BINGO
a better future under
CLIMATE CHANGE

BRINGING INNOVATION TO ONGOING WATER MANAGEMENT

D2.2

Data downscaled to 12km/daily, Europe, for the period 2015-2024

June 2016

www.projectbingo.eu

The BINGO project has received funding from the European Union's Horizon 2020 Research and Innovation programme, under the Grant Agreement number 641739.

Horizon 2020 Societal challenge 5:
Climate action, environment, resource efficiency and raw materials

BINGO

Bringing INnovation to onGOing water management – a better future under climate change

Grant Agreement n° 641739, Research and Innovation Action

Deliverable number:	**D2.2**
Deliverable name:	**Data downscaled to 12km/daily, Europe, for the period 2015-2024**
WP / WP number:	WP2: Climate predictions and downscaling to extreme weather
Delivery due date:	Project month 12 (30/06/2016)
Actual date of submission:	30/06/2016
Dissemination level:	Public
Lead beneficiary:	FUB
Responsible scientist/administrator:	Uwe Ulbrich, Henning Rust
Estimated effort (PM):	12PM
Contributor(s):	Komlan A. Kpogo-Nuwoklo, Edmund Meredith, Henning Rust, Uwe Ulbrich (all FUB)
Estimated effort contributor(s) (PM):	12PM
Internal reviewer:	Paula Lorza, March Scheibel (both Wupperverband), George Zittis (CYI)

Changes with respect to the DoW

With justification if applicable In the DoW, it was stated that the decadal climate predictions (2015-2024) would be downscaled to 12km/1d over the entire EURO-CORDEX region. In order to reduce computational cost, we have instead carried out the downscaling over two European sub-regions (see Figure 1, page 7). These regions cover Iberia and north-west Europe, encompassing all five research sites that the FUB is responsible for providing data for. An additional benefit of this approach is that the reduced computational expense allows us to downscale more realizations of the decadal predictions than would be possible for the whole EURO-CORDEX domain.

Dissemination and uptake

These data can be used by all partners within the project, though cannot be publicly disseminated without prior consent.

Short Summary of results

This deliverable describes the decadal predictions (2015-2024) produced by WP2, to be subsequently used as input data for the hydrological models used in WP3. The data have been dynamically downscaled to a horizontal resolution of 0.11° (12 km) and are available at a daily frequency. As with all data produced in Work Package 2 (see, e.g., Deliverable 2.1), the data are available to all project partners via the online Freva portal, where the data is converted to a format suitable for specific hydrological models and the option of applying a bias correction to the data is also available.

Evidence of accomplishment

1. BINGO-DECO plug-in of the FreVa evaluation system https://freva.met.fu-berlin.de/plugins/bingo_deco/detail/ (log-in required)
2. DECO Documentation: DECO – A plug-in for data extraction and conversion developed within and for BINGO, H.W. Rust, A. Richling, E. Meredith, M. Fischer, C. Vagenas, C. Kadow, and U. Ulbrich, 2016, Technical Report. Available online at http://users.met.fu-berlin.de/~HenningRust/BINGO/DECO_docu.pdf (See also Appendix 2)

TABLE OF CONTENTS

1. Introduction

Work Package 2 – Climate predictions and downscaling to extreme weather – of the BINGO Project is tasked with producing high-resolution climate data, which will be subsequently used as input data for the hydrological models being used in the other work packages. Deliverable 2.2, the focus of this document, is specifically tasked with providing these data for the period 2015-2024, based on decadal climate predictions of the MiKlip decadal prediction system.

The work carried out for Deliverable 2.2 builds on what was previously implemented for Deliverable 2.1. As such, the same approach to data processing, storage and distribution is adopted, which can be read in detail in Deliverable 2.1.

This document describes the data produced for Deliverable 2.2, how it was produced, and how it can be accessed by project partners.

2. Data downscaled to 12km/daily, Europe, for the period 2015-2024

This chapter describes the regional climate simulations carried out for Deliverable 2.2, the data produced for BINGO partners (including their bias correction), and how these data can be accessed by project partners.

