ibicus: a new open-source Python package and comprehensive interface for statistical bias adjustment and evaluation in climate modelling (v1.0.1)

Fiona Spuler[1,*], Jakob Wessel[2,*], Edward Comyn-Platt[3], James Varndell[3], and Chiara Cagnazzo[3]

[1]Department of Meteorology, University of Reading
[2]Department of Mathematics and Statistics, University of Exeter
[3]European Centre for Medium-Range Weather Forecasts (ECMWF)
[*]These authors contributed equally to this work.

Correspondence: Jakob Wessel (jw1301@exeter.ac.uk)

Abstract. Statistical bias adjustment is commonly applied to climate models before using their results in impact studies. However, different methods, based on a distributional mapping between observational and model data, can change the simulated trends, as well as the spatiotemporal and inter-variable consistency of the model, and are prone to misuse if not evaluated thoroughly. Despite the importance of these fundamental issues, researchers who apply bias adjustment currently do not have

5   the tools at hand to compare different methods or evaluate the results sufficiently to detect possible distortions. Because of this, widespread practice in statistical bias adjustment is not aligned with recommendations from the academic literature. To address the practical issues impeding this, we introduce *ibicus*, an open-source Python package for the implementation of eight different peer-reviewed and widely used bias adjustment methods in a common framework and their comprehensive evaluation. The evaluation framework introduced in *ibicus* allows the user to analyse changes to the marginal, spatiotemporal

10   and inter-variable structure of user-defined climate indices and distributional properties, as well as any alteration of the climate change trend simulated in the model. Applying *ibicus* in a case study over the Mediterranean region using seven CMIP6 global circulation models, this study finds that the most appropriate bias adjustment method depends on the variable and impact studied and that even methods that aim to preserve the climate change trend can modify it. These findings highlight the importance of a use-case-specific choice of method and the need for a rigorous evaluation of results when applying statistical

15   bias adjustment.

1   Introduction

Even though climate models have greatly improved in recent decades, simulations of present-day climate still exhibit biases. This means that there are systematic discrepancies between model output and observations that become especially relevant when using the output of climate models for local impact studies, for example by running agricultural or hydrological models.

20   To account for and potentially correct these biases, it has become common practice to post-process climate models using statistical bias adjustment before using their output for impact studies. The idea behind statistical bias adjustment is to calibrate an empirical transfer function between the observed and climate model distribution of a chosen variable. A variety of statistical

bias adjustment methods have been developed and published in recent years, ranging from simple adjustments to the mean, to trend-preserving adjustments by quantile and multivariate methods (Michelangeli et al., 2009; Li et al., 2010; Cannon et al., 2015; Vrac and Friederichs, 2015; Maraun, 2016; Switanek et al., 2017; Lange, 2019, and more).

25

Despite widespread use, ranging from the IPCC AR6 WGI&II (IPCC, 2021, 2022) reports to the climate scenarios used by central banks across the world (NGFS, 2021), Maraun et al. (2017) and others highlight fundamental issues with statistical bias adjustment and show that the approach can destroy the spatiotemporal and inter-variable consistency of the climate model and is prone to misuse. To avoid misuse, Maraun et al. (2017) recommend the evaluation of non-calibrated aspects, the development of process-informed bias adjustment methods based on an understanding of climate model errors, and the selection of climate models that represent the large-scale patterns and feedback relevant to the impact sufficiently well.

30

We argue that the remedies mentioned above are not common practice due to practical issues with statistical bias adjustment. As Ehret et al. (2012); Maraun (2016); Casanueva et al. (2020) highlight, different bias adjustment approaches are appropriate for different use cases. However, methods that exist in the academic literature are published either only as papers, bias adjusted datasets (Dumitrescu et al., 2020; Mishra et al., 2020; Navarro-Racines et al., 2020; Xu et al., 2021, and more) or as stand-alone packages across multiple programming languages (Iturbide et al., 2019; Lange, 2021b; Michelangeli, 2021; Cannon, 2023, and more), often without accompanying evaluation or evaluation frameworks. This gives users who are not necessarily experts in these methods limited options to choose the bias adjustment method most appropriate for their use case and evaluate the results sufficiently to detect issues.

35

In this paper, we introduce *ibicus*, an open-source Python package for the implementation, comparison and evaluation of bias adjustment for climate model outputs. The contribution of *ibicus* is two-fold: It provides a unique unified interface to apply eight different peer-reviewed and widely used bias adjustment methodologies, including Scaled Distribution Matching (Switanek et al., 2017), CDFt (Michelangeli et al., 2009) and ISIMIP3BASD (Lange, 2019). Further, it develops an evaluation framework for assessing distributional properties and user-defined climate indices (covering but not limited to the ETCCDI indices – Zhang et al., 2011) not only along marginal but also temporal, spatial and multivariate dimensions. Applying *ibicus* in a case study over the Mediterranean region, we find that the most appropriate method indeed depends on the variable and impact studied and that the evaluation of spatiotemporal metrics can identify issues with bias adjustment that would not be found when only marginal, i.e. calibrated aspects are evaluated. Further, we find that even methods that aim to preserve the trend of the climate model can modify it, and that bias adjustment modifies the overall climate model ensemble spread.

40

45

The remainder of this paper is structured as follows. Section 2 gives an introduction to statistical bias correction methodologies and section 3 presents *ibicus*, covering both the details of different bias adjustment methodologies and evaluation metrics implemented, as well as the software design of the package. In section 4, we present the results of the case study before drawing conclusions in section 5.

50

2 Background

2.1 Statistical bias adjustment of climate models

Climate model biases can be defined as "systematic difference between a simulated climate statistic and the corresponding real-world climate statistic" (Maraun, 2016). These biases mostly stem from the imperfect representation of physical processes such as orographic drag, convection, or land-atmosphere interactions. This leads to the incorrect representation of features such as the mean and variance of observed temperature or the spatial properties of extreme rainfall over a certain area.

Bias adjustment methods for climate models have their origin in methods developed for the post-processing of Numerical Weather Prediction (NWP) models. The rationale is to calibrate an empirical transfer function between model simulations and observations over the historical period, that is then applied to the model simulation for the period of interest, often in the future. However, in contrast to NWP models, there is no direct correspondence between the time series of observations and the climate model in historical simulations. This means that typical regression-based approaches used for NWP are not applicable. Rather, properties of the statistical distribution of the two variables, such as the mean or quantiles, are mapped to each other when bias adjusting climate models. Furthermore, the magnitude of biases in climate models can be much larger, whereas NWP forecasts are tightly constrained by recent observations.

The most common approaches to the bias adjustment of climate models include a simple adjustment of the mean (Linear Scaling or Delta Change), a mapping of the two entire cumulative distribution functions (Quantile Mapping), or more advanced methods that also aim to preserve the trend projected in the climate model (such as CDFt or ISIMIP3BASD). The practice of using bias adjustment methods to also downscale the climate model has been criticized in various publications (von Storch, 1999; Maraun, 2013; Switanek et al., 2022), therefore this paper focuses on bias adjustment of climate models purely for the purpose of reducing biases at constant resolution.

The use of bias adjustment methods has become standard practice in academic climate impact studies, and increasingly outside of academia in national assessment reports or other climate services. For example, the ISIMIP3BASD methodology (Lange, 2019) is implemented as the only bias adjustment method as a standard pre-processing step in the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) impact modelling framework that is used in the climate risk scenarios published by central banks (NGFS, 2021). However, applying statistical bias adjustment to climate models raises a number of important considerations and issues which we categorize into *fundamental* and *practical* issues for the purpose of this paper.

2.2 Fundamental issues with statistical bias adjustment and evaluation

Climate model biases in statistics at the grid-cell level can stem from larger-scale biases of the model such as biases in larger drivers such as El Niño, the lack of local feedback to these drivers or the misplacement of storm tracks in a region. However, univariate statistical bias adjustment methods are only as capable as their assumptions and input data and therefore correct only the impact these larger-scale biases have on the distribution of the variables at grid cell level (Maraun et al., 2017).

Univariate bias adjustment might also deteriorate the spatial, temporal or multivariate structure of the climate model. This is particularly problematic for compound events which have been argued to be of particularly high societal relevance due to their

elevated impacts and neglect in standard extreme event evaluation approaches (Zscheischler et al., 2018, 2020). As this issue will not be detected in location-wise cross-validation approaches, it is necessary to evaluate bias adjusted data with a particular focus on spatial, temporal and multi-variable components (Maraun et al., 2017; Maraun and Widmann, 2018).

90   Furthermore, bias adjustment can modify the climate change trend, in particular, that of threshold-sensitive climate indices (Dosio, 2016; Casanueva et al., 2020). This holds overall for non-trend-preserving methods, as well as for trend-preserving methods if underlying assumptions are not met. Trend modification might be justifiable in specific cases (Boberg and Christensen, 2012; Gobiet et al., 2015), but is not justified as a default practice, therefore requiring a decision on a case-by-case basis.

95   There are some options available to cope with these fundamental issues in impact studies: the first is to discard climate models that misrepresent large-scale circulation relevant to the problem at hand. The second is to conduct a careful evaluation of multivariate aspects of the bias adjusted climate model to identify potential artefacts and discard methods that introduce these before proceeding with the impact study. The third is to develop process-informed multivariate bias adjustment methods that for example include large-scale covariates such as weather patterns (Maraun et al., 2017; Verfaillie et al., 2017; Manzanas

100   and Gutiérrez, 2019). These more elaborate methods require an even more careful case-by-case model selection and evaluation.

2.3 Practical issues with bias adjustment and the availability of open-source software

Addressing these fundamental issues and improving the application of bias adjustment is impeded by a number of practical issues.

The first practical issue is that the comparison of different bias adjustment methods and their adaptation to a specific applica-

105   tion is not easily possible for a user. This is because the code to implement different methodologies is published, if at all, across different software packages and languages, impeding interoperability. Users also have the option of downloading already bias adjusted datasets which improves ease of access but does not allow for any custom adjustments (Dobor et al., 2015; Famien et al., 2018; Dumitrescu et al., 2020; Xu et al., 2021). The second practical issue is that available software packages are not accompanied by evaluation methods beyond marginal aspects. As the evaluation of bias adjustment is not straightforward, this

110   makes it difficult for a user to detect artefacts or identify improper results by assessing multivariate properties of the climate model, rendering bias adjustment prone to misuse (Maraun et al., 2017).

These practical issues jeopardize the current implementation of statistical bias adjustment. Addressing these issues does not solve the more fundamental issues but can improve common practice and enhance transparency.

