CMhyd User Manual

Documentation for preparing simulated climate change data for hydrologic impact studies

May, 2016

Hendrik Rathjens, Katrin Bieger, Raghavan Srinivasan, Indrajeet Chaubey, Jeffrey G. Arnold

In association with

About the software and this manual

Watershed models are often used to simulate the impact of future climate conditions on hydrologic processes. However, Teutschbein and Seibert (2012) state that simulations of temperature and precipitation often show significant biases due to systematic model errors or discretization and spatial averaging within grid cells, which hampers the use of simulated climate data as direct input data for hydrological models. Bias correction procedures are used to minimize the discrepancy between observed and simulated climate variables on a daily time step so that hydrological simulations driven by corrected simulated climate data match simulations using observed climate data reasonably well. CMhyd is a tool that can be used to extract and bias-correct data obtained from global and regional climate models. It is highly recommended to apply an ensemble approach, i.e. to use bias-corrected data provided by several climate models and downscaling methods (e.g., Teutschbein and Seibert, 2010, 2012).

This manual explains how to prepare simulated climate change data for watershed-based hydrologic impact studies using the CMhyd (**C**limate **M**odel data for **hyd**rologic modeling) tool. For background information, the reader is referred to the literature. Detailed explanations and discussions of the theory of linking climate and hydrologic models have been published (e.g., Salathé, 2003; Bates et al., 2008; Christensen et al., 2008; Teutschbein and Seibert, 2010, 2012; Lafon et al., 2013).

CMhyd was written in Python 2.7 using several Python packages (mainly NetCDF4[1], NumPy (van der Walt et al., 2011), SciPy (Oliphant, 2007; Millman and Aivazis, 2011)) and the PyQt4[2] application framework.

Example data

This document is accompanied by example files with observed and simulated climate data (precipitation and temperature). A step-by-step guideline for downscaling (or bias-correcting) the simulated climate data using the observed data is provided in these boxes.

[1] https://pypi.python.org/pypi/netCDF4
[2] https://riverbankcomputing.com/software/pyqt/intro

Processing framework

CMhyd was designed to provide simulated climate data that can be considered representative for the location of the gauges used in a watershed model setup. Therefore, climate model data should be extracted and bias-corrected for each of the gauge locations.

Bias correction procedures employ a transformation algorithm for adjusting climate model output. The underlying idea is to identify biases between observed and simulated historical climate variables to parametrize a bias correction algorithm that is used to correct simulated historical climate data (see Figure 1). Bias correction methods are assumed to be stationary, i.e. the correction algorithm and its parametrization for current climate conditions are assumed to be valid for future conditions as well. Thus, the same correction algorithm is applied to the future climate data. However, it is unknown how well a bias correction method performs for conditions different from those used for parametrization. A good performance during the evaluation period does not guarantee a good performance under changed future conditions. Teutschbein and Seibert (2012) provide a detailed discussion and state that a method that performs well for current conditions is likely to perform better for changed conditions than a method that already performs poorly for current conditions.

Example data - The data set

The example data set is solely provided (1) to give an example of the data format, and (2) to demonstrate the general functioning of CMhyd. Do not use the observed data to evaluate climate model performance or bias correction methods.
Unzip the example dataset to a location on your hard drive (`.../example_data`). Please create a backup of the input files before getting started.

Step 1: Preparing observed data

CMhyd uses ASCII format for observed data. The data for the gauges are saved in individual files, all of which are listed in a location file. Separate location files are required for precipitation and temperature. The *NAME* fields in the location files point to the corresponding data files and the fields *LAT* and *LON* specify the location of the gauges. In the data files, the first line is reserved for the starting date of the time series, while each of the following lines represents one day. Data gaps (i.e. missing values) have to be included using a no-data value (-99.9 or -99.0). The precipitation data file contains one record per day (daily sum of precipitation [mm]) and the temperature data file two (daily maximum and minimum temperature [°C]). The data file names have to match the names specified in the *NAME* field of the location file. All files have the extension *txt* (see Figure 2 and Table 1 and 2).

The ASCII format is also used by the ArcSWAT interface (Winchell et al., 2010), which facilitates the use of simulated climate data in SWAT (Soil and Water Assessment Tool, Arnold et al., 1998) model applications.

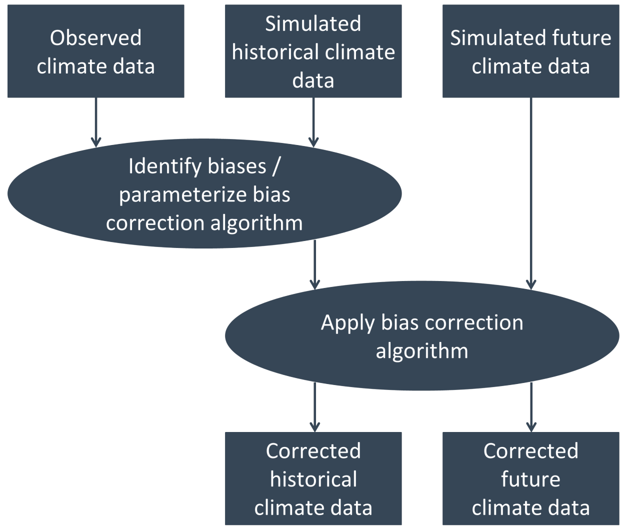

Figure 1: Bias correction framework.