2.1. MiKlip forced decadal predictions

The relatively new field of decadal climate prediction, e.g. Smith et al. (2007), aims to simulate both the climate response to future anthropogenic forcing and the future evolution (from the present) of the climate due to internal climate variability (Marotzke et al., 2016). This differs from the approach taken in climate projections, e.g. the CMIP5 project (Taylor et al., 2012), where the focus is on the response of the climate to anthropogenic forcing and the impacts of internal climate variability are (supposed to be) nullified via multi-decadal climate model integrations. The earth system models (ESMs) used in decadal prediction systems are initialized with an observed state of the climate system, i.e. ocean, atmosphere, soil, ice, etc. Skill in predicting internal climate variability on a decadal scale is derived from the long-term memory (i.e. sensitivity to the initial state) of certain components of the climate system, predominantly the ocean. As such, decadal predictions (unlike climate projections) are reliant on a high-quality initialization of the ESM for those components which exhibit long-term memory.

The MiKlip project (http://www.fona-miklip.de) is funded by the German Ministry for Education and Research with the aim of developing a world-class decadal prediction system. The MiKlip decadal prediction system is based on the Max Planck Institute's earth system model, MPI-ESM (http://www.mpimet.mpg.de/en/science/models/mpi-esm/), and has an atmospheric horizontal resolution of T63 (1.875°). The first phase of the MiKlip project showed significant skill (e.g. Müller et al., (2012), Pohlmann et al., (2013)) in the MiKlip system, based on the evaluation of decadal hindcast simulations which were initialized yearly between 1960 and 2010. In addition to this, the MiKlip system was also used for future decadal prediction running up to 2024, for 10 realizations and with an initialization in 2015. Module C of the first phase of the MiKlip project was devoted to the regionalisation of the MiKlip global model output, via dynamical downscaling. This was carried out over a European domain for the entire MiKlip period (1960-2024) using the CCLM at 0.44° resolution.

For the BINGO project, the FUB have further dynamically downscaled the first four realizations of the future decadal predictions (2015-2024) from 0.44° to 0.11° using the CCLM (see the attached Technical Report (Rust et al.), Appendix 2, for a description of the CCLM). To reduce computational expense, this has been carried out over two sub-domains (Fig. 1):

(1) **NW-EUR-11**: which contains the research sites at Bergen, Veluwe and Wupper.

(2) **IBERIA-11**: which contains the research sites at Tagus and Badalona.

The 0.11° resolution matches that used in the downscaling carried out for D2.1, creating a total set of simulations spanning 1979-2024, crucially all at the same spatial resolution.

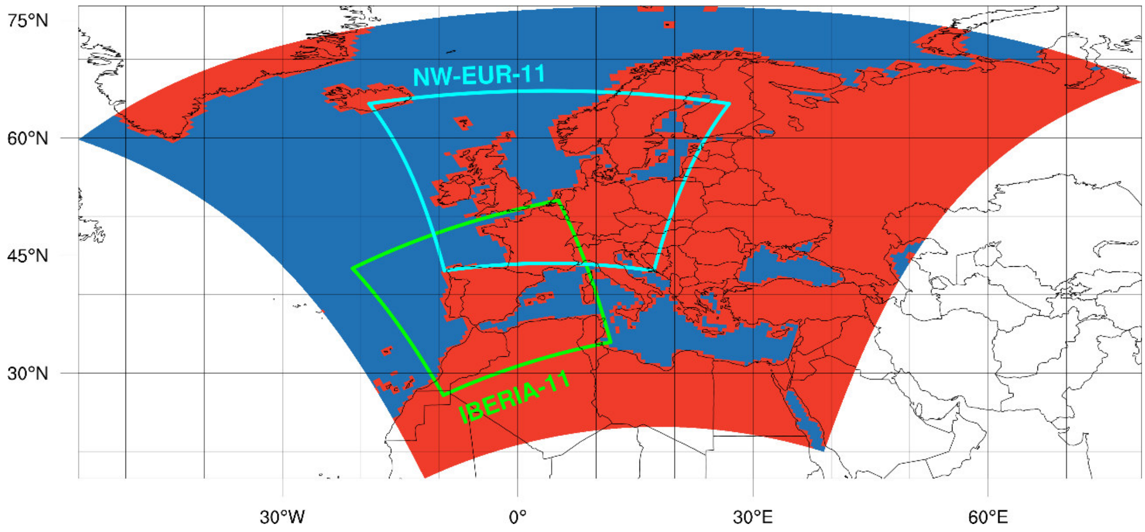

Figure 1 | BINGO domains for MiKlip decadal predictions. The area covered by the MiKlip 0.44° regionalisation is shown in colour, with red representing the 0.44° land mask and blue the sea. The IBERIA-11 and NW-EUR-11 domains of the BINGO project are shown in green and cyan.