An example of good practice is the MiDAS package which introduces a new bias adjustment method that is compared

115   to other methods in Berg et al. (2022). However, even though the package is in principle extendable, other methods are not implemented in practice, nor is an adjustable evaluation framework developed.

3 ibicus – an open-source software package for bias adjustment

To address the practical issues outlined in the previous section we introduce *ibicus*, an open-source Python package for the bias adjustment of climate models and evaluation thereof. *ibicus* implements eight state-of-the-art, peer-reviewed bias adjustment
120 methodologies in a common interface that enables the user to modify aspects of an individual methodology to suit their target variable, region and impact of interest. The package provides an extensive evaluation framework covering spatial, temporal and multivariate aspects. As part of this, we develop a generalized threshold metric class that allows the user to evaluate both frequently used climate metrics such as frost days or dry days, as well as define their own threshold metrics targeted to the specific impact study. The spatiotemporal evaluation of threshold metrics enables the user to detect artefacts and evaluate
125 compound events before and after bias adjustment. *ibicus* is designed to be flexible and easy to use, facilitating both the "off the shelf" use of methods as well as their customization and allowing use in notebook environments all the way up to the integration with high-performance computing (HPC) packages such as dask (Rocklin, 2015). This chapter provides an overview of the key features of *ibicus*. A more complete user guide and tutorials can be found on the documentation page of the package.

3.1 Data input

130 Bias adjustment requires observational data and climate model simulations during the same historical period and climate model simulation for the (future) period of interest. *ibicus* operates on a numerical level, taking three-dimensional (time, latitude, longitude) numpy arrays as input and returning arrays of the same shape and type. This choice was made to ensure interoperability with different geoscientific computing packages such as xarray (Hoyer and Hamman, 2017) or iris (Met Office, 2010), as well as operation in different computing environments and integration with dask (Rocklin, 2015).

135 ### 3.2 Bias adjustment

ibicus represents each bias adjustment methodology as a class which inherits generic functionalities from a base 'debiaser' class, such as the common initialization interface and a function applying the 'debiaser' in parallel over a grid of locations. The base 'debiaser' class makes the package easily extendable, as a new bias adjustment methodology can inherit these generic functionalities and requires only the specification of a function which applies the methodology for a given location ('ap-
140 ply_location').

Each 'debiaser' object is initialized separately for each variable and requires several class parameters. These are specific to the bias adjustment methodology and include parameters such as the distribution used for a parametric fit or the type of trend preservation applied. For a number of methodology-variable combinations, default settings exist that are described in the documentation. Default settings are labelled 'experimental' if they have not been published in the peer-reviewed literature but
145 are proposed by the package authors after extensive evaluation. It is possible and encouraged to modify the parameters even when default settings exist to adapt the method to a given use case. For example, if precipitation extremes are of special interest, the user could choose to modify the parametric fit for this variable as the gamma distribution – an often used default – might underestimate precipitation extremes (Katz et al., 2002). After initialization, each debiaser object has an 'apply' method to

Table 1. Distinctions between different bias adjustment methods and important considerations motivating the choice.

Statistic / quantiles	Methods for bias adjustment range from simple adjustments to the mean (Linear Scaling – LS) or mean and variance (LS) to adjustments to all quantiles of the distribution.
Parametric or non-parametric Method	Non-parametric methods are restricted to the range of observed/modelled data in their "historical period" and might not handle extremes well, while parametric methods introduce additional assumptions. *ibicus* allows users to implement all methods non-parametrically by modifying method attributes. Based on the default arguments, QM, QDM, ECDFM and SDM are parametric methods while CDFt is non-parametric and ISIMIP3BASD is semi-parametric. For each method using a parametric distribution, it is possible to exchange it with a different one.
Time-window	Some methods include a running window to calculate different transfer functions depending on seasonality (QDM, ISIMIP3BASD, CDFt is applied by month) whilst others do not account for seasonality explicitly.
Trend-preservation and stationarity assumption	Methods such as quantile mapping can modify the trend in the climate model. This might be sensible if the trends are taken to be unrealistic or due to state-dependent biases which need correction (Boberg and Christensen, 2012; Gobiet et al., 2015; Doblas-Reyes et al., 2021). However, in other cases, the trend might be considered credible and should be preserved. Methods can be designed to preserve trends in the mean (DC, LS, dQM), mean and variance (dQM) or all quantiles (CDFt, ECDFM, QDM, ISIMIP3BASD, SDM) - although even then they are not guaranteed to do so. Often trends are distinguished between additive trends (as for temperature) and multiplicative trends (as for precipitation where trends in intensity occur), however not all methods share this distinction. The question of trend preservation is related to the assumption made that the bias is 'stationary'. The assumption is explicitly made by Quantile Mapping. SDM explicitly relaxes the assumption, CDFt and QDM account for it by including a running window over the future period in addition to one over the year.

apply bias adjustment to climate model data. This takes a 3-dimensional numpy array of observations, as well as historical and
150 future climate model simulations as input, together with optional date information for running windows. The apply function
can be run in parallel to speed up execution and integrates with dask for deployment in HPC environments.

Table A1 in the appendix provides an overview of the methodologies currently implemented in *ibicus*. The methods were
chosen to cover the most widely used bias adjustment methods. These methods are based on different assumptions, making
them suitable for different applications. For example, ISIMIP3BASD is a parametric trend preserving quantile mapping which
155 might be appropriate if the variable approximately follows a known parametric structure and the climate change trend in all
quantiles is judged to be realistic. If these assumptions are not valid, a non-parametric method such as CDFt or a non-trend
preserving method such as Quantile Mapping might be more appropriate. Alternatively, if changes in extremes are of special
interest, a parametric method based on extreme value theory might be adequate. In table 1, we highlight the key methodological
considerations differentiating these methods.

Table 2. Attributes of the threshold metrics class.

Threshold Attribute	Description
Name	Name of the threshold
Value(s)	Values defining the threshold (to compare climate model or observations against).
Description	Brief description of the threshold.
Type	Whether values shall fall above, below, outside or between threshold(s).
Scope	Whether the threshold(s) is defined daily, monthly, seasonally, or overall (different for each time category, or not).
Locality	Whether the threshold is defined location-wise or globally (different at each location or not).

3.3 Evaluation

Physical consistency in space, time or between variables is not ensured when using univariate bias adjustment methods. Furthermore, the trend of the climate model might be modified, and the bias of some statistics or impact metrics might be increased through some bias adjustment methods – even if it is removed in certain quantiles. The *ibicus* evaluation framework offers a collection of tools to identify these issues and compare the performance of different bias adjustment methods for variables of interest.

3.3.1 Metrics and design

The evaluation framework consists of two components: 1) the evaluation of bias adjustment on a validation/testing period that enables comparison of the bias adjusted model with observations, and 2) the analysis of trend preservation between the validation and future, or any two future periods. The latter component is necessary as bias adjustment methods can modify the climate change trend, even with methods that are designed to preserve it, as demonstrated by the case study in section 4. In the absence of evidence to the contrary, trend-preserving methods should be preferred as statistical bias adjustment methods usually do not have an underlying physical reasoning for modifying a particular trend.

In both components of the evaluation framework, there are two kinds of metrics that can be evaluated using *ibicus*, termed statistical properties and threshold metrics. Statistical properties allow the user to compare properties of the observational distribution and the climate model distribution - such as the mean or different quantiles - before and after bias adjustment. Threshold-based climate indicators are often of special interest for climate impact studies – for example, frost days, by time of year, could be of interest for agricultural or biodiversity impacts – and where the success of bias adjustment methods is particularly desirable (Dosio et al., 2012; Dosio, 2016). A number of threshold metrics are implemented by default in the package. A new threshold metric can be specified by the user along the dimensions in table 2. Accumulations such as monthly total precipitation can also be estimated. Using these definitions, the evaluation module covers but is not limited to the indices developed by the Expert Team on Climate Change Detection and Indices (ETCCDI - Zhang et al., 2011) used in many application studies.

Table 3. Overview of evaluation categories implemented in ibicus.

	Statistical Properties	Threshold Metrics
Location-wise	**Calculation**: location-wise bias (absolute and percentage) in different distributional properties (quantile, mean) of climate model before and after applying different bias adjustment methods. **Visualization**: boxplot across locations and spatial plot. Plotting functions for visual inspection of observed and climate model distribution (histogram and CDF).	**Calculation**: location-wise bias (absolute in days/year and percentage) in the frequency of singular threshold exceedance events in climate model before and after bias adjustment methods. **Visualization**: boxplot across locations and spatial plot.
Temporal	-	**Calculation**: distribution of spell lengths of threshold exceedances (for example dry spell length). **Visualization**: plot of empirical CDF.
Spatial	RMSE of between spatial correlation matrices at each location.	**Calculation**: distribution of spatial cluster size of threshold exceedances (for example spatial size of heatwaves). **Visualization**: plot of empirical CDF.
Spatio-temporal	-	**Calculation**: distribution of spatiotemporal cluster size of threshold exceedances (for example spatiotemporal extent of heatwaves). **Visualization**: plot of empirical CDF.
Multivariate	**Calculation**: correlations between chosen pair of variables at each location. **Visualization**: spatial plot.	**Calculation**: conditional probability of threshold co-occurrence (such as dry and hot days) in observations and climate model before and after bias adjustment. **Visualization**: boxplot.
Trend	**Calculation**: location-wise bias in the multiplicative or additive trend of a threshold metric or distributional property (mean, quantiles) – percentage change between climate model before and after bias adjustment. **Visualization**: boxplot across locations and spatial plot.	

Since location-wise evaluation is not sufficient to decide whether a bias adjustment method is fit for the use-case, the module offers the functionality to evaluate location-wise, as well as spatiotemporal and multivariate metrics both in terms of threshold metrics and statistical properties. The table 3 gives an overview of the implemented methods.

Finally, different bias adjustment methods rely on different assumptions such as certain parametric distributions providing suitable fits. The evaluation framework includes functions to assess the fit of parametric distributions and the seasonality of the variable to help the user make decisions on how to customize the bias adjustment method to their application.

4   Implementation of ibicus in the Mediterranean region

We demonstrate the comparison and evaluation of different bias adjustment methods by applying *ibicus* over the Mediterranean. Rather than conducting a comprehensive evaluation for a single use case, our aim is to highlight the use-case dependency of the method choice more broadly and hence the necessity of targeted evaluation beyond marginal aspects. We, therefore, choose to limit this case study to the bias adjustment of global climate models, even though specific impact studies often but not always (IPCC, 2021) use higher-resolution models over the target region.

