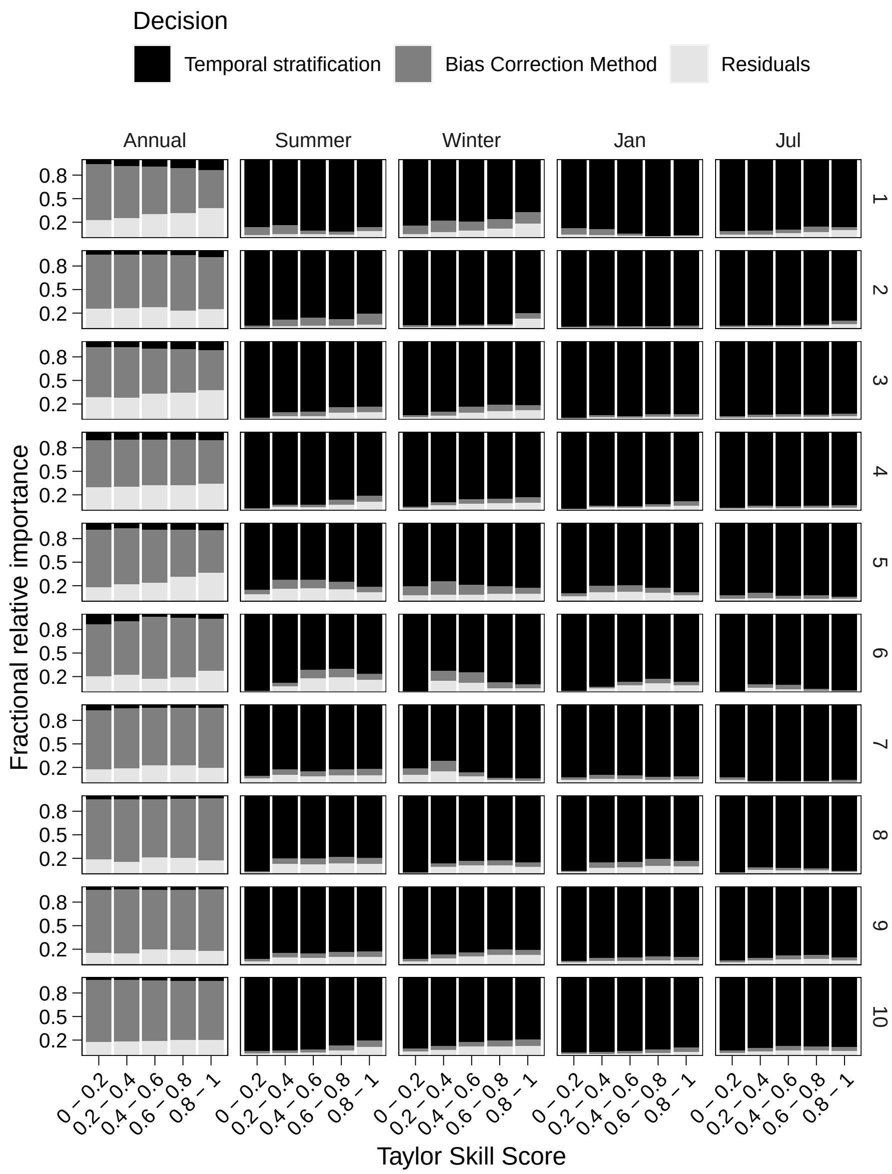

Figure S14. Same as in Figure S11, but for the coefficient of variation of inter-annual precipitation.