Table 1: Climate data ASCII format: Location file.

Field name	Field format	Description
ID	integer	Gauge / raster identification number (1, 2, ...)
NAME	string	Data file name (points to the data file)
LAT	real	Latitude in decimal degrees
LON	real	Longitude in decimal degrees
Elevation	real	Elevation of the gauge (not used)

Table 2: Climate data ASCII format: Precipitation and temperature data files.

Line	Format	Description
First	date (YYYYmmdd)	Starting date of the time series
Following lines		Daily data (one line per day):
	real	Daily precipitation [mm]
	real,real	Daily max. and min. temperature [°C] (comma separated)

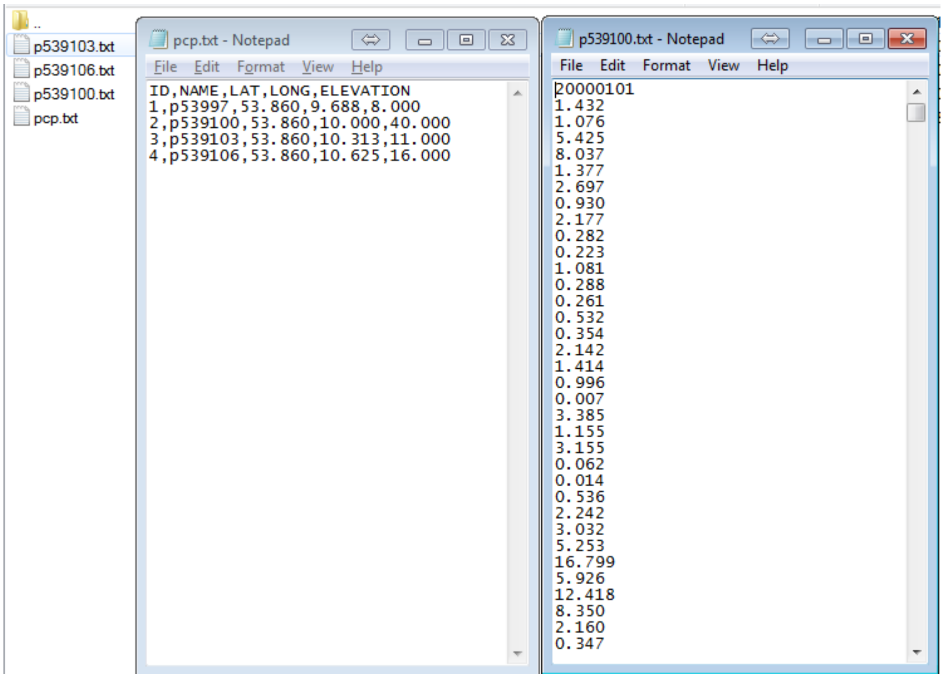

Figure 2: Precipitation data in ASCII format.

Example data - Observed data

Precipitation and temperature time series for four gauges in Northern Germany are provided in the example data set (`.../example_data/observed`). They include data for the time period from January 1st 2000 to December 31st, 2010. The precipitation and temperature location files are called *pcp.txt* and *tmp.txt*, respectively.

Step 2: Extracting historical and scenario climate model data

Climate model time series are typically provided in netCDF[3] (Network Common Data Form) format. The data is stored in several, multi-dimensional, binary-decoded **.nc* files. netCDF is a self-describing, machine-independent data format that supports the creation, access, and sharing of array-oriented scientific data. The format is an open standard of the Open Geospatial Consortium. Each file contains a set of variables, dimensions, and attributes.

CMhyd uses the meta information of a netCDF file to (1) find the climate model grid cells that overlay the gauge locations and (2) convert the precipitation and temperature data into millimeters and degrees Celsius, respectively. Finally, CMhyd extracts time series of the relevant grid cells by reading from the netCDF files.

Download simulated data The *CMIP – Data Access – Getting Started*[4] section is a good starting point for downloading simulated climate data. Attributes and organization of meta information might vary with the data source and CMhyd will not work with any netCDF data structure. The tool has been tested using the CORDEX (COordinated Regional climate Downscaling Experiment) archive, which is a reliable

[3] `http://www.unidata.ucar.edu/software/netcdf/`
[4] `http://cmip-pcmdi.llnl.gov/cmip5/data_getting_started.html`

source for regional climate models. Within this project several data nodes are available. Links are provided on the CORDEX web page[5]. The tool has also been tested with data from the *Downscaled CMIP3 and CMIP5 Climate Projections*[6] archive. The data server require regular maintenance and might be temporarily unavailable.

Download the data for your region and time period of interest. Even though CMhyd is able to handle fragmented time series it is recommended to use complete time series. The climate model name conventions for precipitation and temperature are

- *pr* for precipitation and
- *tasmax* and *tasmin* for maximum and minimum temperature, respectively.