4.1   Data and Methods

We consider the Mediterranean region, between 35-45°N latitude and 18°W to 45°E longitude and apply bias adjustment to seven Coupled Model Intercomparison Project Phase 6 (CMIP6) models, selected based on the use in previous studies in the Mediterranean region (Zappa and Shepherd, 2017; Babaousmail et al., 2022). The chosen models include ACCESS-CM2, CMCC-ESM2, IPSL-CM6A-LR, MIROC6, MPI-ESM1-2-LR, MRI-ESM2-0 and NORESM2-MM. Table B1 in the appendix provides more details on these models. We compare four widely used bias adjustment methods that are implemented in *ibicus*: ISIMIP3BASD (Lange, 2019), applied amongst others by Jägermeyr et al. (2021); Pokhrel et al. (2021) as well as impact models run under the ISIMIP framework), Scaled Distribution Mapping (Switanek et al. (2017), applied amongst others as pre-processing step to assess changes in high impact weather events over the UK in Hanlon et al. (2021)), as well as Quantile Mapping (applied in impact studies such as Babaousmail et al., 2022) and Linear Scaling as reference methods. These four methods are applied to daily total precipitation (pr) and daily minimum near-surface air temperature (tasmin), chosen to cover two different types of variables (bounded vs unbounded, different distributions etc) that are both highly relevant for many impact studies. We use ERA5 reanalysis data (Hersbach et al., 2020) as an observational reference, conservatively regridded to match the resolution of the selected climate models. The data ranges from January 1st, 1959 to December 31st, 2005, with the initial 30-year period (1959-1989) serving as the historical/reference period and used as a training dataset and the subsequent 15-year period (1990-2005) used for validation purposes.

We demonstrate four bespoke impact metrics related to daily minimum temperature and daily total precipitation, defined using the *ibicus* threshold metrics class.

- tasmin < 10°C (283.15K) which was chosen based on Droulia and Charalampopoulos (2022) who estimate climate impacts to viniculture noting that above >10°C grapevines are in their optimal photosynthesis zone.

- tasmin greater than the seasonal 95th percentile of the temperature in each grid cell during the historical period (1959-1989). This can be an indicator of the impacts of heatwaves (Raei et al., 2018).

- Dry days (daily precipitation <1mm) and very wet days (daily precipitation >10mm) as two ETCCDI indices.
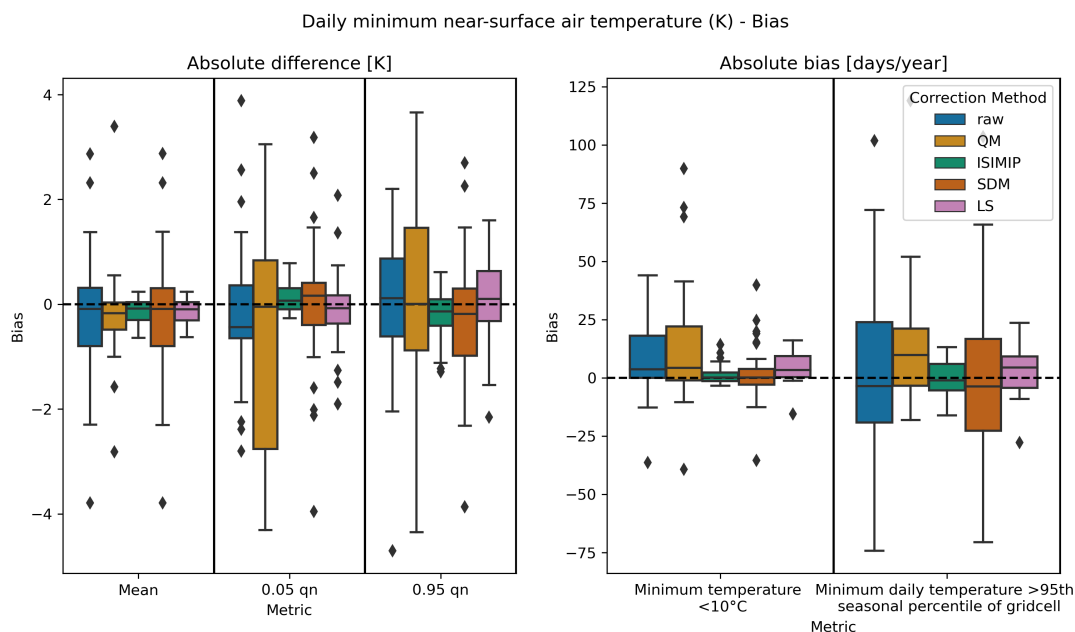
Figure 1. Distribution across locations of marginal minimum daily temperature bias of the ACCESS-CM2 climate model before bias adjustment (raw) and after applying the ISIMIP3BASD bias adjustment method (ISIMIP), Quantile Mapping (QM), Scaled Distribution Mapping (SDM) and Linear Scaling (LS). The left panel displays the distribution of the absolute bias (in Kelvin) in the mean and 0.05 and 0.95 quantiles. The right panel displays the distribution of the absolute bias in the threshold metrics: minimum daily temperature below 10°C and minimum daily temperature above the 95th seasonal percentile defined for this grid cell, both in units of days per year. Bias (location-wise) is defined as the difference between the metric for the (bias adjustment) climate model in the validation period and the metric for the observational data in the validation period (in each grid cell, metrics calculated in the temporal dimension). This figure shows the standard *ibicus* output distribution of location-wise bias for a set of specified statistics and threshold metrics. The boxplot shows the median, the first and third quartiles as a box, the outer range (defined as Q1 - 1.5 × IQR and Q3 + 1.5 × IQR) as whiskers, and any points beyond this as diamonds.

4.2 Results

4.2.1 Evaluation of the location-wise bias on the validation period

220 Figures 1-3 show the marginal bias of the climate model with respect to observations over the validation period before and after bias adjustment across locations in the study area.

We find that most methods reduce but do not eliminate the marginal bias in the mean, shown for the ACCESS-CM2 model and minimum daily temperature in figure 1, while the range of reduction is varied: ISIMIP and Linear Scaling achieve more significant reductions in the bias than Quantile Mapping or Scaled Distribution Mapping. In extremal quantiles and threshold

225 metrics however, certain methods can also inflate the bias of the raw climate model even if the bias in the mean is reduced, as we see most prominently for quantile mapping in figures 1 and 2.
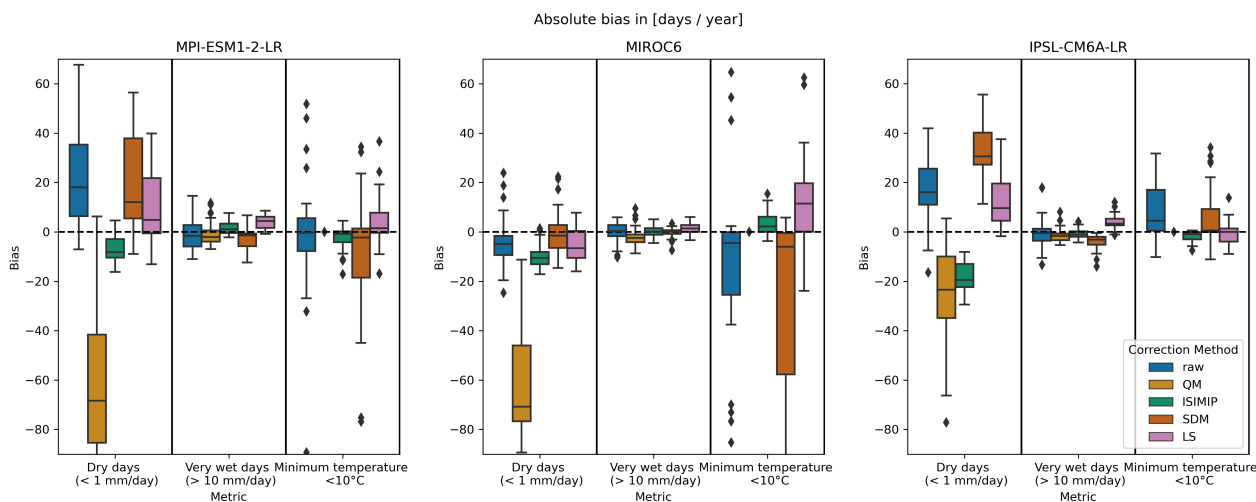
Figure 2. Distribution of marginal bias across locations before bias adjustment (raw) and after applying the ISIMIP3BASD bias adjustment method (ISIMIP), Quantile Mapping (QM), Scaled Distribution Mapping (SDM) and Linear Scaling (LS). Three climate models (MPI-ESM1-2-LR, MIROC6 and IPSL-CM6A-LR) and three threshold metrics (minimum daily temperature below 10°C, dry days defined as total precipitation below 1mm and very wet days defined as total precipitation above 10mm) are evaluated. The bias in minimum temperature <10°C of the climate models after applying quantile mapping is particularly large, exceeding 300%. For improved readability of the plot, we have omitted this bias adjustment - metric combination here but show the full plot in the appendix.

Furthermore, in figure 2 we see that that the success of a bias adjustment method depends on the use case, meaning the variable, metric and climate model studied. While Scaled Distribution Mapping somewhat reduces the bias in dry days for two of the climate models, it inflates the bias in dry days for the third. On the other hand, the method reduces bias in the minimum temperature threshold metric for the IPSL-CM6A-LR model but inflates the bias in this metric for the MIROC6 model. ISIMIP3BASD on the other hand reduces the bias in dry days for the MPI-ESM1-2-LR model but increases it for the MIROC6 model. Quantile Mapping performs reasonably well for the wet-day metric but quite badly for the dry-day and minimum temperature metrics. These differences in the performance of bias adjustment methods can be due to their assumptions (a parametric distribution fit might not replicate the correct tail behaviour), and method (whether they are tailored to a specific variable or whether event frequency adjustment is implemented), as well as the physical source of the bias in the climate model.