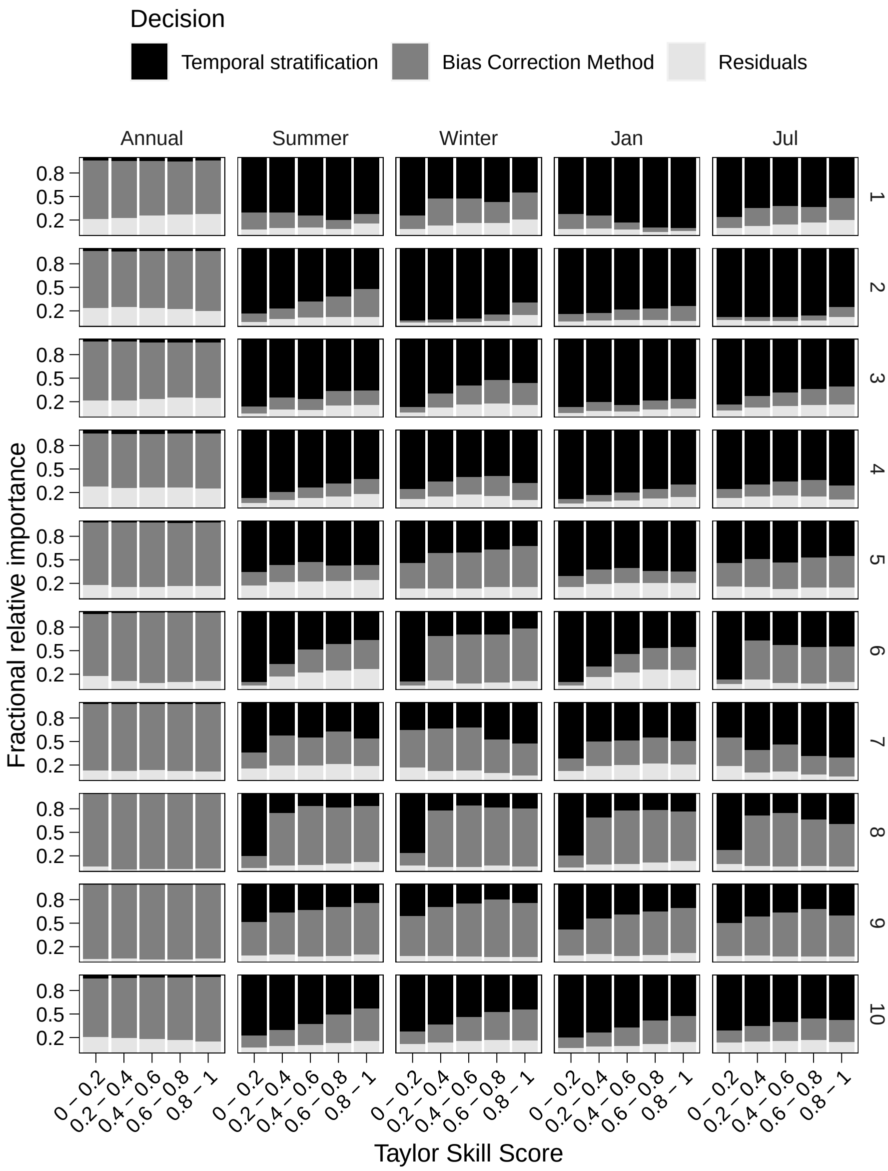

Figure S15. Same as in Figure S11, but for the highest 1% daily precipitation.