The CORDEX data archive CMhyd was designed to work with the CORDEX data archive, which provides downscaled regional climate model data. The data can be downloaded for specific domains. An overview of the domains including their coverages is given on the CORDEX web page[7]. The web interface provides a filter that allows the user to efficiently browse the data archive. Four filters are recommended for downloading the data.

1. Domain (e.g., EUR-44 for Europe; do not select the *XXX*-i* domains).
2. Time Frequency (day).
3. Experiment (one historical or evaluation and one or more future experiments).
4. Variable (*pr*, *tasmax*, and *tasmin*).

Check data format and meta information While CMhyd was designed to work with the CORDEX data archive, it has also been tested using other data sources. Please contact the developers if the structure of a netCDF data format is not supported by the software.

Errors in the meta data are possible and it is strongly recommended to check the meta information and data attributes of the netCDF files using *ncBrowse*[8] (e.g., name of variable, time period, units, . . .). An example is given in Figure 3. If everything appears to be correct CMhyd can be used to extract and bias correct the data.

If no meta data has been incorporated, the file name can be used to pass the missing information to CMhyd using the syntax `<VariableName>_<Frequency>_<ModelName>_<Experiment>.nc` (e.g., `pr_day_ICHEC-EC-EARTH_rcp45.nc`).

Example data - CORDEX simulated historical and scenario data

In the example data set precipitation and temperature time series of historical and scenario climate model (ICHEC-EC-EARTH) data from the CORDEX archive is provided. The historical and scenario (rcp45) data cover the time period from January 1st 1996 to December 31st 2005 and January 1st, 2091 to December 31st 2100. Each **.nc* file includes one climate variable (i.e. *pr*, *tasmax*, and *tasmin*) and covers a 5-year time period.
Use *ncBrowse* to check if the meta information of the **.nc* files is correct (see Figure 3).

[5] http://www.cordex.org/
[6] http://gdo-dcp.ucllnl.org/
[7] http://www.cordex.org/index.php/community/domains/
[8] http://www.epic.noaa.gov/java/ncBrowse/

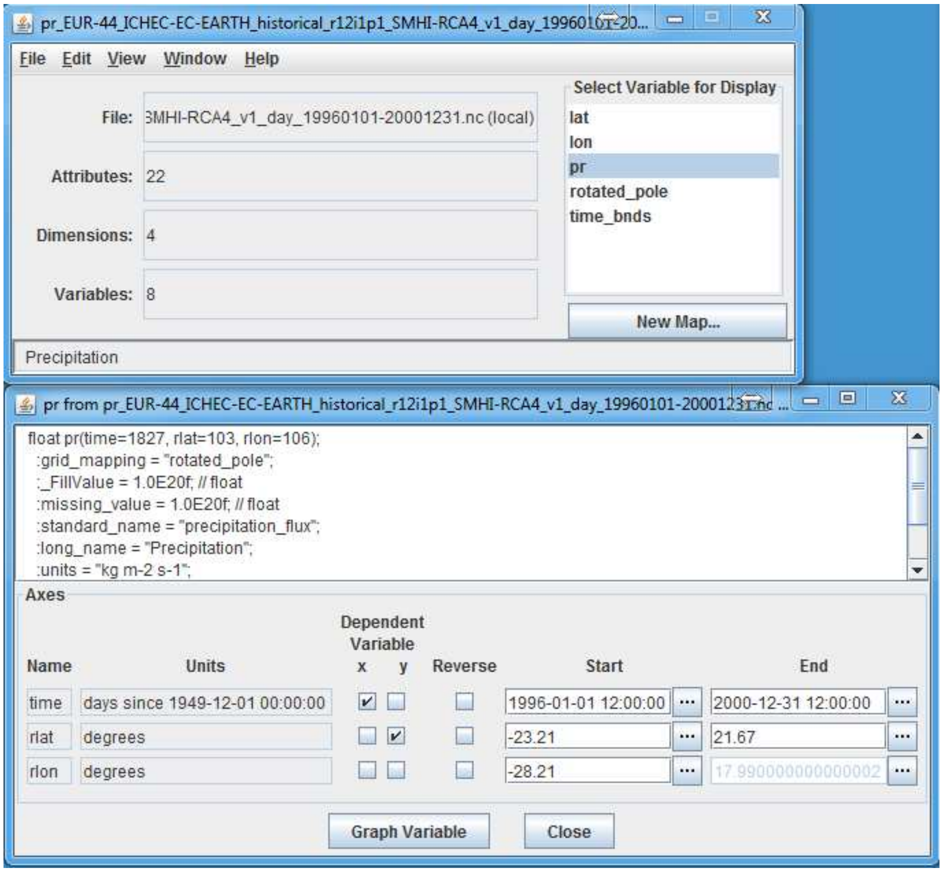

Figure 3: Use *ncBrowse* to access the meta information of a **.nc* file.

Step 3: Processing the data using CMhyd

The *Processing* tab of the graphical user interface (GUI) guides the user through the data extraction and bias correction process. Basically, five steps are needed (see Figure 4).

A command prompt window opens in addition to the GUI. This window should be ignored unless an error occurs or CMhyd shows unexpected behavior. Often, the error message shown in the command prompt is useful to track down the error. Check if the CMhyd files are in the correct format prior to contacting the developers.