When investigating the spatial distribution of the bias (figure 3), we find that certain methods can assimilate the spatial pattern of the bias across climate models. For example, linear scaling (LS) shifts climate models to an overestimation of very wet days in similar regions, even models like NORESM2-MM which previously underestimated these days. In other cases, methods can perform well in certain regions, but not in others. Quantile mapping (QM) seems to perform reasonably well over the Iberian peninsula, but has difficulties over Italy, especially for MPI-ESM1-2-LR where a strong underestimation is shifted
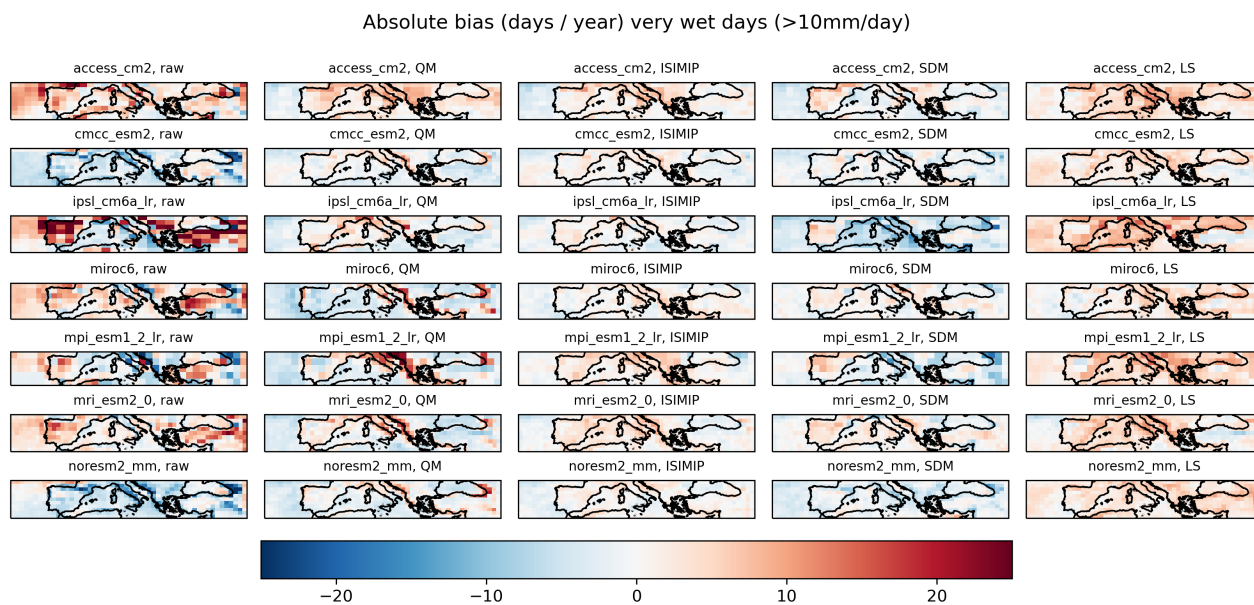
Figure 3. Spatial plot of marginal absolute bias in very wet days defined as total precipitation above 10mm given in [days / year]. Results are shown for seven climate models (ACCESS-CM2, CMCC-ESM2, IPSL-CM6A-LR, MIROC6, MPI-ESM1-2-LR, MRI-ESM2-0 and NORESM2-MM) before bias adjustment (raw) and after applying the ISIMIP3BASD bias adjustment method (ISIMIP), Quantile Mapping (QM), Scaled Distribution Mapping (SDM) and Linear Scaling (LS).

into a strong overestimation. This highlights the importance of investigating the spatial distribution of the marginal bias as this varies across the different regions in the Mediterranean.

4.2.2 Evaluation of the bias in spatiotemporal characteristics on the validation period

245 Moving on to the investigation of spatiotemporal characteristics, figures 4 and 5 show the cumulative distribution of spell length and spatial extent for the dry-day and minimum temperature heatwave days metric, respectively. The plots depict the standard visualization output that the *ibicus* software package produces for this type of evaluation.

The spatiotemporal characteristics investigated exhibit biases between the reanalysis data and raw climate model output. For example, it is ∼1.6 times more likely for a dry spell to exceed 20 days in the raw climate model IPSL-CM6A-LR compared to 250 the reanalysis data.

We find that the bias in these spatiotemporal metrics can be reduced with some bias adjustment methods: for example, ISIMIP3BASD reduces the spell length bias for dry days, and Scaled Distribution Mapping reduces the bias in both spell length and spatial extent for minimum temperature heatwave days. However, this result is again inconsistent across methods and variables, and different bias adjustment methods frequently appear to increase the spatiotemporal bias: Scaled Distribution
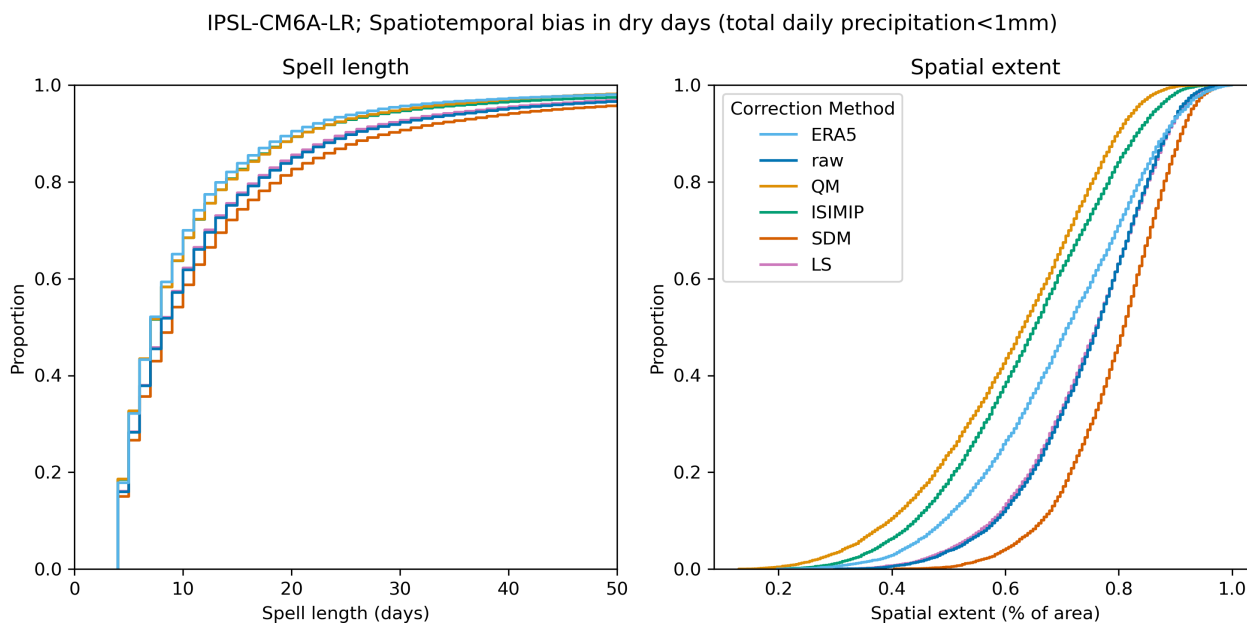
Figure 4. Cumulative distribution functions of spell length (left panel) and spatial extent of dry days (right panel). The spell length is defined as the length of a temporal sequence longer than three days during which a single grid cell exceeds the specified threshold. The spatial extent is defined as the fraction of cells exceeding the specified threshold, given that a single cell exceeds the threshold. This plot shows the cumulative distribution function of individual spell lengths and spatial extents at single points in time across the entire Mediterranean region in the observational data (ERA5), in the climate model IPSL-CM6A-LR before bias adjustment (raw) and after applying the ISIMIP3BASD bias adjustment method (ISIMIP), Quantile Mapping (QM), Scaled Distribution Mapping (SDM) and Linear Scaling (LS).

Mapping increases the bias in spell length and spatial extent of dry days, as do Quantile Mapping and ISIMIP3BASD when investigating the spatial extent.

These results are to some extent expected, as the selected methods are univariate methods, meaning they are calibrated location-wise and do not incorporate spatiotemporal information. However, the results highlight the need to evaluate how bias adjustment changes spatiotemporal characteristics, as these are often implicitly used in impact downstream impact studies.

4.2.3 Evaluation of the climate change trend before and after bias adjustment

As mentioned in the background section, the modification of the climate change signal through bias adjustment has been reported and discussed in various publications and stimulated the development of methods that aim to preserve the climate signal.

In the analysis of the dry day trend, shown in figure 6, we find that a non-trend-preserving method such as quantile mapping significantly alters the climate change trend. The axes in figure 6 were limited to +-100 for the sake of readability, however,
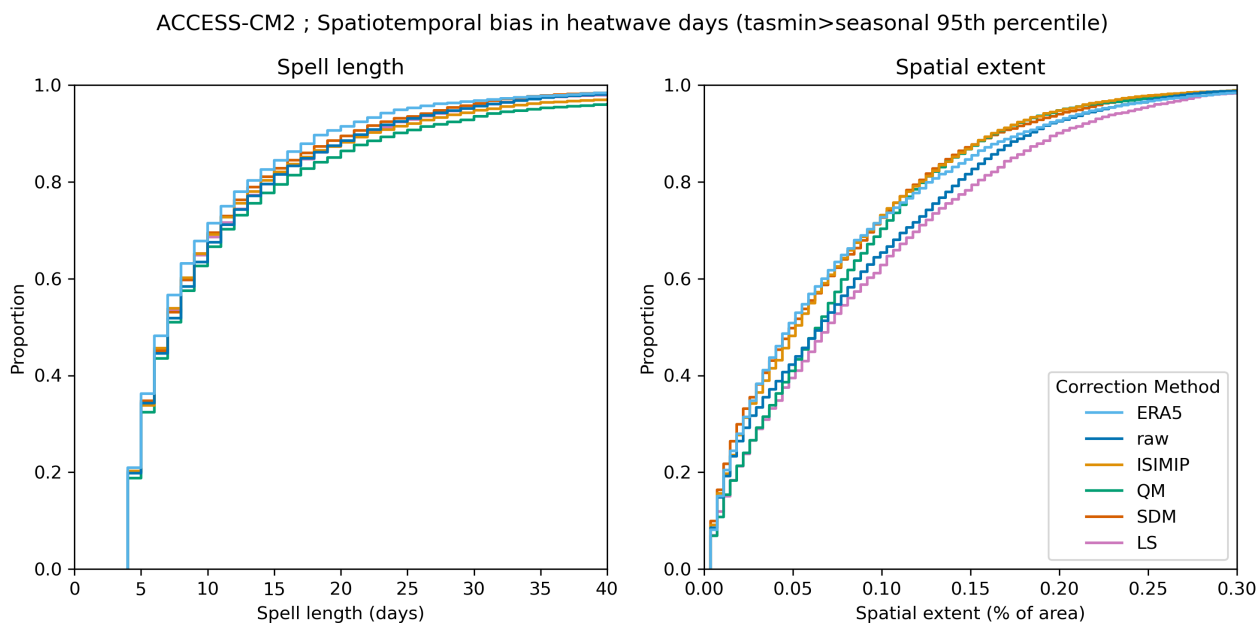
13

Figure 5. As Figure 4, but investigating the threshold of minimum daily temperature exceeding its 95th seasonal percentile defined per grid cell for the climate model ACCESS-CM2.

a limited number of data points show even larger biases after applying quantile mapping. The unrestricted version of this plot can be found in the appendix.

