SoftwareX 22 (2023) 101379

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

Original software publication

Introduction of the BiasAdjustCXX command-line tool for the application of fast and efficient bias corrections in climatic research

Benjamin Thomas Schwertfeger [a,b,*], Gerrit Lohmann [a,c], Henrik Lipskoch [b]

[a] Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany
[b] University of applied Sciences Bremerhaven, Bremerhaven, Germany
[c] University of Bremen, Bremen, Germany

ARTICLE INFO

Article history:
Received 11 January 2023
Received in revised form 13 March 2023
Accepted 22 March 2023

Keywords:
Command-line tool
Bias correction
Climatic research
C++

ABSTRACT

Bias correction algorithms for modeled climate variables such as temperature, precipitation, and barometric pressure are used to approximate certain aspects of the distribution characteristics to the actual observed values. Thus, modeled climate data estimating future climate scenarios can be bias-adjusted using data from past periods so that climate variables and their distribution, as well as their variability, can be represented more realistically within the bias-adjusted time series. This document aims to introduce the command-line tool BiasAdjustCXX that enables the application of different scaling- and distribution-based bias adjustment techniques to minimize bias which can be estimated from time-series climate data.

Code metadata

Current code version	v1.8.1
Permanent link to code/repository used for this code version	https://github.com/ElsevierSoftwareX/SOFTX-D-23-00031
Permanent link to Reproducible Capsule	none
Legal Code License	GNU GPL v3
Code versioning system used	git (Github)
Software code languages, tools, and services used	C++11 (tested with gcc 14.0.0)
Compilation requirements, operating environments & dependencies	MacOS, Linux, CMake>= 3.10, NetCDF-4 C++ library>= 4.3.1
If available Link to developer documentation/manual	none
Support email for questions	development@b-schwertfeger.de

1. Motivation and significance

Due to the limited resolution of general circulation, regional climate as well as earth system models, and their large-scale grid cells, the topography and generalization of the processes cannot be mapped equally everywhere, which can lead to high deviations between reanalyzed or observed and modeled data on a local scale, since small-scale processes cannot be fully represented. Also, the distribution of certain quantiles and the extremes can only be determined for one cell individually at a time in a climate model, which means that these cells only represent averaged values for a relatively large area, depending on the model resolution. These statistical deviations and errors also called "bias" can be observed by estimating different aspects of the distributional properties of a climate variable, such as the mean, the variance, the standard deviation, and even the correlation.

If these deviations between observed and modeled data exist, procedures to minimize these errors can be applied. Due to the fact that these so-called bias corrections do not eliminate all errors, but can only reduce them to a minimum, the term "bias adjustment" is also valid.

The development of the command-line tool "BiasAdjustCXX" [1] started with the bachelor thesis of Benjamin T. Schwertfeger [2]. This software is now able to minimize the mentioned discrepancies by providing five different methods to approximate time series of modeled climate variables to the actually occurred or future values based on reference data of the control period.

* Corresponding author at: Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany.
E-mail address: development@b-schwertfeger.de (Benjamin Thomas Schwertfeger).

https://doi.org/10.1016/j.softx.2023.101379
2352-7110/

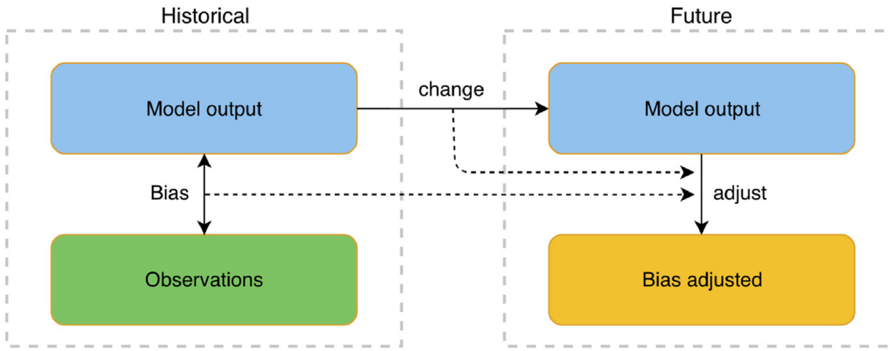

Fig. 1. Schematic representation of the data sets and information flow during bias adjustments.