Example data - Applying the distribution mapping and linear scaling method on precipitation data

The example section will explain how to apply the distribution mapping and linear scaling method to the example precipitation data.

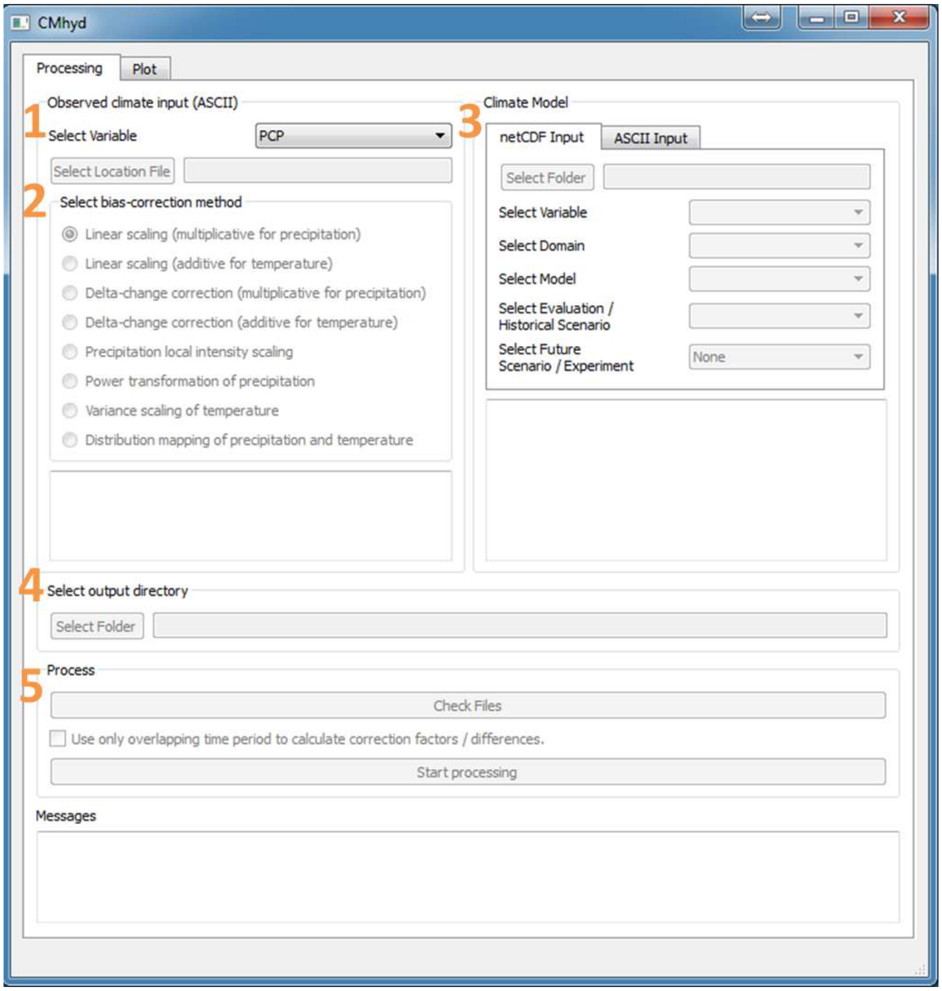

Figure 4: The CMhyd graphical user interface. The orange numbers indicate the steps 3.1 to 3.5.

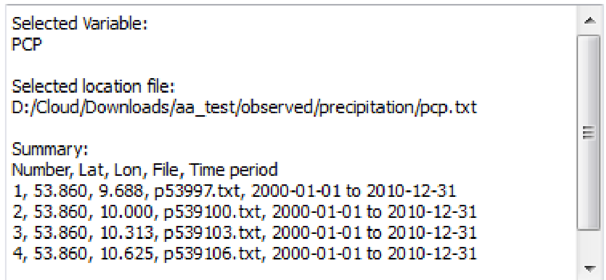

Figure 5: CMhyd Example summary of observed data.

Step 3.1 Select climate variable and observed data

CMhyd is able to extract and correct precipitation (PCP) and temperature (TMP) data. The user needs to select a variable from the combo box and then provide the path to the gauge location file. Information about the observed data (e.g., number and location of gauges, extent of time periods) is given in the text box below (see Figure 5).

Example data - Select climate variable and location file

Select *PCP* from the combo box and click on the *Select Location File* button. Then, navigate to the location of the unzipped example data set and select the precipitation location file (`../example_data/observed/precipitation/pcp.txt`).
The summary shown in Figure 5 will be provided in the text box.

3.2 Select bias-correction method

Eight bias correction methods have been implemented into CMhyd. Teutschbein and Seibert (2012) provide mathematical descriptions of all methods. They also evaluate and discuss the advantages and disadvantages of each method in a hydrological context.

Some bias-correction methods have been designed to either work with precipitation or temperature data (see, Teutschbein and Seibert, 2012). Only the methods that work with the selected variable will be available for selection.

Example data - Select bias-correction method

Select the *Distribution mapping of precipitation and temperature* radio button.
Solely the methods *Linear scaling (multiplicative)*, *Delta-change correction (multiplicative)*, *Precipitation local intensity scaling*, *Power transformation of precipitation*, and *Distribution mapping of precipitation and temperature* have been designed to work with precipitation data and are available for selection.