The variables, all of which can be bias corrected, produced for BINGO partners and their temporal resolution also match those of D2.1. That is, D2.2 provides the following variables at a daily frequency, as well as their daily min/max values:

Total Cloud Fraction

Near-Surface Relative Humidity

Near-Surface Specific Humidity

Precipitation

Surface Air Pressure

Sea Level Pressure

Surface Downwelling Longwave Radiation

Surface Downwelling Shortwave Radiation

Near-Surface Wind Speed

Near-Surface Air Temperature

Near-Surface Dew Point Temperature

Eastward Near-Surface Wind

Northward Near-Surface Wind

2.2. Bias correction of decadal predictions

As discussed in detail in Deliverable 2.1, climate models often show systematic biases with respect to observations in their representation of certain variables. As such, the later work packages are best served by having the option to input bias corrected data to their hydrological models.

Bias correction of the 0.11° decadal predictions is carried out using a method based on the cumulative distribution functions of the model output and historical observations. Broadly speaking, this involves estimating the bias of historical MiKlip data and bias correcting the future MiKlip data based on this. This method also takes account of the fact that the highest resolution historical data available from the MiKlip project is 0.44°, and that having produced 0.11° data for the decadal future predictions, BINGO does not have the computational resources to additionally produce 0.11° resolution data for the historical period. The concept behind the method is described in detail in Appendix 1 and in section 3.3 of Appendix 2.

2.3. Data dissemination

The dissemination method for the data provided for D2.2 is via the Freva system, using the DECO plugin. This is identical to the method adopted for D2.1; users simply need to select the correct experiment (“decadal-predictions”) from the DECO menu. As such, the method will not be described in detail again here. For a recap, the reader is referred to either Appendix 2 or Deliverable 2.1.

3. References

J Marotzke, WA Müller, FSE Vamborg, P Becker, U Cubasch, H Feldmann, F Kaspar, C Kottmeier, C Marini, I Polkova. Miklip – a national research project on decadal climate prediction. Bulletin of the American Meteorological Society, 2016.

WA Mueller, J Baehr, H Haak, JH Jungclaus, J Kröger, D Matei, H Pohlmann, JS Storch, and J Marotzke. Forecast skill of multi-year seasonal means in the decadal prediction system of the max planck institute for meteorology. Geophysical Research Letters, 39(22), 2012.

H. Pohlmann, W. A. Mueller, K. Kulkarni, M. Kameswarrao, D. Matei, FSE Vamborg, C. Kadow, S. Illing, and J. Marotzke. Improved forecast skill in the tropics in the new miklip decadal climate predictions. Geophysical Research Letters, 40(21):5798-5802, 2013.

D. M. Smith, S. Cusack, A. W. Colman, C. K. Follard, G. R. Harris, and J. M. Murphy. Improved surface temperature prediction for the coming decade from a global climate model. Science, 317(5839):796-799, 2007.

K. E. Taylor, R. J. Stouffer, and G. A. Mehl. An overview of CMIP5 and the experiment design. Bull. Amer. Meteor. Soc., 94(4):485-498, 2012.

Appendix 1

Bias Correction of MiKlip Decadal Predictions

A proposed scheme for MiKlip future simulations bias correction

June 29, 2016

Our goal is to bias correct the Miklip future simulations for a spatial resolution of 0.11° using the CDF-Transform method. This bias correction method can be seen as an extension of the classical quantile mapping in which model simulations are corrected so that their Cumulative Distribution Functions (CDF) match the ones of the observations. The CDF-Transform method, in addition, aims to provide approximative CDF of the observations for the future period based on information learned from the historical period (see Appendix 2, section 3.3, for more details).

In the Table 1, the available data for historical and future periods are summarized. WATCH data serves here as observations.

	Resolution 0.44°		Resolution 0.11°	
	Historical	Future	Historical	Future
MiKlip	$F_h^{(M44)}$	$F_f^{(M44)}$	$F_h^{(M11)}$	$F_f^{(M11)}$
WATCH	$F_h^{(W44)}$	$F_f^{(W44)}$	$F_h^{(W11)}$	$F_f^{(W11)}$

Table 1: Table summarizing the data: F stands for CDFs. CDFs that can not be computed directly from the available data are indicated in blue.