These methods can be applied to time series of both past and future climate scenarios, provided that modeled and reference data of a control period are available. This can significantly improve the results of climate studies, which use time series from climate models that can realistically represent the atmospheric or ocean dynamics, but which, for example, represent temperatures that are too cold on average.

2. Software description

The command-line tool BiasAdjustCXX provides different scale- and distribution-based bias correction techniques that can be applied to 1- and 3-dimensional NetCDF-based data sets. These methods usually have an additive and multiplicative variant, so that ratio-based variables such as precipitation can be adjusted with relative rather than absolute factors, as is the case with interval variables such as air temperatures. In addition, further parameters for individual adjustments can be submitted at the start of the program (cf. Section 2.2)

Statistical transformations in the form of a bias adjustment on modeled climate data require observed and modeled data of the same climate variable. Modeled data for a future period require modeled and observed data covering a historical time period as control period. Since most adjustment techniques assume that the error between modeled and observed data remains persistent, i.e., the bias also exists in future periods [3], this can be used to adjust modeled time series of the future. For example by taking into account the bias between the observed and modeled data of the control period or the change between modeled historical and future periods, and thus minimizing the deviations mentioned earlier (Fig. 1).

This presupposes the basic assumption that there is an existing relationship between the generated data of modeled climate scenarios and the real, actual climatic processes and conditions, and that topographic influences in observed, reanalyzed as well as modeled time series are also related to each other.

The mathematical basis of the implemented scaling-based bias adjustment procedures linear scaling (LS), variance scaling (VS), and the delta method (DM), are derived from the articles by [4,5]. For the implementation of the quantile mapping (QM) and quantile delta mapping (QDM) procedures, the description of [6,7] was followed.

2.1. Software architecture

The file main.cxx is the entry point of this modular software, where the class "Manager" is instantiated, which controls the program flow. The mentioned bias correction techniques are part of the custom class "CMethods". All adjustment methods implemented inside are declared static to enable the standalone usage of this functions in custom scripts or Jupyter Notebooks, like in the provided example Notebook named "examples.ipynb". This can be optionally used after installing the required dependencies specified in "environment.yml".

There are several other classes including utility functionalities (Utils) and mathematical formulas (MathUtils) like the mean, standard deviation and interpolation (cf. Fig. 2). This also includes data structures for loading and saving (NcFileHandler) which depend on the Unidata Program Center's programming interfaces for NetCDF-based climate data [8], as well as usage and execution examples provided in "example_all_methods.run.sh". This script serves as an example on how to apply multiple techniques with different configurations by using the provided sample data sets within the input_data directory of the repository. The Jupyter Notebook "Hands-On-BiasAdjustCXX.ipynb" is particularly well suited for getting started and shows the download of the project, the compilation of the source files, the application of the tool using supplied sample data, and the subsequent visualization of the results.

In addition, a Dockerfile is provided which can be used to build a Docker [9] image and run the BiasAdjustCXX tool within a container. The latest image (btschwertfeger/biasadjustcxx:latest) and older versions can also be pulled directly from the Dockerhub registry. Thus, no local installation of the NetCDF-4 library, CMake, and BiasAdjustCXX is required and allows the application on all operating systems where Docker can be set up and executed.

BiasAdjustCXX will gracefully shut down in case of invalid shaped data sets, missing or wrong inputs as well as invalid conditions for the respective adjustment method and the user will receive enlightening command-line output of the current problem.

2.2. Software functionalities

As mentioned earlier, this software enables the application of different bias correction techniques on time series climate data sets. The input data sets must map a climate variable over the dimensions "time", "lat" and "lon" (i.e., time, latitudes, longitudes) in exactly this order in case the variable of interest has more than the time dimension within the data set.