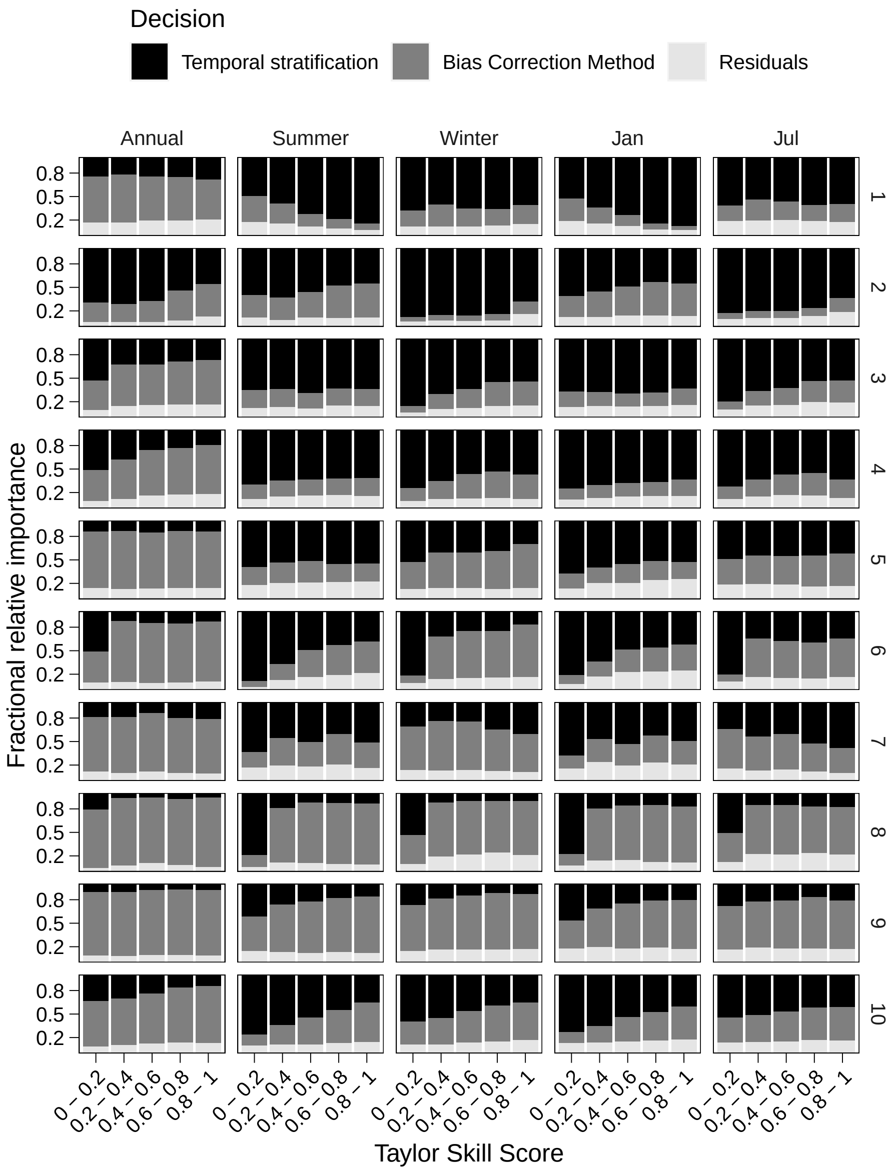

Figure S16. Same as in Figure S11, but for dry spell length.