3.3 Select simulated climate data

CMhyd supports two data formats for simulated climate data

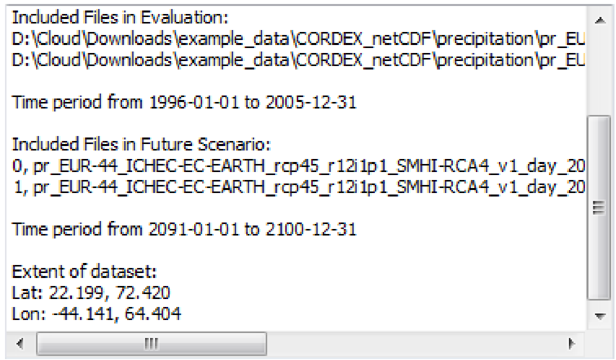

Figure 6: CMhyd example summary of simulated climate data used in bias correction run.

1. netCDF (**.nc*) files (e.g., from the CORDEX data archive) and
2. ASCII input (see Figure 2 and Table 1 and 2).

The second option can be useful if the simulated climate data has already been extracted. With both options, the use of future data is optional and the user can opt to extract and correct historical climate simulation data only.

The netCDF option If the netCDF file option is chosen, the path to the folder containing all **.nc* files needs to be provided. All files used in the current bias correction run need to be stored in the same folder. The combo boxes are used to specify the variable, domain, model, evaluation and future scenario for the current run. Information about the climate simulation data used is summarized in the text box below once the future scenario has been selected (see Figure 6).

The ASCII option If the ASCII option is chosen, the paths to the location files of the historical and (optional) future simulated climate data are required (see Figure 7). Information about the climate simulation data used is summarized in the text box below (see Figure 7).

Example data - Selection of simulated climate data (netCDF format)

Select the *netCDF Input* tab, click on the *Select Folder* button and provide the path to the folder that contains the netCDF precipitation example data (`.../example_data/CORDEX_netCDF/precipitation`). In the example data set only one option is available in the first four combo boxes. In these cases, CMhyd will automatically select the variable (*PCP*), the domain (*EUR-44*), the model (*ICHEC-EC-EARTH*), and the evaluation experiment (*historical*). Finally, use the combo box at the bottom to select the *RCP4.5* scenario. After the scenario has been selected, CMhyd will provide a summary of the simulated climate data used in the current bias correction run in the text box below (see Figure 6).
This manual explains how to use the netCDF option. However, ASCII formatted historical and future simulated climate data is also provided in the example data set.

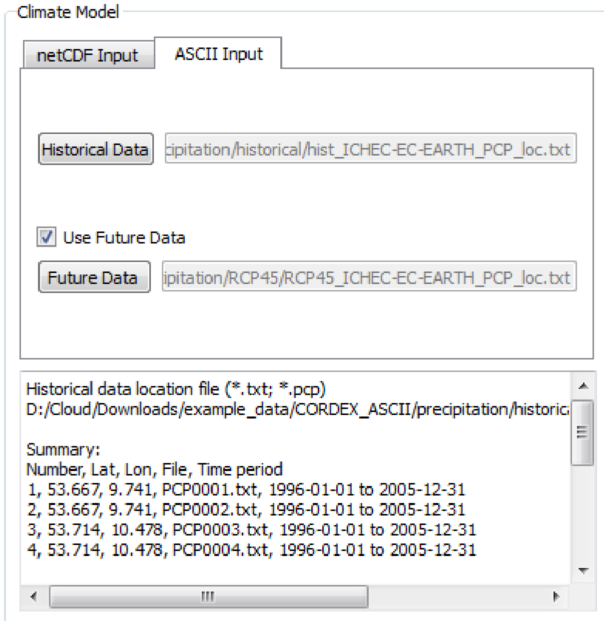

Figure 7: CMhyd ASCII input tab and example summary of simulated climate data used in bias-correction run.

3.4 Select output folder

Select a folder to save the output.

Example data - Select output folder

Create a new sub-folder named *output* in the example data directory (`.../example_data/output`). Then, click the *Select Folder* button and navigate to the *output* folder.

3.5 Process

Pre-processing (*Check Files*) Pre-processing is necessary prior to performing the bias correction. Clicking the *Check Files* button will start pre-processing the data. During this step, CMhyd identifies the climate model grid cells located closest to the gauges, checks the distance between the center of the grid cell and the gauge location, and compares the observed and simulated historical time series. Summarized information is shown in the *Message* text box at the bottom of the GUI (see Figure 8).

Overlapping time periods It is strongly recommended to use observed and simulated historical climate data with a long overlapping time period (i.e. two or three decades). If such an overlapping time period is available, the user can specify whether the parametrization of the bias correction algorithm should be based on (1) the entire observed and simulated time periods, or (2) the time period during which the observed and simulated historical data overlap.