The bias correction of MiKlip future simulations for spatial resolution of 0.11° is quite challenging because historical simulations for the same resolution are not available. However, we have Miklip simulations for both historical and future periods at a spatial resolution of 0.44°. Thanks to the CDF-Transform, it is possible to estimate CDFs of MiKlip historical simulations for spatial resolution of 0.11° by modelling the mathematical link between these two resolutions. Thus, the proposed method to correct the MiKlip future simulations for spatial resolution of 0.11°, using CDF-Transform can be listed as followed (see Figure 1 for graphic illustration):

1. Estimate CDFs of the MiKlip historical simulations for a spatial resolution of 0.11° ($F_h^{(M11)}(x)$) using information provided by the large scale (0.44°) MiKlip data (Section 1).

2. Estimate CDFs of the WATCH data (0.11°) for the future period ($F_f^{(W11)}(x)$) using the MiKlip historical and future simulations and WATCH historical data for spatial resolution of 0.11° (Section 2).

3. Once CDFs of the WATCH data (0.11°) for the future period are estimated, a quantile mapping can finally be applied between $F_f^{(M11)}(x)$ and $F_f^{(W11)}(x)$. The corrected value x_{bc} is then

$$x_{bc} = F_f^{(W11)-1}\left[F_f^{(M11)}(x)\right]. \tag{1}$$

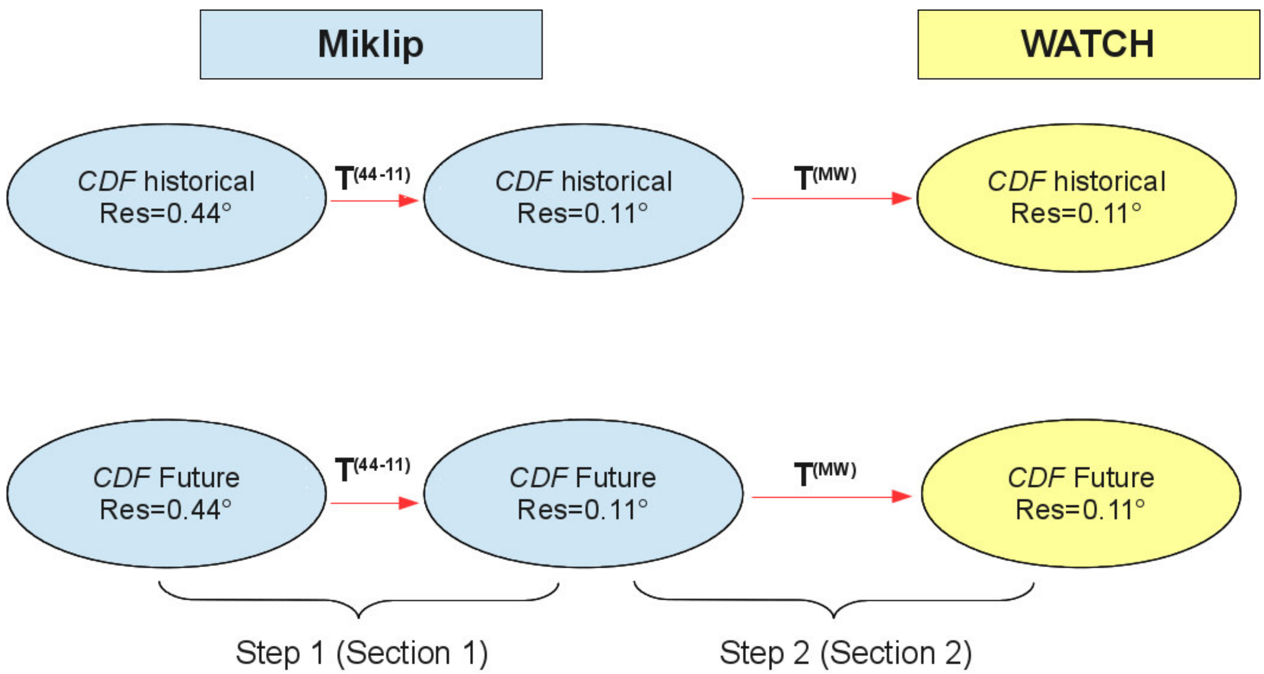

Figure 1: Illustration of transformations involved in the proposed method

1 Estimation of MiKlip historical CDF ($F_h^{(M11)}(x)$) from the MiKlip $0.44°$ data: probabilistic downscaling

We also make the assumption that there is a mathematical transformation $T^{(44-11)}$ that allows to go from the CDF of the large scale MiKlip historical data ($F_h^{(M44)}(x)$) to the CDF of the small scale MiKlip historical data ($F_h^{(M11)}(x)$) ,