When using scaling-based methods, time series can be scaled on the basis of long-term monthly means [4] or based on long-term 31-day moving windows. Month-dependent scaling cannot avoid unrealistic monthly transitions on long-term means, since surrounding months are scaled by different factors. This can be avoided using long-term 31-day moving windows where each day of the year is assigned an individual scaling factor based on the nearest 31 days in every year. The 31-day moving window variant was developed as this weak point of the monthly-dependent scaling was noticed.

The distribution-based bias adjustment techniques implemented here, consider the individual time series as a whole, so

```
├─ include
│  ├─ CMethods.hxx
│  ├─ colors.h
│  ├─ Manager.hxx
│  ├─ MathUtils.hxx
│  ├─ NcFileHandler.hxx
│  └─ Utils.hxx
├─ src
│  ├─ CMethods.cxx
│  ├─ main.cxx
│  ├─ Manager.cxx
│  ├─ MathUtils.cxx
│  ├─ NcFileHandler.cxx
│  └─ Utils.cxx
├─ input_data
│  └─ *.nc
├─ examples
│  ├─ environment.yml
│  ├─ example_all_methods.run.sh
│  ├─ examples.ipynb
│  └─ Hands-On-BiasAdjustCXX.ipynb
├─ CMakeLists.txt
└─ Dockerfile
```

Fig. 2. BiasAdjustCXX project structure.

that their distributional properties can be adjusted month- and interval-independent.

Except for the variance scaling, all included adjustment techniques have not only an additive but also a multiplicative variant, so that ratio variables, such as precipitation, can be adjusted with relative instead of absolute factors like temperatures.

As per default in this software, the scaling-based techniques are applied based on long-term 31-day moving windows. This requires that all input files exclude the 29th February and every year has 365 entries. This is not required when using the `--no-group` flag which can be used to apply the scaling in such a way that the scaling factors are based on the whole time series at once. This enables the possibility to apply the BiasAdjustCXX tool to data sets with custom time scales, for example to adjust monthly separated time series individually to match the techniques described in [4,5]. Table 1 lists the available arguments which can be passed at program start.

2.3. Sample code snippets analysis

Once the software is compiled using CMake, the executable file "BiasAdjustCXX" can be used to adjust time series climate data.

In Listing 1 the tool is used to adjust the data set "scenario.nc" that maps the variable "tas" (i.e., 2 m air temperatures) over the dimensions "time", "lat", and "lon" using the additive delta method as described by [5] but using long-term 31-day moving windows instead of month-dependent scaling factors.

```
BiasAdjustCXX                    \
    --ref observations.nc  \ # reference data (control
    period)
    --contr control.nc     \ # modeled data (control
    period)
    --scen scenario.nc     \ # data to adjust (scenario
    period)
    -o tas_dm_adjusted.nc  \ # output file
    -v tas                 \ # variable of interest
    -m delta_method        \ # adjustment method
    -k "+"                   # additive variant
```

Listing 1: Execution example for adjusting a 3-dimensional data set using the additive delta method based on long-term 31-day moving windows

The application of a distribution-based technique on a 1-dimensional data set (i.e., only one time series is included) that maps a ratio-based climate variable like precipitation (i.e., "pr") over the time dimension is shown in Listing 2.

```
BiasAdjustCXX                        \
    --ref 1d_observations.nc   \ # reference data (
    control period)
    --contr 1d_control.nc      \ # modeled data (control
     period)
    --scen 1d_scenario.nc      \ # data to adjust (
    scenario period)
    -o pr_qdm_adjusted_1d.nc   \ # output file
    -v pr                      \ # variable of interest
    -m quantile_delta_mapping  \ # adjustment method
    -k "*"                     \ # multiplicative
    variant
    --1dim                     \ # input files only
    contain one time series
    -q 250                       # quantiles to respect
```

Listing 2: Execution example for adjusting a 1-dimensional data set using the multiplicative quantile delta mapping technique

3. Illustrative examples

Daily mean temperature values, as well as daily mean precipitation rates of the control period from 1951 to 1980 are used to adjust modeled time series of the scenario period from 1981 to 2010. The modeled data are those from CMIP6.CMIP.MPI-M.MPI-ESM1-2-HR.historical [10]. The NOAA/CIRES/DOE 20th Century Reanalysis Project V3 provide the reference data sets [11].