We also find that methods that aim to preserve the trend such as ISIMIP3BASD or Scaled Distribution Mapping modify it up to 50% at some locations. For the ISIMIP method, this is presumably due to the fact that the 'future observations' through which the trend preservation is implemented are mapped using empirical CDFs, whereas the bias adjustment itself is parametric. It has been argued that the normal distribution for temperature or the gamma distribution for precipitation might not adequately capture the tail behaviour of these variables (Katz et al., 2002; Nogaj et al., 2006; Sippel et al., 2015; Naveau et al., 2016). This is particularly relevant when investigating the trend of high or low quantiles, as well as threshold metrics that do not sit at the centre of the distribution. Additionally, for bounded variables such as precipitation, the frequency beyond two outer thresholds is adjusted separately in the ISIMIP3BASD methodology which could lead to the change in the dry day trend shown in figure 6.

We find a much smaller change in the trend of the mean minimum daily temperature across methods, shown in figure 7. In fact, linear scaling barely modifies the trend at all, which is to be expected since the method only subtracts the mean bias from the future and the validation period, based on the strong assumption that the bias affects the mean only and is stationary over time.
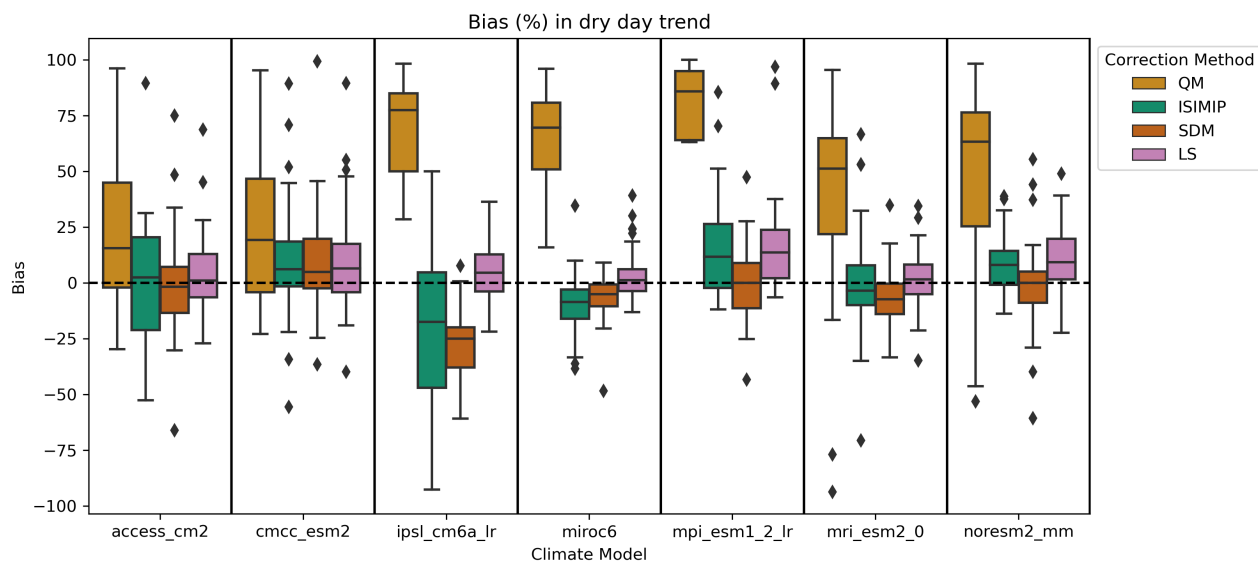
Figure 6. Distribution of location-wise biases in the additive climate change trend in dry days introduced through the bias adjustment method, computed by computing the additive trend between the validation period and the future period in both the raw and the bias adjusted model and taking the percentage difference between the two trends.
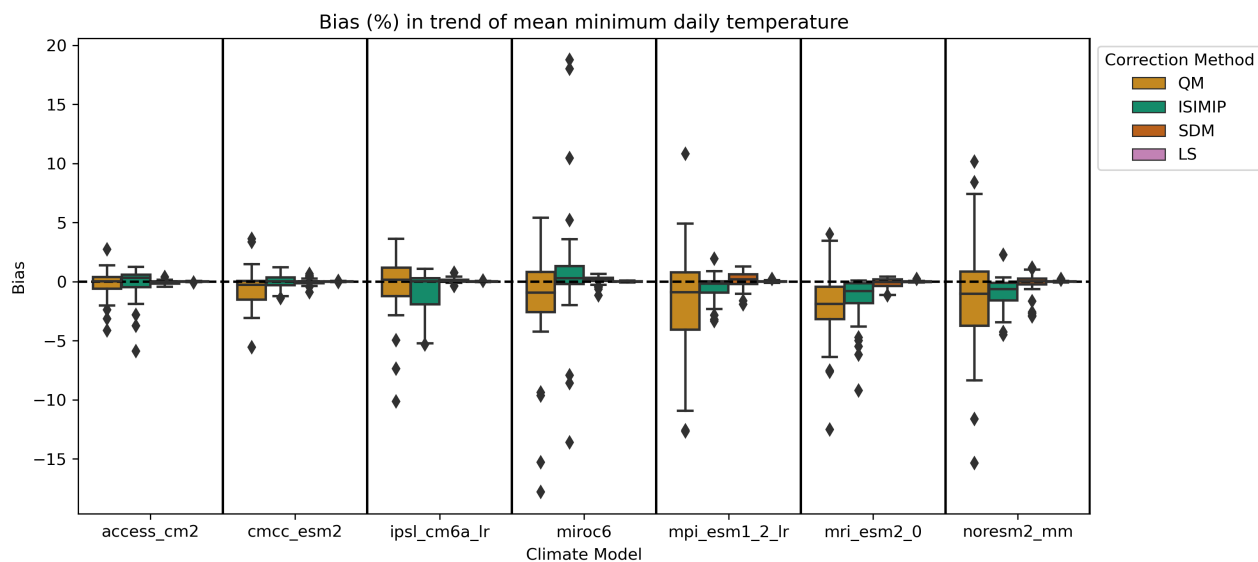

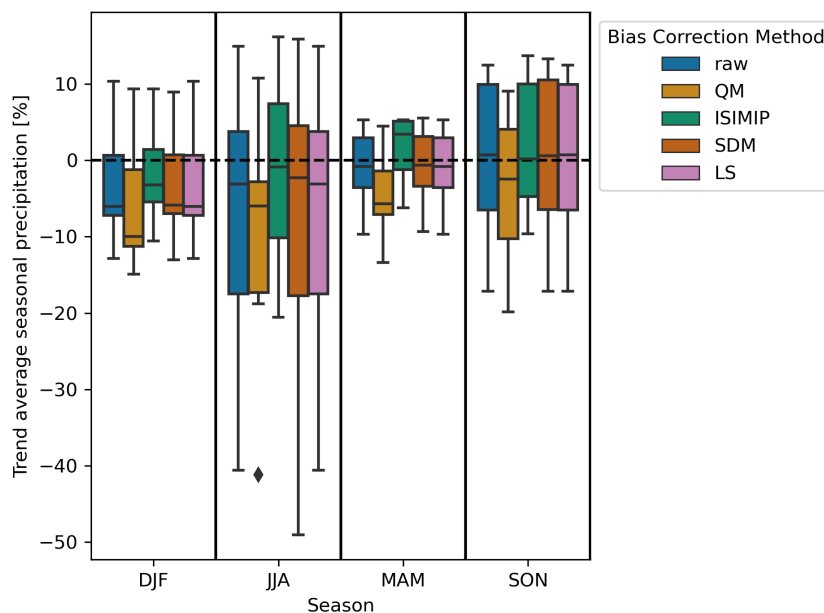**Figure 7.** As Figure 6 but for the trend in mean minimum daily temperature.

Figure 8. Ensemble spread of seven selected climate models (ACCESS-CM2, CMCC-ESM2, IPSL-CM6A-LR, MIROC6, MPI-ESM1-2-LR, MRI-ESM2-0 and NORESM2-MM), showing the trend in average seasonal precipitation between the validation and future period, without applying bias adjustment (raw) and after applying ISIMIP3BASD, Quantile Mapping and Scaled Distribution Mapping.

4.2.4 Evaluation of the uncertainty in the climate model ensemble before and after bias adjustment

Figure 8 shows that the climate model ensemble spread of the trend in mean seasonal precipitation is modified when applying bias adjustment. This means that not only the trend but also the range of uncertainty and possible worst-case scenarios analysed in impact studies depend on the bias adjustment method used to pre-process the climate model. As shown in the previous

285 sections, the 'best' bias adjustment method for a given use case depends on the variable, region and impact variable studied. The result shown in figure 8 demonstrates that bias adjustment can add an additional source of uncertainty if the method is applied blindly and not evaluated properly. Interestingly the uncertainty range is not necessarily narrowed as has been postulated by some authors (Ehret et al., 2012), but even extended and shifted in some cases.

5 Conclusions

290 Statistical bias adjustment is a useful method when working with climate models to understand future climate impacts. However, there are fundamental as well as practical issues in how bias adjustment is currently used both in academic research and by practitioners in the private and government sector. One practical issue impeding good practice is the availability of open-source software to compare different bias adjustment methods and evaluate non-calibrated aspects.

This paper demonstrates that the success of a bias adjustment method depends on the variable and impact studied, and bias
295 adjustment should therefore be evaluated and adapted targeted to the region and use-case at hand. Depending on the climate
model and variable of interest different methods can reduce or also increase biases by a large range, can impair or leave
spatiotemporal coherence relatively unaffected. This is non-systematic across bias adjustment methods, climate models and
variables/metrics of interest. Furthermore, we find that even trend-preserving methods can modify the trend in statistical prop-
erties and climate indices, and each bias adjustment method changes the climate model ensemble spread slightly differently.