$$T^{(44-11)}\left[F_h^{(M44)}(x)\right] = F_h^{(M11)}(x). \tag{2}$$

Under assumption that $T^{(11-44)}$ still valid for the future period,

$$T^{(44-11)}\left[F_f^{(M44)}(x)\right] = F_f^{(M11)}(x), \tag{3}$$

it can be modelled by replacing x by $F_f^{(M44)-1}(u)$ in (3), where u is any probability in $[0, 1]$,

$$T^{(44-11)}(u) = F_f^{(M11)}\left[F_f^{(M44)-1}(u)\right]. \tag{4}$$

Inserting (4) in (2) leads to a modelling of $F_h^{(M11)}(x)$,

$$F_h^{(M11)}(x) = F_f^{(M11)}\left\{F_f^{(M44)-1}\left[F_h^{(M44)}(x)\right]\right\}. \tag{5}$$

2 Estimation of WATCH future CDF ($F_f^{(W11)}(x)$)

We make assumption that there is a mathematical transformation $T^{(MW)}$ that allows to go from the CDF of the MiKlip historical data ($F_h^{(M11)}(x)$) to the CDF of the WATCH historical data ($F_h^{(W11)}(x)$),

$$T^{(MW)}\left[F_h^{(M11)}(x)\right] = F_h^{(W11)}(x). \tag{6}$$

We suppose that this transformation is still valid for the future period,

$$T^{(MW)}\left[F_f^{(M11)}(x)\right] = F_f^{(W11)}(x). \tag{7}$$

By modelling $T^{(MW)}$, it is then possible to approximate the CDF of WATCH future data ($F_f^{(W11)}(x)$). The simple way to model $T^{(MW)}$ is to replace x by $F_h^{(M11)-1}(u)$ in (6), where u is any probability in $[0, 1]$. We then obtain

$$T^{(MW)}(u) = F_h^{(W11)}\left[F_h^{(M11)-1}(u)\right], \tag{8}$$

corresponding to the simple definition of $T^{(MW)}$. Inserting (8) in (7) leads to a modelling of $F_f^{(W11)}(x)$,

$$F_f^{(W11)}(x) = F_h^{(W11)}\left\{F_h^{(M11)-1}\left[F_f^{(M11)}(x)\right]\right\}. \tag{9}$$

In equation (9) the inverse CDF of the historical Miklip data $F_h^{(M11)-1}$ is unknown and needs to be estimated. One way to do so is to apply the CDF-t as a probabilistic downscaling to relate CDFs from the large scale ($0.44°$) to CDFs from the small scale ($0.11°$).

Appendix 2

Technical Report: Available data and the Freva system/DECO plugin through which it is distributed

BINGO Work package 2
Climate predictions and downscaling to extreme weather

DECO

—

A plug-in for data extraction and conversion developed within and for BINGO

Henning W. Rust, Andy Richling, Edmund Meredith, Madlen Fischer, Christos Vagenas, Christopher Kadow and Uwe Ulbrich

Freie Universität Berlin

Version from June 28, 2016

The BINGO project has received funding from the European Union's Horizon 2020 Research and Innovation programme, under the Grant Agreement number 641739.

Abstract

DECO is a plug-in to the *Freie Universität Berlin Evaluation Framework for Earth System Science* (FreVa) and aims at extracting and converting COSMO-CLM regional climate model simulations from a central data storage on demand via a web-based platform. The data is to be used as meteorological driving data for hydrological models at the six BINGO Research Sites (BINGO deliverable D2.1). Currently, the spatial resolution of the regional model is 12km and data will be made available as daily values. The climatology for the simulated meteorological parameters, i.e. their seasonally varying mean (and higher moments) differ from the observed one, a bias correction can be optionally applied before converting the data and file format to the particular needs of the individual modeling groups at the Research Sites. The latter implies a conversion from the native COSMO-CLM grid to station locations or to a different grid, a change of units, as well as writing the data to the desired file format. This on-demand post-processing and conversion approach allows for an efficient data storage, maximal reproducibility and transparency, as well as transferability to new data sets. The application developed here can be accessed via a web-platform or a command line interface. Development is subjected to a strict version control to ensure reproducibility.