The modeled data sets have a global grid resolution of 384 *x* 192 longitudes/latitudes. To apply BiasAdjustCXX, these grids must be scaled to match the grid resolution of the reference data with 360 *x* 181 longitudes/latitudes. The bias thus artificially created is to be disregarded, since the focus in this example is only on the effects of the methods and not on the evaluation of the underlying data sets. In addition the European region from −45 degrees West to 65 degrees East and from 23 to 72 degrees North is chosen.

Bias adjustment algorithms can have a large impact on the mean values of climate time series. The long-term mean values of the modeled daily mean 2 m air temperatures of the MPI-ESM, averaged over all time series of the European region, are significantly cooler than those of the reference time series of

Table 1
List of available arguments for the BiasAdjustCXX command-line tool.

Argument	Description
`--ref`, `--reference`	path to observational/reference data set (control period)
`--contr`, `--control`	path to modeled data set (control period)
`--scen`, `--scenario`	path to data set that is to be adjusted (scenario period)
`-v`, `--variable`	variable to adjust
`-k`, `--kind`	kind of adjustment; one of: "`+`" or "`add`" and "`*`" or "`mult`"
`-m`, `--method`	adjustment method name one of: "`linear_scaling`", "`variance_scaling`", "`delta_method`", "`quantile_mapping`" and "`quantile_delta_mapping`"
`-q`, `--n_quantiles`	[optional] number of quantiles to respect (only for distribution-based methods)
`--1dim`	[optional] required if the data sets have no spatial dimensions (i.e., only one time dimension)
`--no-group`	[optional] Disables the adjustment based on long-term 31-day moving windows for the scaling-based methods. Scaling will be performed on the whole data set at once, so it is recommended to separate the input files for example by month and apply this program to every long-term month. (only for scaling-based methods)
`--max-scaling-factor`	[optional] Define the maximum scaling factor to avoid unrealistic results when adjusting ratio based variables for example in regions where heavy rainfall is not included in the modeled data and thus creating disproportional high scaling factors. (only for multiplicative methods, except QM; default: 10)
`-p`, `--n_processes`	[optional] How many threads to use (default: 1)
`-h`, `--help`	[optional] display usage example, arguments, hints and exits the program