300 With the Python package *ibicus*, we aim to provide a resource to address these practical issues and for better and more
targeted use of bias adjustment. By facilitating the easy comparison and extensive evaluation of different methods, we hope
to contribute to a change in practice in the use of bias adjustment and mainstream use-case dependent and evaluated bias
adjustment.

So far, the package implements univariate bias adjustment methods, meaning that the bias adjustment is calibrated and
305 applied on each grid point separately. Multivariate bias adjustment methods that correct spatial, temporal, or inter-variable
structures next to marginal aspects have been published, amongst others by Piani and Haerter (2012); Vrac and Friederichs
(2015); Sippel et al. (2016); Cannon (2016, 2018); Vrac (2018); François et al. (2020). As bias adjustment can also impair the
relationship between large-scale circulation and local feedback, more research on processes-based bias adjustment is needed to
explicitly capture these relationships (Maraun et al., 2017; Verfaillie et al., 2017; Manzanas and Gutiérrez, 2019). We have so
310 far chosen to focus on univariate methods as the need for careful model selection and evaluation becomes even more pertinent
when using multivariate methods. Our aim was therefore to first establish a robust workflow and evaluation for simpler, yet
still widely used, univariate methods to address the practical issues in the application of bias adjustment. However, the package
remains under active development and maintenance and we would like to invite collaboration from the community to extend
and further develop its functionalities. It is available via standard channels like *PyPI* with extensive documentation on *Read
315 the Docs*.

Both the choice of validation period as well as the choice of observational dataset and uncertainty therein have been shown
to affect the results of bias adjustment (Casanueva et al., 2020). This is not explicitly explored in this publication or package,
however, the evaluation tools available through *ibicus* enable the investigation of these issues. Furthermore, the implications
of bias adjustment on the uncertainty of the climate model ensemble spread should be explored further and accounted for in
320 impact studies. Fundamentally however the performance of a bias adjusted climate model is always dependent on the driving
climate model itself. Even the best bias adjustment cannot correct a fundamental misrepresentation of physical processes in a
GCM or RCM and continued research on improving these is necessary.

tions presented in this paper (doi:10.5281/zenodo.8101842, Wessel and Spuler, 2023). The ERA5 and CMIP6 data used were accessed via the Copernicus Climate Data Store under the Copernicus licence: https://doi.org/10.24381/cds.143582cf and https://doi.org/10.24381/cds.c866074c respectively.

Table A1. Bias adjustment methods currently implemented in ibicus with variables covered and details on their functioning. Here x refers to observations x_{obs} or climate model values during the historical / reference $x_{\text{cm, hist}}$ or future period $x_{\text{cm, fut}}$ and F to a Cumulative Distribution Function (CDF) fitted either parametrically or non-parametrically. Covered variables indicate variables for which the bias adjustment method currently has default settings and climatic variables with a * are variables with experimental default settings. Those are settings that were not published in the peer-reviewed literature but were found to give good performance.

Name	References	Details
ISIMIP3BASD	Hempel et al. (2013); Lange (2019, 2021a)	**Method**: semi-parametric quantile mapping-based method that aims to be trend-preserving in all quantiles. Generates "pseudo future observations" by applying the models' climate change trend to observations either additively, multiplicatively or in an alternative way. Applies quantile-mapping between the modelled future values and the pseudo future observations, either non-parametrically or parametrically, depending on the variable, optionally with an event likelihood adjustment as in Switanek et al. (2017). The core method is applied in a running window to account for seasonality, and trends in both observations and model are removed prior to applying the method. **Covered variables**: hurs, pr, prsnratio, psl, rlds, rsds, sfcWind, tas, tasrange, taskew.
CDFt	Michelangeli et al. (2009); Vrac et al. (2012, 2016); Famien et al. (2018)	**Method**: non-parametric quantile mapping that aims to be trend-preserving in all quantiles. CDFt constructs a CDF of future observations and then applies a quantile mapping between the CDF of the future climate model values and the CDF of the future observations: $$x_{\text{cm, fut}} \rightarrow F_{\text{obs, fut}}^{-1}\big(F_{\text{cm, fut}}(x_{\text{cm, fut}})\big) = F_{\text{cm, fut}}^{-1}\big(F_{\text{cm, hist}}\big(F_{\text{obs, hist}}^{-1}(F_{\text{cm, fut}}(x_{\text{cm, fut}}))\big)\big).$$ Because non-parametric CDFs will not be able to map values outside the range of the data an additive or multiplicative shift can be applied to the future and historical climate model data prior to fitting CDFs: the additive or multiplicative bias in the mean can be subtracted / divided out first. CDFt can be run separately for each month of the year to account for seasonality as well as in a running window over the future period, to smooth discontinuities and relax the stationarity assumption. To correct precipitation occurrences in addition to amounts Stochastic Singularity Removal (Vrac et al., 2016) can be applied. **Covered variables**: hurs*, pr, psl*, rlds*, rsds*, sfcwind*, tas, tasmin, tasmax, tasrange*, taskew*.
Scaled Distribution Matching (SDM)	Switanek et al. (2017)	**Method**: parametric quantile mapping that aims to be trend-preserving in all quantiles. Conceptually similar to Quantile Delta Mapping and ECDFM. Scales CDFs by projected absolute (temperature) or relative (precipitation) changes, whilst at the same time also adjusting the likelihood of individual events, by adjusting return intervals, prior to mapping. **Covered variables**: pr, tas, tasmin*, tasmax*.

Table A2. Table A1 cont.

Name	References	Details
(Detrended) Quantile Mapping (dQM)	Cannon et al. (2015); Maraun (2016)	**Method**: quantile by quantiles mapping of observational and climate model distribution. Forms the basis of most other methods listed. Trends in the mean can be adjusted for using detrended quantile mapping, removing trends before quantile mapping and reapplying them afterwards, either additively or multiplicatively. $$x_{\text{cm, fut}} \rightarrow F_{\text{obs}}^{-1}\big(F_{\text{cm, hist}}(x_{\text{cm, fut}})\big).$$ **Covered variables**: hurs*, pr, psl*, rlds*, sfcWind*, tas, tasmin*, tasmax*.
Quantile Delta Mapping (QDM) / Equidistant CDF Matching (ECDFM)	Li et al. (2010); Cannon et al. (2015)	**Method**: parametric quantile mapping methods that aim to be trend preserving in all quantiles, with special focus on high quantiles. Quantile Delta Mapping applies the following transformation to the future climate model values $x_{\text{cm, fut}}$ if relative changes are to be preserved (eg. for precipitation): $$x_{\text{cm, fut, bc}}(t) = x_{\text{cm, fut}}(t) \cdot \frac{F_{\text{obs}}^{-1}(\hat{F}_{\text{cm, fut}}^{(t)}(x_{\text{cm, fut}}(t)))}{F_{\text{cm, hist}}^{-1}(\hat{F}_{\text{cm, fut}}^{(t)}(x_{\text{cm, fut}}))},$$ and the following for absolute changes (eg. for temperature): $$x_{\text{cm, fut, bc}}(t) = x_{\text{cm, fut}}(t) + F_{\text{obs}}^{-1}(\hat{F}_{\text{cm, fut}}^{(t)}(x_{\text{cm, fut}}(t))) - F_{\text{cm, hist}}^{-1}(\hat{F}_{\text{cm, fut}}^{(t)}(x_{\text{cm, fut}})).$$ Quantile Delta Mapping for absolute changes is equivalent to the ECDFM method by Li et al. (2010), however the parameters chosen, especially the distributions used for the CDF fits are different. In Quantile Delta Mapping the CDF for future climate model values is fitted in a running window going over the future period to account for long term changes in the trend. Also a running window over the year is included to account for seasonality. This is not the case for ECDFM. **Covered variables**: hurs*, pr, psl*, rlds*, sfcwind*, tas, tasmin*, tasmax*.
Linear Scaling (LS)	Maraun (2016)	**Method**: simple correction method adjusting biases in the mean (additive case): $$x_{\text{cm, fut}} \rightarrow x_{\text{cm, fut}} - \big(\bar{x}_{\text{cm, hist}} - \bar{x}_{\text{obs}}\big),$$ or mean and variance (multiplicative case): $$x_{\text{cm, fut}} \rightarrow x_{\text{cm, fut}} \cdot \frac{\bar{x}_{\text{obs}}}{\bar{x}_{\text{cm, hist}}}.$$ **Covered variables**: hurs*, pr, psl*, rlds*, rsds*, sfcWind*, tas, tasmin, tasmax.
Delta Change (DC)	Maraun (2016)	**Method**: technically not a bias adjustment method. Adds a climate model trend to observations either additively or multiplicatively. Similar to Linear Scaling, however it adjusts the observations and not the climate model. **Covered variables**: hurs*, pr, psl*, rlds*, rsds*, sfcWind*, tas, tasmin, tasmax.

Table B1. Overview of CMIP6 models and their model developers used in the case study in chapter 4.