Perform bias correction (*Start Processing*) If the linear scaling or distribution mapping method has been chosen, CMhyd will extract the parameters of the bias correction algorithm in addition to the

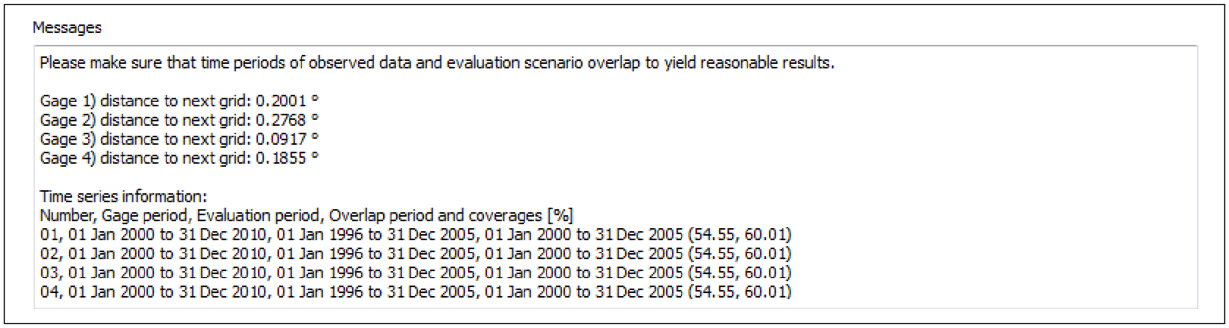

Figure 8: CMhyd example pre-processing summary.

precipitation / temperature time series. These parameters can be used to evaluate the performance of the climate model for each gauge. In particular, the degree of agreement of the gamma and Gaussian distribution have been proven useful indicators of the ability of the climate model to reproduce observed precipitation and temperature, respectively.

The *Message* text box It is recommended to review the information provided in the *Message* text box. Unusual distances between gauges and the nearest climate model grid cell or non-overlapping time periods between observed and simulated climate data might indicate an error in the ASCII formatted input data or in the meta information provided in the netCDF files. In addition, unusually high correction parameters will raise a warning message in the text box if the linear scaling method is used.

CMhyd will raise a warning message if it is not possible to find an adequate parametrization for the bias correction algorithm (e.g., zero monthly mean precipitation values will cause a division by zero error for most of the methods). In these cases, the affected gauges and months will be listed in the *Message* text box and no bias correction will be performed for them. Instead, the original simulated precipitation and temperature values will be used in the historical and future time series.

In any case, it is highly recommended to compare observed, raw simulated, and corrected simulated time series (see next section) prior to their further use.

Output organization – Folder structure The output is stored in the directory specified in Step 3.4. CMhyd generates several sub-folders to organize the output by (1) climate variable, (2) climate model / observed data, and (3) time period. The structure of the output folder varies with the methods and options selected by the user. An example is provided in Figure 9. It is recommended to use the same output folder for additional bias correction runs if the same observed data (but a different climate variable, model, or bias-correction method) is used. CMhyd will automatically save the new output to the appropriate folder and create missing sub-folders.

Output organization – Location and data files CMhyd was designed to provide simulated climate data that can be considered representative for the location of the gauges (see Step 1). Time series are extracted from the climate model defined in Step 3.3. Each of the gauges included in the location files is associated with the climate model grid cell whose center is closest to the gauge. The order of the gauges and climate model grid cells listed in the location files determines the pairing of the data files. The first simulated climate time series listed in the corresponding location file (e.g., *PCP00001*) is associated with first gauge listed in the gauge location file (e.g., *p53997* in Figure 10).

Output organization – File names Several suffixes are used to name the files

- climate variable
  - *PCP* precipitation and
  - *TMP* temperature,
- file type
  - *loc* location files and
  - *obs* observed data,
- time period
  - *ovl* overlapping time period,
  - *hist* simulated historical data, and
  - *sce* simulated future / scenario data,
- original or corrected simulated climate data
  - *raw* original,
  - *LS* linear scaling,
  - *DC* delta change,
  - *DM* distribution mapping,
  - *LI* local intensity scaling
  - *PT* power transformation, and
  - *VS* variance scaling.

Example data - Extracting data and performing the bias correction

Click the *Check Files* button to start the pre-processing procedure, review the message in the *Messages* text box (see Figure 8), check the *Use only overlapping time period to calculate correction parameters* box, and click the *Start Processing* button to apply the distribution mapping bias correction algorithm.
CMhyd will create a *PCP* sub-folder in the selected output directory (see Step 3.4) and the output folder will be organized as shown in Figure 9.
Perform a second precipitation bias correction run using the linear scaling method. CMhyd will automatically add a *LinearScaling* folder (sub-folder of *ICHEC-EC-EARTH*) to the existing folder structure.

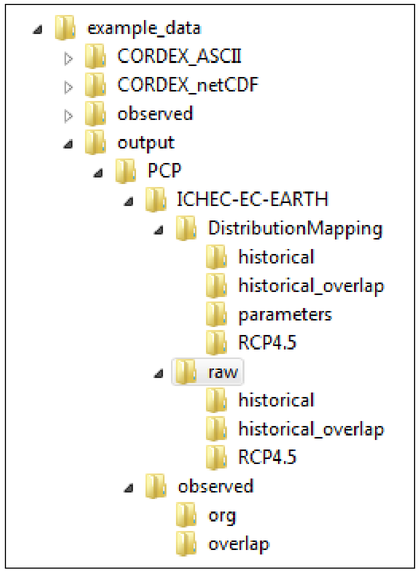

Figure 9: Structure of the *output* folder after performing one precipitation bias correction run using the distribution mapping method and the overlapping time period option.