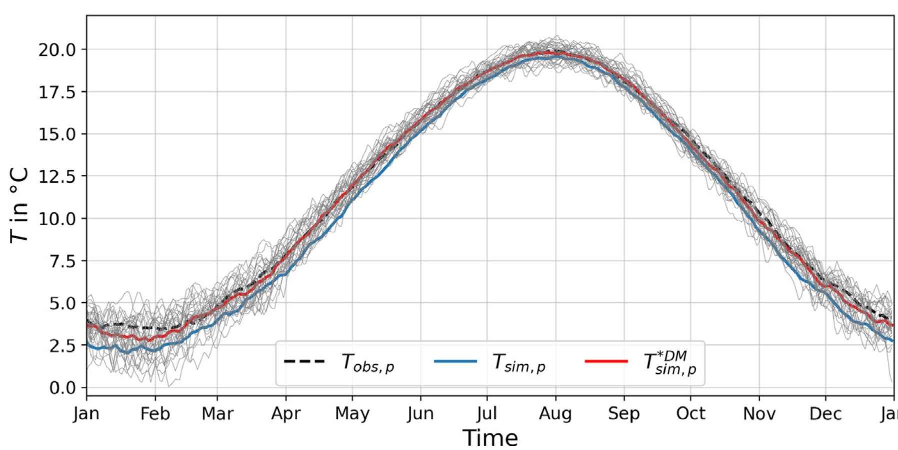

Fig. 3. NOAA 20thC Reanv3 ($T_{obs,p}$), MPI-ESM ($T_{sim,p}$), and delta method-adjusted ($T^{*DM}_{sim,p}$, based on long-term 31-day moving windows) unweighted mean 2 m air temperatures per day of the year in Europe; gray lines represent daily mean 2 m air temperatures of the bias-adjusted time series of every year within the scenario period (−45°W to 65°E and 23°N to 72°N; control period: 1951–1980, scenario/presented period: 1981–2010).

the reanalysis project (Fig. 3, Fig. 4). The time series adjusted using the additive delta method was able to archive a clear approximation to the mean values of the reference data from the scenario period.

In Fig. 3, the subscripts in $T_{obs,p}$, $T_{sim,p}$, $T^{*DM}_{sim,p}$ stand for observed data (*obs*), simulated data (*sim*), and predicted data (*p*). The superscript $*DM$ in $T^{*DM}_{sim,p}$ indicates the data adjusted using the delta method.

The Mean Bias Error (MBE) is used to determine the mean deviation of all time steps between two time series. Since bias corrections also aim to minimize mean deviations, the MBE between the observed and corrected data of the scenario period is in general closer to zero than the underlying modeled data of the scenario time series (Fig. 4).

Also daily mean precipitation rates can be approximated to the reference data by applying the multiplicative delta method. The

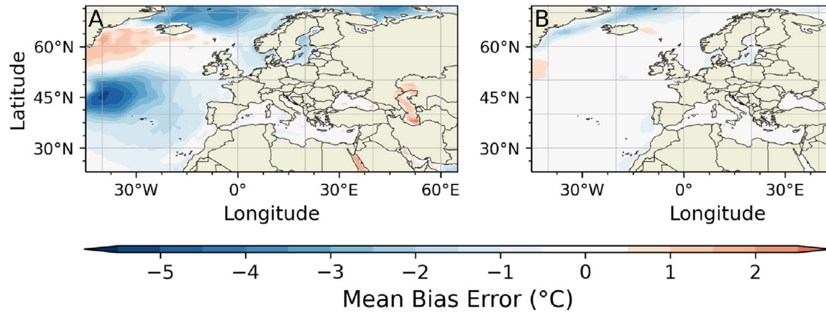

Fig. 4. (A) MBE of daily mean 2 m air temperatures in NOAA 20thC Reanv3 and MPI-ESM (1981–2010); (B) MBE of daily mean 2 m air temperatures in NOAA 20thC Reanv3 and in delta method adjusted data (1981–2010).

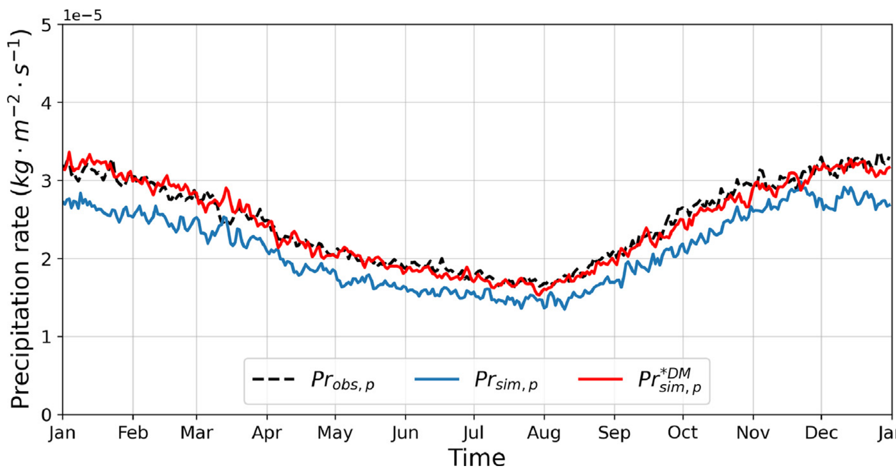

Fig. 5. NOAA 20thC Reanv3 ($Pr_{obs,p}$), MPI-ESM ($Pr_{sim,p}$) and delta method-adjusted ($Pr^{*DM}_{sim,p}$, based on long-term 31-day moving windows) unweighted mean precipitation rates per day of the year in Europe (−45° W to 65°E, 23°N to 72°N; control period: 1951–1980, scenario/presented period: 1981–2010).