Contents

Chapter 1

DECO – Aims and strategy

The BINGO Work Package WP2 *Climate predictions and downscaling to extreme weather* aims at providing high resolution meteorological driving data for various hydrological models. For most models, this includes precipitation, temperature, pressure, wind speed, incoming solar radiation and others. This data is generated on a regional level (for this deliverable at a European level) by dynamically downscaling coarse resolution global data, see also the red box of Fig. 1.1. To be usable for hydrological models, this data

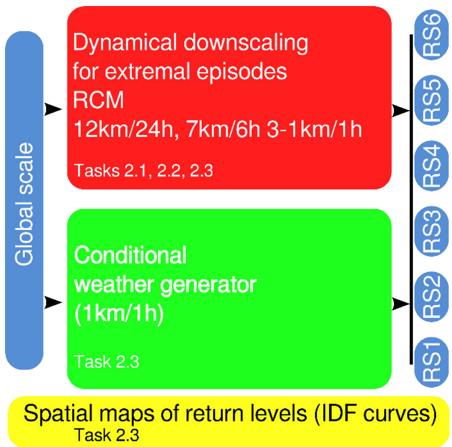

Figure 1.1: Sketch of the strategy in WP2.

needs to be appropriately post-processed and converted to the needs of the 15 and more individual models used at the 6 BINGO Research Sites.

For transparency and reproducibility within BINGO, we must ensure that the very same driving data is available throughout the project and beyond. Thus, it must be either stored in the various different file formats needed by the individual models, or – to be more efficient on memory consumption – one data set is stored in a common standardized format for climate models and additionally data conversion algorithms are specifically tailored for all the individual hydrological models. Besides the efficient use of memory, there are other advantages to the latter procedure: I) these conversion algorithms can be reused for more data to come with the following deliverables associated with WP2, II) bias correction can be exchange with a more sophisticated one if available, and III) many other available data sets in this standardized format can be used additionally to the data produced particularly for BINGO.

Given these advantages, WP2 leaders decided to develop a hybrid application (command-line-interface and web-based) to extract the data needed for the Research Sites, post-process and convert it to the needs of all individual modeling groups. These conversion routines can also be adapted to incorporate new models or to changing needs of existing models. Furthermore, users can also get back to data generated earlier and extract these using, e.g., a new bias correction method or a slightly changed conversion routine. Additionally to efficiently archiving climate data in a central place, the plug-in developers in WP2 use git[1] to ensures a proper versioning of the bias correction and conversion algorithms.

During the development of the plug-in comments were requested directly from the modeling groups at the Research Site. Hydrological modelers were also asked to test the output of the plug-in with their models and review the plug-in during development. As for all pieces of software, also the development of DECO is not finished but open to being continued. Conversion routines for other models could be integrated, as well as new bias correction schemes.

This document describes the data generated with the COSMO-CLM driven by ERA-Interim (Sect. 2), the bias correction used (Sect. 3) and the extraction and conversion algorithms for the individual models at the research Sites (Sect. 4).

[1] https://en.wikipedia.org/wiki/Git_%28software%29

Chapter 2

Description of the climate simulations

All simulations described in this chapter have been carried out using the COSMO-CLM regional climate model (Rockel et al., 2008), for a domain centred over Europe. In this chapter, a description of the COSMO-CLM regional model is provided, followed by details of the simulations performed.

2.1 Model Description

The COSMO-CLM (CCLM) is a state-of-the-art nonhydrostatic regional climate model, that is the climate version of the COSMO numerical weather prediction model used by the German Weather Service (DWD). CCLM is developed for climate purposes by the CLM-Community (`http://www.clm-community.eu/`) and features a software architecture allowing for computational parallelism and system extensibility. It is suitable for a broad spectrum of applications across scales ranging from hundreds of metres to thousands of kilometres. The model components include an atmospheric model directly coupled to a land-surface/soil model, and an aerosol model. Sea surface temperatures must be prescribed as a boundary condition.

2.2 Description of Simulations

2.2.1 Reanalysis-forced evaluation runs

To create a climatology and continuous time series of key hydorological variables, downscaling simulations have been carried out with the CCLM over the EURO-CORDEX domain (Fig. 2.1), with a horizontal resolution of 0.11° (about 12km) and spanning the time period 1979-2015. At the domain lateral boundaries, the model is constrained by the latest generation of reanalyses from the European Centre for Medium-range Weather Forecasting, ERA-Interim (Dee et al., 2011), which provide updated lateral boundary conditions every 6 hours. The entire period covered by the ERA-Interim reanalysis has been downscaled (1979-2015). Initial conditions, sea surface temperature, and sea-ice cover also come from the ERA-Interim reanalysis.