| Model Name	Institution
ACCESS-CM2	Commonwealth Scientific and Industrial Research Organisation / Australia
CMCC-ESM2	Euro-Mediterranean Centre on Climate Change / Italy
IPSL-CM6A-LR	Institut Pierre-Simon Laplace / France
MIROC6	Japan Agency for Marine-Earth Science, University of Tokyo, National Institute for Environmental and RIKEN Centre for Computational Science / Japan
MPI-ESM1-2-LR	Max Planck Institute for Meteorology / Germany
MRI-ESM2-0	Meteorological Research Institute / Japan
NORESM2-MM	Norwegian Climate Centre / Norway
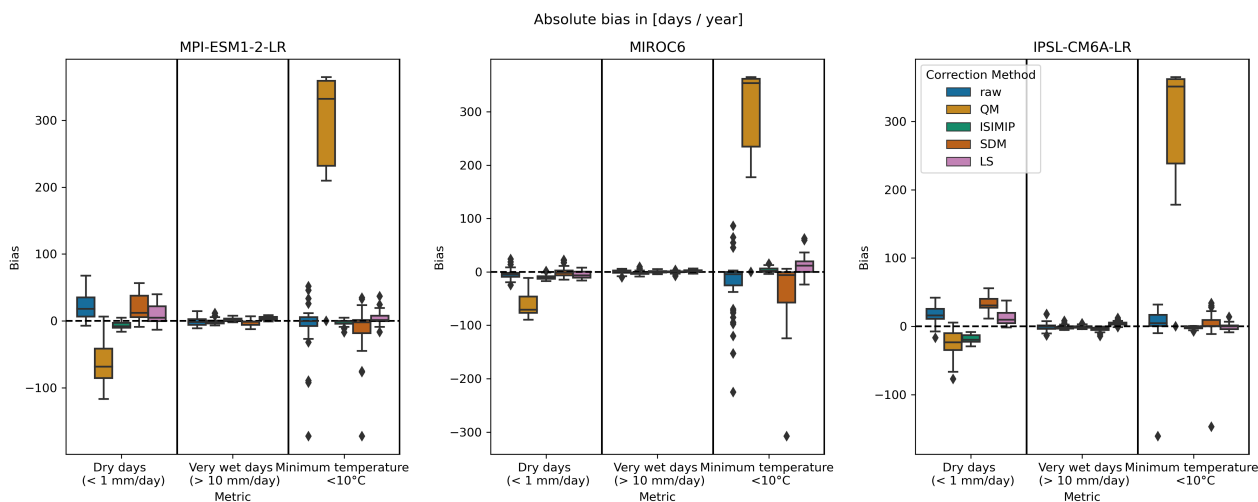
Figure A1. As figure 2, but including Quantile Mapping for minimum temperature.
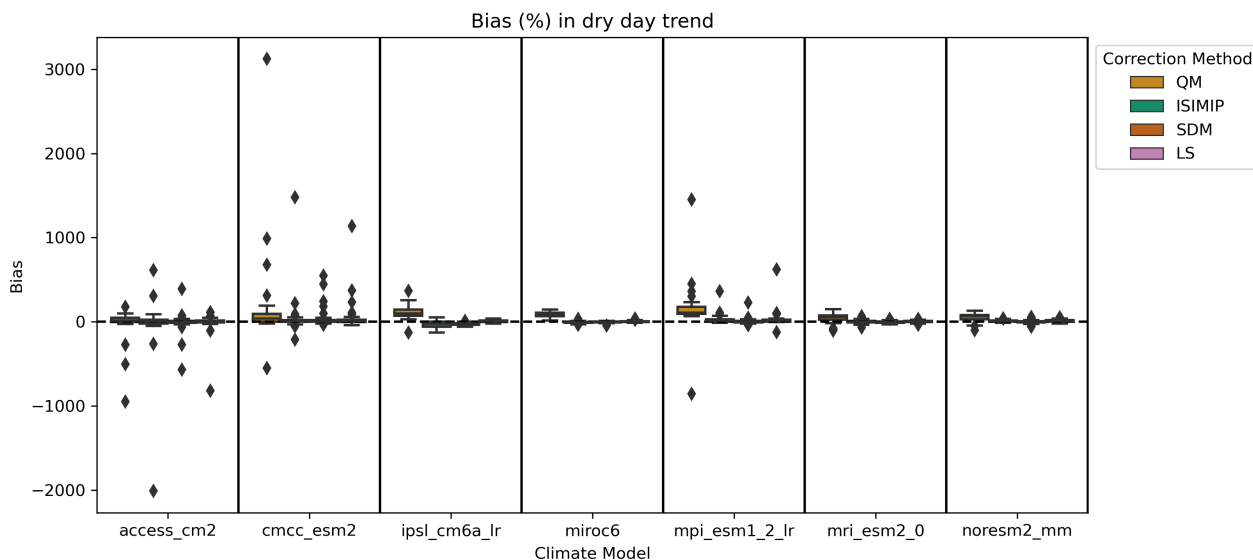

Figure A2. As figure 6, but without axis limits at +-100.

References

Babaousmail, H., Hou, R., Ayugi, B., Sian, K. T. C. L. K., Ojara, M., Mumo, R., Chehbouni, A., and Ongoma, V.: Future changes in mean and extreme precipitation over the Mediterranean and Sahara regions using bias-corrected CMIP6 models, International Journal of Climatology, 42, 7280–7297, https://doi.org/10.1002/joc.7644, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.7644, 2022.

Berg, P., Bosshard, T., Yang, W., and Zimmermann, K.: MIdASv0.2.1 – MultI-scale bias AdjuStment, Geoscientific Model Development,
350   15, 6165–6180, https://doi.org/10.5194/gmd-15-6165-2022, publisher: Copernicus GmbH, 2022.

Boberg, F. and Christensen, J. H.: Overestimation of Mediterranean summer temperature projections due to model deficiencies, Nature
Climate Change, 2, 433–436, https://doi.org/10.1038/nclimate1454, number: 6 Publisher: Nature Publishing Group, 2012.

Cannon, A. J.: Multivariate Bias Correction of Climate Model Output: Matching Marginal Distributions and Intervariable Dependence
Structure, Journal of Climate, 29, 7045–7064, https://doi.org/10.1175/JCLI-D-15-0679.1, publisher: American Meteorological Society
355   Section: Journal of Climate, 2016.

Cannon, A. J.: Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model
simulations of multiple variables, Climate Dynamics, 50, 31–49, https://doi.org/10.1007/s00382-017-3580-6, 2018.

Cannon, A. J.: MBC: Multivariate Bias Correction of Climate Model Outputs, https://cran.r-project.org/web/packages/MBC/index.html,
2023.

360   Cannon, A. J., Sobie, S. R., and Murdock, T. Q.: Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods
Preserve Changes in Quantiles and Extremes?, Journal of Climate, 28, 6938–6959, https://doi.org/10.1175/JCLI-D-14-00754.1, 2015.

Casanueva, A., Herrera, S., Iturbide, M., Lange, S., Jury, M., Dosio, A., Maraun, D., and Gutiérrez, J. M.: Testing bias adjustment methods
for regional climate change applications under observational uncertainty and resolution mismatch, Atmospheric Science Letters, 21, e978,
https://doi.org/10.1002/asl.978, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asl.978, 2020.

365   Doblas-Reyes, F., Sörensson, A., Almazroui, M., Dosio, A., Gutowski, W., Haarsma, R., Hamdi, R., Hewitson, B., Kwon, W.-T., Lamptey,
B., Maraun, D., Stephenson, T., Takayabu, I., Terray, L., Turner, A., and Zuo, Z.: Linking Global to Regional Climate Change, in:
Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Inter-
governmental Panel on Climate Change, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud,
N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi,
370   O., Yu, R., and Zhou, B., pp. 1363–1512, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA,
https://doi.org/10.1017/9781009157896.012, type: Book Section, 2021.

Dobor, L., Barcza, Z., Hlásny, T., Havasi, , Horváth, F., Ittzés, P., and Bartholy, J.: Bridging the gap between climate mod-
els and impact studies: the FORESEE Database, Geoscience Data Journal, 2, 1–11, https://doi.org/10.1002/gdj3.22, _eprint:
https://onlinelibrary.wiley.com/doi/pdf/10.1002/gdj3.22, 2015.

375   Dosio, A.: Projections of climate change indices of temperature and precipitation from an ensemble of bias-adjusted high-
resolution EURO-CORDEX regional climate models, Journal of Geophysical Research: Atmospheres, 121, 5488–5511,
https://doi.org/10.1002/2015JD024411, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2015JD024411, 2016.

Dosio, A., Paruolo, P., and Rojas, R.: Bias correction of the ENSEMBLES high resolution climate change projections for use by impact mod-
els: Analysis of the climate change signal, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2012JD017968,
380   _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2012JD017968, 2012.

Droulia, F. and Charalampopoulos, I.: A Review on the Observed Climate Change in Europe and Its Impacts on Viticulture, Atmosphere, 13,
837, https://doi.org/10.3390/atmos13050837, number: 5 Publisher: Multidisciplinary Digital Publishing Institute, 2022.

Dumitrescu, A., Amihaesei, V.-A., and Cheval, S.: RoCliB– bias-corrected CORDEX RCMdataset over Romania, Geoscience Data Journal,
n/a, https://doi.org/10.1002/gdj3.161, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/gdj3.161, 2020.

385   Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions "Should we apply bias correction to global and regional climate model data?", Hydrology and Earth System Sciences, 16, 3391–3404, https://doi.org/10.5194/hess-16-3391-2012, publisher: Copernicus GmbH, 2012.

Famien, A. M., Janicot, S., Ochou, A. D., Vrac, M., Defrance, D., Sultan, B., and Noël, T.: A bias-corrected CMIP5 dataset for Africa using the CDF-t method – a contribution to agricultural impact studies, Earth System Dynamics, 9, 313–338, https://doi.org/10.5194/esd-9-313-2018, publisher: Copernicus GmbH, 2018.

François, B., Vrac, M., Cannon, A. J., Robin, Y., and Allard, D.: Multivariate bias corrections of climate simulations: which benefits for which losses?, Earth System Dynamics, 11, 537–562, https://doi.org/10.5194/esd-11-537-2020, publisher: Copernicus GmbH, 2020.

Gobiet, A., Suklitsch, M., and Heinrich, G.: The effect of empirical-statistical correction of intensity-dependent model errors on the temperature climate change signal, Hydrology and Earth System Sciences, 19, 4055–4066, https://doi.org/10.5194/hess-19-4055-2015, publisher: Copernicus GmbH, 2015.

Hanlon, H. M., Bernie, D., Carigi, G., and Lowe, J. A.: Future changes to high impact weather in the UK, Climatic Change, 166, 50, https://doi.org/10.1007/s10584-021-03100-5, 2021.

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., and Piontek, F.: A trend-preserving bias correction – the ISI-MIP approach, Earth System Dynamics, 4, 219–236, https://doi.org/10.5194/esd-4-219-2013, publisher: Copernicus GmbH, 2013.

400   Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3803, 2020.

Hoyer, S. and Hamman, J.: xarray: N-D labeled Arrays and Datasets in Python, Journal of Open Research Software, 5, 10, https://doi.org/10.5334/jors.148, number: 1 Publisher: Ubiquity Press, 2017.

IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, vol. In Press, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/9781009157896, type: Book, 2021.

IPCC: Climate Change 2022: Impacts, Adaptation and Vulnerability, Summary for Policymakers, Cambridge University Press, Cambridge, UK and New York, USA, type: Book, 2022.

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

Jägermeyr, J., Müller, C., Ruane, A. C., Elliott, J., Balkovic, J., Castillo, O., Faye, B., Foster, I., Folberth, C., Franke, J. A., Fuchs, K., Guarin, J. R., Heinke, J., Hoogenboom, G., Iizumi, T., Jain, A. K., Kelly, D., Khabarov, N., Lange, S., Lin, T.-S., Liu, W., Mialyk, O., Minoli, S., Moyer, E. J., Okada, M., Phillips, M., Porter, C., Rabin, S. S., Scheer, C., Schneider, J. M., Schyns, J. F., Skalsky, R., Smerald, A., Stella, T., Stephens, H., Webber, H., Zabel, F., and Rosenzweig, C.: Climate impacts on global agriculture emerge earlier in new generation of climate and crop models, Nature Food, 2, 873–885, https://doi.org/10.1038/s43016-021-00400-y, number: 11 Publisher: Nature Publishing Group, 2021.