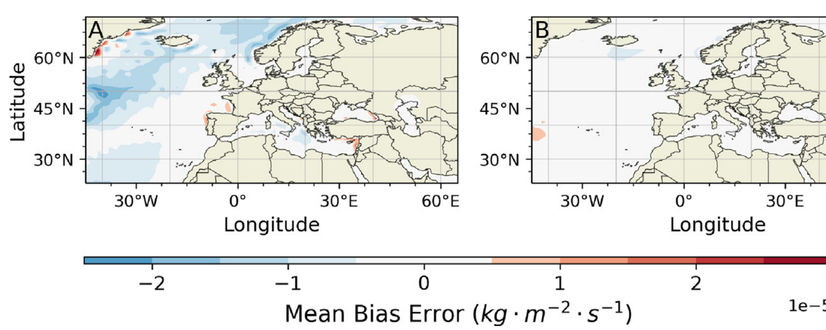

Fig. 6. (A) Mean Bias Error of daily mean precipitation rates in NOAA 20thC Reanv3 and MPI-ESM (1981–2010); (B) Mean Bias Error of daily mean precipitation rates in NOAA 20thC Reanv3 and the delta method adjusted data (1981–2010.

bias-adjusted time series now archive a higher similarity to the reference data than the raw modeled time series (Figs. 5 and 6).

4. Impact

The basis of hydrological and climate impact studies [4,12,13] are data sets coming from such models that are subject to exactly the same sources of error as those mentioned in Section 1. Therefore it is particularly important to prepare them beforehand by means of bias correction procedures, so that their distributional properties can be realistically adjusted on the basis of historical observations [3,6]. This is exactly what BiasAdjustCXX offers, a fast and computationally efficient open source tool for post-processing climate model output to approximate modeled time series to the actual occurring values for both past and future time periods. To show this, the execution time for different sized data sets has been measured and visualized in [14]. This includes a comparison to the results of the Python modules "xclim" v0.40.0 [15] and "python-cmethods" v0.6.1 [16].

Until now, there has been a lack of easy-to-use and efficient methods for bias correction outside of Python, R, and MATLAB.

BiasAdjustCXX solves this problem by initially providing five different methods and making it easier for scientists and researchers to apply them in an operating system independent way.

Previously implemented methods can also be investigated with the help of linked references. In addition, the modular implementation offers the possibility to add further procedures or to customize existing ones. This can be achieved by adding the appropriate implementation of a new method to the CMethods class and defining the associated call-up conditions within the Manager class.

Thus, this tool can be used in the future as a starting point for the application of bias corrections without high-level programming skills needed. At the same time, the open source nature of the methods allows them to be extended, so that further methods can be accumulated in the future whose mathematical foundations come from valid references.

5. Conclusions

The presented command-line tool BiasAdjustCXX allows the application of different scale- and distribution-based bias correction techniques for climatic research. It tries to fill the gap of missing fast and efficient platform-independent bias correction algorithms for post-processing climate model output. The tool can be executed both from the command-line and also within custom scripts and routines for example by being integrated in post-processing workflows. BiasAdustCXX thus provides an opportunity for researchers to keep the focus on their studies instead of implementing own bias correction algorithms and perhaps even adding new errors or wasting time with testing code and troubleshooting.

The selection of an appropriate method for a particular study is a large separate topic, since it depends heavily on the underlying data and the focus of the research. For this reason, no recommendations for specific methods are made here.