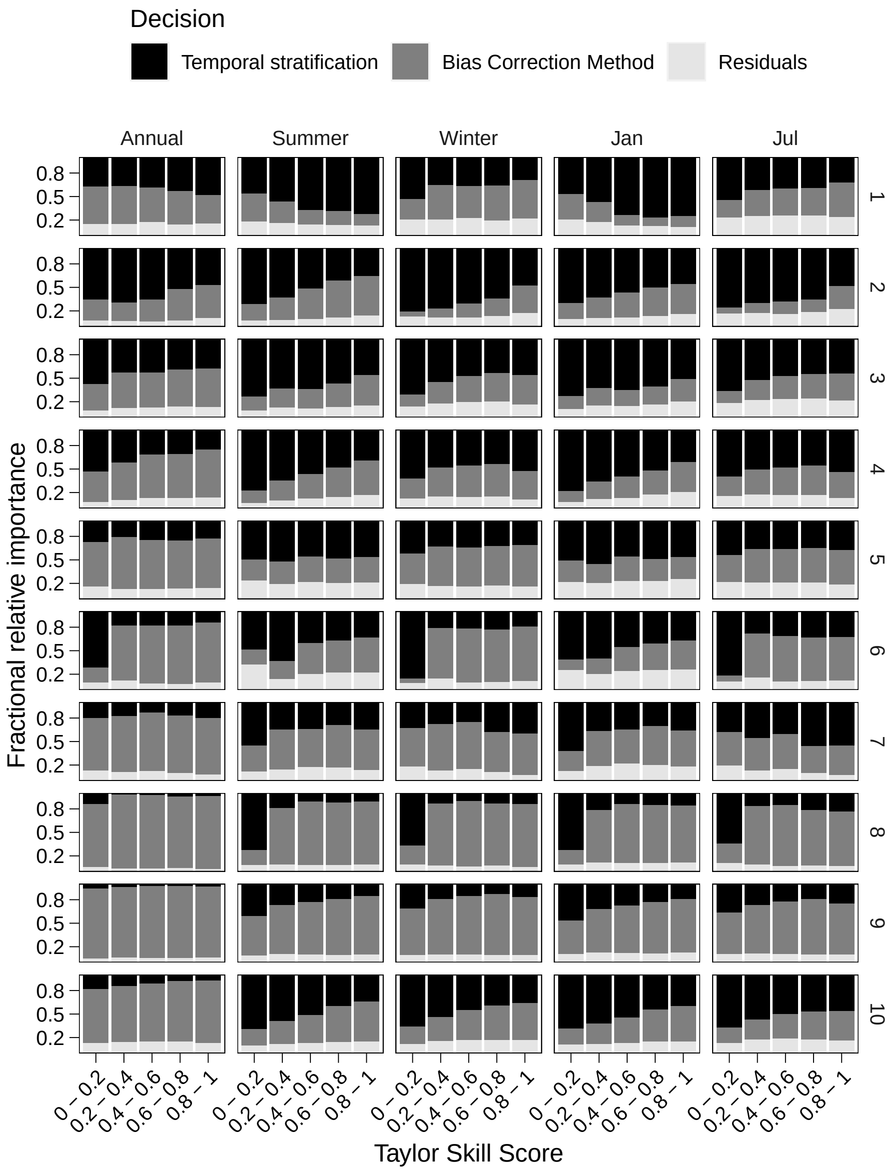

Figure S17. Same as in Figure S11, but for wet spell length.