Katz, R. W., Parlange, M. B., and Naveau, P.: Statistics of extremes in hydrology, Advances in Water Resources, 25, 1287–1304, https://doi.org/10.1016/S0309-1708(02)00056-8, 2002.

Lange, S.: Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0), Geoscientific Model Development, 12, 3055–3070, https://doi.org/10.5194/gmd-12-3055-2019, publisher: Copernicus GmbH, 2019.

Lange, S.: ISIMIP3b bias adjustment fact sheet, Tech. rep., Inter-Sectoral Impact Model Intercomparison Project, 2021a.

Lange, S.: ISIMIP3BASD, https://doi.org/10.5281/zenodo.4686991, 2021b.

Li, H., Sheffield, J., and Wood, E. F.: Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching, Journal of Geophysical Research: Atmospheres, 115, https://doi.org/10.1029/2009JD012882, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2009JD012882, 2010.

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

Maraun, D.: Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue, Journal of Climate, 26, 2137–2143, https://doi.org/10.1175/JCLI-D-12-00821.1, publisher: American Meteorological Society Section: Journal of Climate, 2013.

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

Maraun, D. and Widmann, M.: Cross-validation of bias-corrected climate simulations is misleading, Hydrology and Earth System Sciences, 22, 4867–4873, https://doi.org/10.5194/hess-22-4867-2018, publisher: Copernicus GmbH, 2018.

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., Hagemann, S., Richter, I., Soares, P. M. M., Hall, A., and Mearns, L. O.: Towards process-informed bias correction of climate change simulations, Nature Climate Change, 7, 764–773, https://doi.org/10.1038/nclimate3418, number: 11 Publisher: Nature Publishing Group, 2017.

Met Office: Iris: A powerful, format-agnostic, and community-driven Python package for analysing and visualising Earth science data, Tech. rep., Met Office, Exeter, Devon, https://doi.org/10.5281/zenodo.7386117, 2010.

Michelangeli, M. V. a. P.-A.: CDFt: Downscaling and Bias Correction via Non-Parametric CDF-Transform, https://cran.r-project.org/web/packages/CDFt/index.html, 2021.

Michelangeli, P.-A., Vrac, M., and Loukos, H.: Probabilistic downscaling approaches: Application to wind cumulative distribution functions, Geophysical Research Letters, 36, https://doi.org/10.1029/2009GL038401, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2009GL038401, 2009.

Mishra, V., Bhatia, U., and Tiwari, A. D.: Bias-corrected climate projections for South Asia from Coupled Model Intercomparison Project-6, Scientific Data, 7, 338, https://doi.org/10.1038/s41597-020-00681-1, number: 1 Publisher: Nature Publishing Group, 2020.

Navarro-Racines, C., Tarapues, J., Thornton, P., Jarvis, A., and Ramirez-Villegas, J.: High-resolution and bias-corrected CMIP5 projections for climate change impact assessments, Scientific Data, 7, 7, https://doi.org/10.1038/s41597-019-0343-8, number: 1 Publisher: Nature Publishing Group, 2020.

Naveau, P., Huser, R., Ribereau, P., and Hannart, A.: Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection, Water Resources Research, 52, https://doi.org/10.1002/2015WR018552, 2016.

NGFS: NGFS Climate Scenarios for central banks and supervisors, Tech. rep., Network for Greening the Financial System, https://www.ngfs.net/en/ngfs-climate-scenarios-central-banks-and-supervisors-september-2022, 2021.

Nogaj, M., Yiou, P., Parey, S., Malek, F., and Naveau, P.: Amplitude and frequency of temperature extremes over the North Atlantic region, Geophysical Research Letters, 33, https://doi.org/10.1029/2005GL024251, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2005GL024251, 2006.

Piani, C. and Haerter, J. O.: Two dimensional bias correction of temperature and precipitation copulas in climate models, Geophysical Research Letters, 39, https://doi.org/10.1029/2012GL053839, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2012GL053839, 2012.

Pokhrel, Y., Felfelani, F., Satoh, Y., Boulange, J., Burek, P., Gädeke, A., Gerten, D., Gosling, S. N., Grillakis, M., Gudmundsson, L., Hanasaki, N., Kim, H., Koutroulis, A., Liu, J., Papadimitriou, L., Schewe, J., Müller Schmied, H., Stacke, T., Telteu, C.-E., Thiery, W., Veldkamp, T., Zhao, F., and Wada, Y.: Global terrestrial water storage and drought severity under climate change, Nature Climate Change, 11, 226–233, https://doi.org/10.1038/s41558-020-00972-w, number: 3 Publisher: Nature Publishing Group, 2021.

Raei, E., Nikoo, M. R., AghaKouchak, A., Mazdiyasni, O., and Sadegh, M.: GHWR, a multi-method global heatwave and warm-spell record and toolbox, Scientific Data, 5, 180 206, https://doi.org/10.1038/sdata.2018.206, number: 1 Publisher: Nature Publishing Group, 2018.

Rocklin, M.: Dask: Parallel Computation with Blocked algorithms and Task Scheduling, in: Proceedings of the 14th Python in Science Conference, edited by Huff, K. and Bergstra, J., pp. 130 – 136, 2015.

Sippel, S., Zscheischler, J., Heimann, M., Otto, F. E. L., Peters, J., and Mahecha, M. D.: Quantifying changes in climate variability and extremes: Pitfalls and their overcoming, Geophysical Research Letters, 42, 9990–9998, https://doi.org/10.1002/2015GL066307, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2015GL066307, 2015.

Sippel, S., Otto, F. E. L., Forkel, M., Allen, M. R., Guillod, B. P., Heimann, M., Reichstein, M., Seneviratne, S. I., Thonicke, K., and Mahecha, M. D.: A novel bias correction methodology for climate impact simulations, Earth System Dynamics, 7, 71–88, https://doi.org/10.5194/esd-7-71-2016, publisher: Copernicus GmbH, 2016.

Spuler, F. and Wessel, J.: ibicus v1.0.1, https://doi.org/10.5281/zenodo.8101898, 2023.

Switanek, M., Maraun, D., and Bevacqua, E.: Stochastic downscaling of gridded precipitation to spatially coherent subgrid precipitation fields using a transformed Gaussian model, International Journal of Climatology, publisher: Wiley Online Library, 2022.

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

Verfaillie, D., Déqué, M., Morin, S., and Lafaysse, M.: The method ADAMONT v1.0 for statistical adjustment of climate projections applicable to energy balance land surface models, Geoscientific Model Development, 10, 4257–4283, https://doi.org/10.5194/gmd-10-4257-2017, publisher: Copernicus GmbH, 2017.

von Storch, H.: On the Use of "Inflation" in Statistical Downscaling, Journal of Climate, 12, 3505–3506, https://doi.org/10.1175/1520-0442(1999)012<3505:OTUOII>2.0.CO;2, publisher: American Meteorological Society Section: Journal of Climate, 1999.

Vrac, M.: Multivariate bias adjustment of high-dimensional climate simulations: the Rank Resampling for Distributions and Dependences (R^2D^2) bias correction, Hydrology and Earth System Sciences, 22, 3175–3196, https://doi.org/10.5194/hess-22-3175-2018, publisher: Copernicus GmbH, 2018.

Vrac, M. and Friederichs, P.: Multivariate—Intervariable, Spatial, and Temporal—Bias Correction, Journal of Climate, 28, 218–237, https://doi.org/10.1175/JCLI-D-14-00059.1, publisher: American Meteorological Society Section: Journal of Climate, 2015.

Vrac, M., Drobinski, P., Merlo, A., Herrmann, M., Lavaysse, C., Li, L., and Somot, S.: Dynamical and statistical downscal-
ing of the French Mediterranean climate: uncertainty assessment, Natural Hazards and Earth System Sciences, 12, 2769–2784,
https://doi.org/10.5194/nhess-12-2769-2012, publisher: Copernicus GmbH, 2012.

Vrac, M., Noël, T., and Vautard, R.: Bias correction of precipitation through Singularity Stochastic Removal: Because occur-
rences matter, Journal of Geophysical Research: Atmospheres, 121, 5237–5258, https://doi.org/10.1002/2015JD024511, _eprint:
https://onlinelibrary.wiley.com/doi/pdf/10.1002/2015JD024511, 2016.

Wessel, J. and Spuler, F.: ibicus v1.0.1 - data and additional code for GMD submission, https://doi.org/10.5281/zenodo.8101842, 2023.

Xu, Z., Han, Y., Tam, C.-Y., Yang, Z.-L., and Fu, C.: Bias-corrected CMIP6 global dataset for dynamical downscaling of the historical and
future climate (1979–2100), Scientific Data, 8, 293, https://doi.org/10.1038/s41597-021-01079-3, number: 1 Publisher: Nature Publishing
Group, 2021.

Zappa, G. and Shepherd, T. G.: Storylines of Atmospheric Circulation Change for European Regional Climate Impact Assessment, Journal
of Climate, 30, 6561–6577, https://doi.org/10.1175/JCLI-D-16-0807.1, publisher: American Meteorological Society Section: Journal of
Climate, 2017.

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., Trewin, B., and Zwiers, F. W.: Indices for monitoring changes
in extremes based on daily temperature and precipitation data, WIREs Climate Change, 2, 851–870, https://doi.org/10.1002/wcc.147,
_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.147, 2011.

Zscheischler, J., Westra, S., Van Den Hurk, B. J., Seneviratne, S. I., Ward, P. J., Pitman, A., Aghakouchak, A., Bresch, D. N., Leonard, M.,
Wahl, T., and Zhang, X.: Future climate risk from compound events, Nature Climate Change, 8, https://doi.org/10.1038/s41558-018-0156-
3, 2018.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., van den Hurk, B., AghaKouchak, A., Jézéquel, A.,
Mahecha, M. D., Maraun, D., Ramos, A. M., Ridder, N. N., Thiery, W., and Vignotto, E.: A typology of compound weather and climate
events, Nature Reviews Earth and Environment, 1, https://doi.org/10.1038/s43017-020-0060-z, 2020.