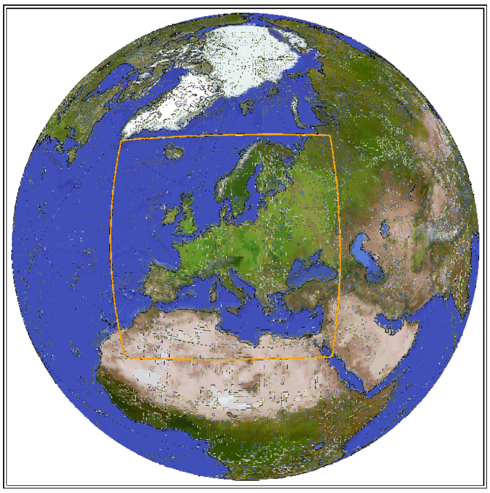

Figure 2.1: EURO-CORDEX Domain. The EURO-CORDEX domain is a standard reference domain for Europe defined as part of the CORDEX Experiment (http://www.cordex.org/).

Computational limitations prohibit simulation at the kilometre-scale over such a large domain and time period. A horizontal resolution of 0.11° has thus been chosen, which represents somewhat of a "breakthrough" resolution and has been shown to add significant value to coarser global model output for the simulation of (non-convective) precipitation extremes (Heikkilä. et al., 2011), particularly in mountainous regions, and can modulate the climate change signal of coarse resolution global models (Torma et al., 2015).

The simulations have been carried out via three separate model runs, each of which is continuous during the time period indicated in Tab 2.1 (middle column). Runs 1 and 3 were carried out at the Freie Universität Berlin; run 2 had previously been carried out by the CLM community, and we utilise this to keep compuational expense to a minimum. Within each model run, it is necessary to allow sufficient time after initialisation for the spin-up of soil moisture and soil-related processes within the higher resolution CCLM. In each run, a minimum of two months is allowed for spin-up. Model output from this period is thus not included in the final data, i.e. the data to be used for research purposes. The final data thus covers the period March 1979 - July 2015, summarised in Tab 2.1 (right column). The model output can be extended beyond July 2015 if there is agreement amongst project partners and if reanalysis data for that period is available.

For most key BINGO variables, 0.11° CCLM output is available at 3-hourly frequency. The exception to this is precipitation, which is available at hourly frequency. BINGO variables available and their temporal frequencies are summarised in Table 2.2.

Model Run	Period of Run	Data Available
1	01.01.1979 - 01.04.1989	01.03.1979 - 01.03.1989
2	01.01.1989 - 01.01.2009	01.03.1989 - 01.12.2009
3	01.09.2008 - 01.08.2015	01.12.2009 - 01.08.2015

Table 2.1: List of model runs.

2.2.2 High-resolution (test) simulations of extremal episodes

Theory and Methods

Extremal weather patterns and individual events of hydrological significance shall be identified from the 0.11° CCLM simulations, for each of the 6 research sites. These shall be subjected to more detailed analysis. In the case of extreme precipitation events, which are a phenomenon with high spatial variability, this involves further dynamic downscaling of the identified events up to a convection-permitting resolution of 0.02° (2.2 km) to provide better representation of the dynamics driving any changes in hydrological extremes and hence a more detailed input for the hydrological models. In convection-permitting simulations, deep convective processes can be explicitly simulated by the model, where they have to be otherwise parametrized in lower-resolution simulations, i.e. our 0.11° CCLM integration. The further downscaling to convection-permitting resolution is a key step, as recent studies have shown that convection-permitting resolution is essential to accurately capture the response of convective precipitation extremes to climatic changes (Kendon et al., 2014; Ban et al., 2015), which can be highly nonlinear (Meredith et al., 2015).

The issue of spatial spinup - meaning the distance from the lateral boundaries at which fine-scale features can be achieved - is an important consideration when designing regional downscaling experiments. For consistency, we intend to use a common domain for all high-resolution simulations at each research site. As the large-scale forcing behind individual extremes can come from any side of the domain, we centre our high-resolution domains over each research site and use a 201 x 201 grid, with 50 vertical levels. This allows at least 100 grid-lengths between the lateral boundaries and the centre of each research site. Brisson et al. (2015) investigated the impact of domain size on the simulation of precipitation in convection-permitting models, using a horizontal grid spacing of 3 km. They concluded that a spatial spinup of at least 40 grid cells is necessary for the realistic simulation of precipitation patterns. For more detailed discussion of convection-permitting modelling the reader is referred to Prein et al. (2015).