CRediT authorship contribution statement

Benjamin Thomas Schwertfeger: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualisation. **Gerrit Lohmann:** Conceptualization, Resources, Writing – review & editing, Supervision, Funding acquisition. **Henrik Lipskoch:** Conceptualization, Resources, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data used are linked in the references. Example data generated by the authors themselves can be found in the repository.

References

[1] Schwertfeger BT. btschwertfeger/BiasAdjustCXX: Release v1.8.1. Zenodo; 2023, http://dx.doi.org/10.5281/zenodo.7722979.

[2] Schwertfeger BT. The influence of bias corrections on variability, distribution, and correlation of temperatures in comparison to observed and modeled climate data in Europe. University of Applied Sciences Bremerhaven, Alfred Wegener Institute, Helmholtz-Center for Polar and Marine Research; 2022, URL https://epic.awi.de/id/eprint/56689/.

[3] Maraun D. Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. Geophys Res Lett 2012;39(6). http://dx.doi.org/10.1029/2012GL051210.

[4] Teutschbein C, Seibert J. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. J Hydrol 2012;456–457:12–29. http://dx.doi.org/10.1016/j.jhydrol.2012.05.052.

[5] Beyer R, Krapp M, Manica A. An empirical evaluation of bias correction methods for palaeoclimate simulations. Clim Past 2020;16(4):1493–508. http://dx.doi.org/10.5194/cp-16-1493-2020.

[6] Tong Y, Gao X, Han Z, Xu Y, Xu Y, Giorgi F. Bias correction of temperature and precipitation over China for RCM simulations using the QM and QDM methods. Clim Dynam 2021;57(5):1425–43. http://dx.doi.org/10.1007/s00382-020-05447-4.

[7] Cannon AJ, Sobie SR, Murdock TQ. Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? J Clim 2015;28(17):6938–59. http://dx.doi.org/10.1175/JCLI-D-14-00754.1.

[8] Unidata. NetCDF programming interfaces: netCDF v4.9.0, netcdf-cxx4 v4.3.1. Boulder, CO: UCAR/Unidata Program Center; 2022, http://dx.doi.org/10.5065/D6RN35XM.

[9] Merkel D. Docker: lightweight linux containers for consistent development and deployment. Linux J 2014;2014(239):2.

[10] Jungclaus J, Bittner M, Wieners K-H, Wachsmann F, Schupfner M, Legutke S, et al. MPI-M MPI-ESM1.2-HR model output prepared for CMIP6 CMIP historical. Earth System Grid Federation; 2019, http://dx.doi.org/10.22033/ESGF/CMIP6.6594.

[11] Slivinski LC, Compo GP, Whitaker JS, Sardeshmukh PD, Giese BS, McColl C, et al. Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. Q J R Meteorol Soc 2019;145(724):2876–908. http://dx.doi.org/10.1002/qj.3598.

[12] Chen J, Yang Y, Tang J. Bias correction of surface air temperature and precipitation in CORDEX East Asia simulation: What should we do when applying bias correction? Atmos Res 2022;280:106439. http://dx.doi.org/10.1016/j.atmosres.2022.106439.

[13] Chen J, Brissette FP, Chaumont D, Braun M. Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. Water Resour Res 2013;49(7):4187–205. http://dx.doi.org/10.1002/wrcr.20331.

[14] Schwertfeger BT. btschwertfeger/BiasAdjustCXX-Performance-test: Release v1.0.1. Zenodo; 2023, http://dx.doi.org/10.5281/zenodo.7678393.

[15] Logan T, Aoun A, Bourgault P, Dupuis É, Huard D, Lavoie J, et al. Ouranosinc/xclim: v0.40.0. Zenodo; 2023, http://dx.doi.org/10.5281/zenodo.7535677.

[16] Schwertfeger BT. btschwertfeger/python-cmethods: Release v0.6.1. Zenodo; 2022, http://dx.doi.org/10.5281/zenodo.7652756.