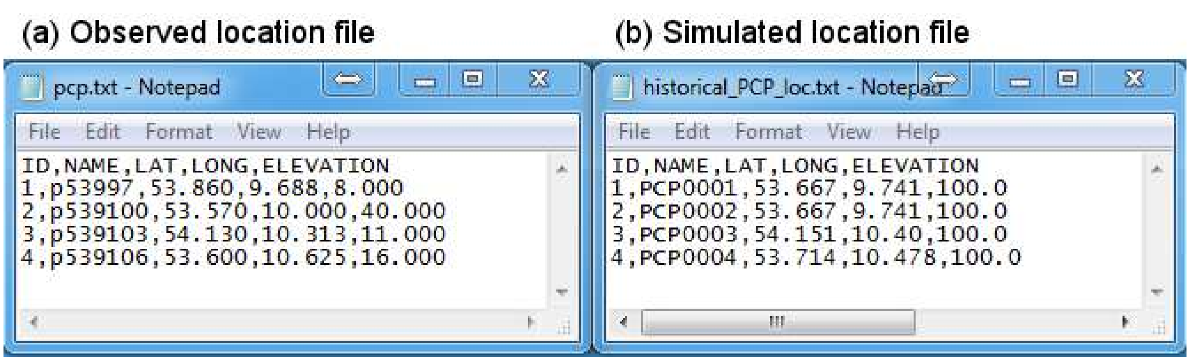

Figure 10: Location files of (a) observed and (b) simulated climate data. The *ID* field determines the pairing of the data files. For example, the simulated data file *PCP0001* is paired with the observed data file *p53997*.

Analyze the data

It is recommended to evaluate and analyze CMhyd output prior to its further use. A detailed evaluation guideline of the climate simulation data and the bias correction methods is provided by Teutschbein and Seibert (2012).

The *Plot* tab can be used to plot a monthly summary, and annual and monthly time series of previously generated CMhyd output for each gauge defined in the location files (see Figure 11). CMhyd compares the data files as listed in the location files (e.g., the first data file of location file one will be compared to the first data file of location file two, . . .). These plots and the parameters provided by the linear scaling and distribution mapping methods are a good starting point for evaluating the bias correction methods and the ability of the climate model to reproduce observed precipitation or temperature values.

The precipitation summary plots include monthly means, 90th percentiles, standard deviations, wet day probabilities, and precipitation intensity. For temperature, monthly means, standard deviation, and 10th and 90th percentiles are plotted for both the minimum and maximum temperature time series. An example of a precipitation summary plot is shown in Figure 13.

It is recommended to use the same output folder as defined in Step 3.4. CMhyd will automatically create a sub-folder *plots* and add folders for the plots (see Figure 12).

Example data - Plotting CMhyd output

To generate plots that show (1) observed precipitation, (2) raw simulated precipitation, and (3) linear scaling-corrected simulated precipitation select *PCP* as the climate variable to be plotted and then set the number of data sets to be plotted to 3. Click the *Select PCP Location File* buttons one to three and provide the path to the location file of the observed[a], raw simulated[b], and linear scaling bias-corrected location[c] files that cover the overlapping time period.
The plots will be saved in the output folder that was defined in Step 3.4. Click the *Create Plots* button to generate the plots. CMhyd will add a sub-folder structure to the output folder (see Figure 12) and save the plots for the four gauges in `.../timeseries` and `.../monthly_summary` in `.../example_data/output/plots/PCP`. The monthly summary and time series plots allow a comparison between observed, raw simulated, and linear scaling corrected time series. The precipitation summary plot for ID 1 is shown in Figure 13.

[a] `.../output/PCP/observed/overlap/obs_ovl_loc.txt`
[b] `.../output/PCP/ICHEC-EC-EARTH/raw/historical_overlap/raw_hist_ovl_loc.txt`
[c] `.../output/PCP/ICHEC-EC-EARTH/LinearScaling/historical_overlap/LS_hist_ovl_loc.txt`

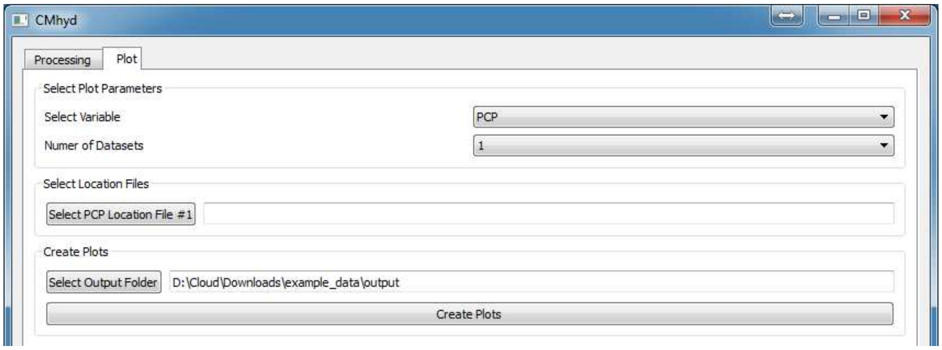

Figure 11: The CMhyd *Plot* tab.