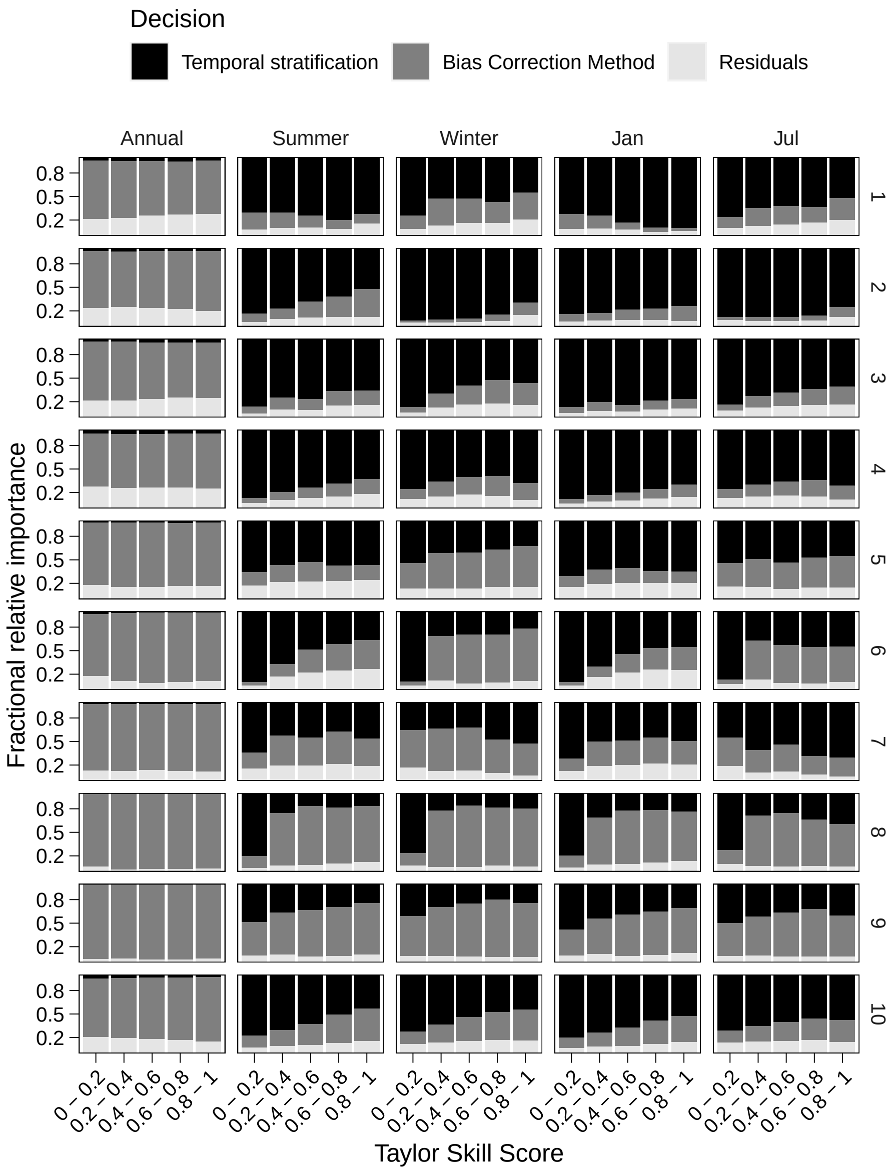

Figure S18. Same as in Figure S11, but for wet day fraction

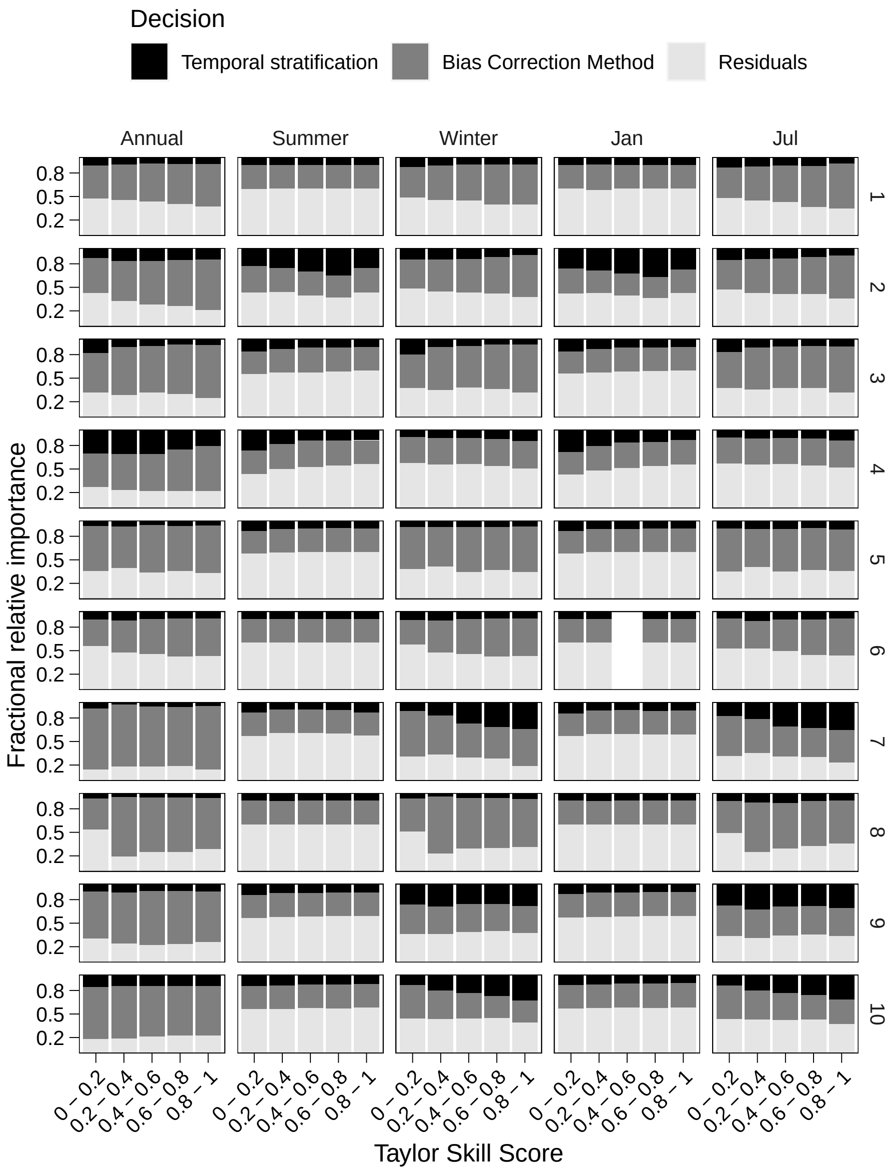

Figure S19. Same as in Figure S11, but for the snowfall fraction.