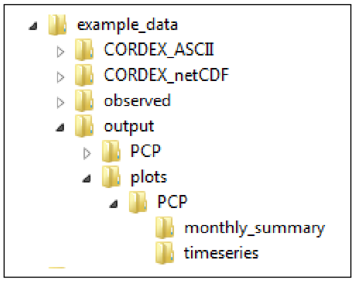

Figure 12: Structure of the *output* folder after creating plots.

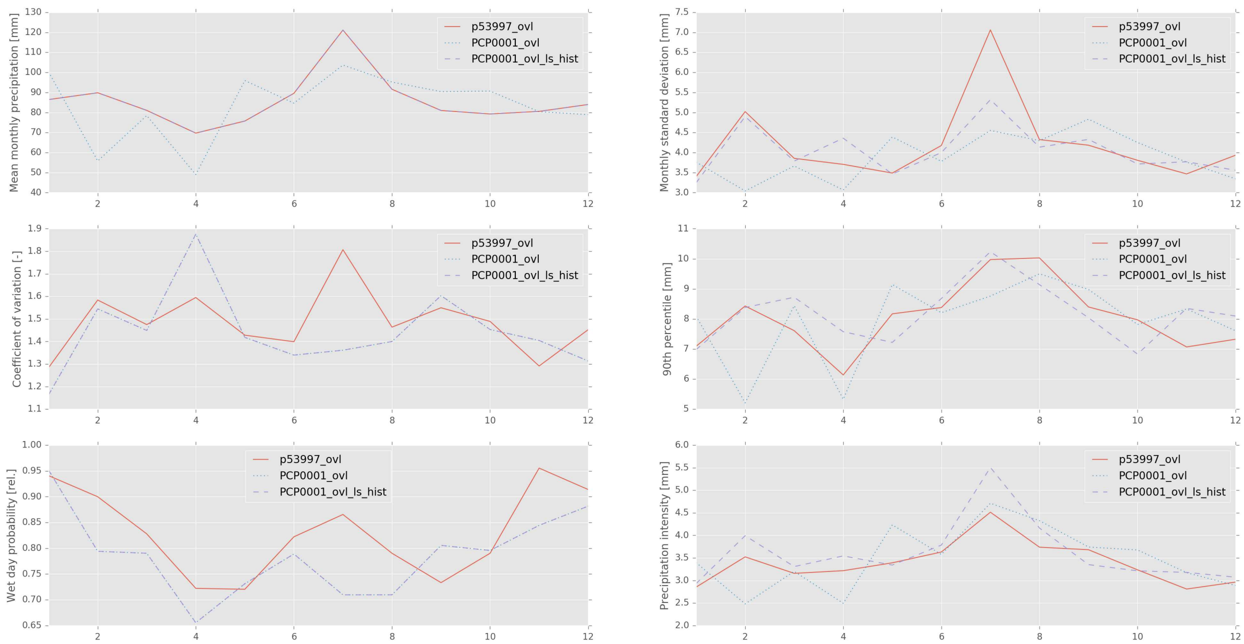

Figure 13: Example of the precipitation summary plot.

References

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment part I: model development. JAWRA Journal of the American Water Resources Association 34, 73–89.

Bates, B., Z.W. Kundzewicz, S.W., (Eds.), J.P., 2008. Climate Change and Water. Technical Paper of the Intergovernmental Panel on Climate Change. IPCC Secretariat.

Christensen, J.H., Boberg, F., Christensen, O.B., Lucas-Picher, P., 2008. On the need for bias correction of regional climate change projections of temperature and precipitation. Geophysical Research Letters 35, n/a–n/a. L20709.

Lafon, T., Dadson, S., Buys, G., Prudhomme, C., 2013. Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. International Journal of Climatology 33, 1367–1381.

Millman, K.J., Aivazis, M., 2011. Python for scientists and engineers. Computing in Science Engineering 13, 9–12.

Oliphant, T.E., 2007. Python for scientific computing. Computing in Science Engineering 9, 10–20.

Salathé, E.P., 2003. Comparison of various precipitation downscaling methods for the simulation of streamflow in a rainshadow river basin. International Journal of Climatology 23, 887–901.

Teutschbein, C., Seibert, J., 2010. Regional climate models for hydrological impact studies at the catchment scale: A review of recent modeling strategies. Geography Compass 4, 834–860.

Teutschbein, C., Seibert, J., 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. Journal of Hydrology 456-457, 12 – 29.

van der Walt, S., Colbert, S., Varoquaux, G., 2011. The numpy array: A structure for efficient numerical computation. Computing in Science Engineering 13, 22–30.

Winchell, M., Srinivasan, R., Di Luzio, M., Arnold, J.G., 2010. ArcSWAT Interface for SWAT 2009 – User's Guide. Texas Agricultural Experiment Station and USDA Agricultural Research Service. Temple (Texas).