The Simulations

With the aid of the questionnaire responses from each research site, **one extreme precipitation event** has been identified from the 0.11°-CCLM simulations for each site

Variable Description	Variable ID	Frequency	Time Method*	Min/Max†
Total Cloud Fraction	clt	3-hr, day	Instantaneous	Daily
Near-Surface Relative Humidity	hurs	3-hr, day	Instantaneous	Daily
Near-Surface Specific Humidity	huss	3-hr, day	Instantaneous	Daily
Precipitation	pr	1-hr, 3-hr, day	Mean	Daily
Surface Air Pressure	ps	3-hr, day	Instantaneous	Daily
Sea Level Pressure	psl	3-hr, day	Instantaneous	Daily
Surface Downwelling Longwave Radiation	rlds	3-hr, day	Mean	Daily
Surface Downwelling Shortwave Radiation	rsds	3-hr, day	Mean	Daily
Near-Surface Wind Speed	sfcWind	3-hr, day	Instantaneous	Daily
Near-Surface Air Temperature	tas	3-hr, day	Instantaneous	Daily
Near-Surface Dew Point Temperature	tdps	3-hr, day	Instantaneous	Daily
Eastward Near-Surface Wind	uas	3-hr, day	Instantaneous	Daily
Northward Near-Surface Wind	vas	3-hr, day	Instantaneous	Daily

Table 2.2: BINGO Variables. *Time method refers to the highest frequency data; daily means are calculated by averaging the highest frequency data. †Variable IDs for daily min/max are formed by appending either 'min' or 'max' to the variable ID, e.g. clt[min|max].

A "day" is defined as 24-hours from 00:00 UTC. All time-stamps in the output data are in UTC.

(excluding Cyprus), and has been further downscaled to 0.02° resolution (2.2 km) with the CCLM. The events for these **test-simulations** were subjectively identified from the 0.11° degree model output, based on the questionnaire descriptions of past extremes at each site and the 0.11° modelled precipitation.

The output variables are at an hourly frequency and have been made available for the same parameters as shown in Table 2.2, though there are obviously no daily min/max's provided. All data are available through the Freva DECO plugin, and are best accessed by selecting "**test-events**" in the *experiment* field and then the appropriate research site (i.e. Badalona, Bergen, Tagus, Veluwe, Wupper) in the *product* field. As of 10-06-2016, the test-events for the Tagus research site are not yet available through the DECO plugin. We hope to remedy this asap, after clarifying the input requirements of the particular hydrological model being used for the site.

Project partners are asked to download the high-resolution test-simulations for their respective research sites and test the data on their hydrological models. Feedback should then be provided as soon as possible. The earlier feedack is received, the more likely that any concerns raised can be satisfactorily addressed. Feedback on the test-simulations is best provided to Edmund Meredith (edmund.meredith@met.fu-berlin.de).

2.2.3 MiKlip forced decadal predictions

The relatively new field of decadal climate prediction, e.g. Smith et al. (2007), aims to simulate both the climate response to future anthropogenic forcing and the future evolution (from the present) of the climate due to internal climate variability (Marotzke et al., 2016). This differs from the approach taken in climate projections, e.g. the CMIP5 project (Taylor et al., 2012), where the focus is on the response of the climate to anthropogenic forcing and the impacts of internal climate variability are (supposed to be) nullified via multi-decadal climate model integrations. The earth system models (ESMs) used in decadal prediction systems are initialized with an observed state of the climate system, i.e. ocean, atmosphere, soil, ice, etc. Skill in predicting internal climate variability on a decadal scale is derived from the long-term memory (i.e. sensitivity to the initial state) of certain components of the climate system, predominantly the ocean. As such, decadal predictions (unlike climate projections) are reliant on a high-quality initialization of the ESM for those components which exhibit long-term memory.

The MiKlip project (`http://www.fona-miklip.de`) is funded by the German Ministry for Education and Research with the aim of developing a world-class decadal prediction system. The MiKlip decadal prediction system is based on the Max Planck Institute's earth system model, MPI-ESM, and has an atmospheric horizontal resolution of T63 (1.875°). The first phase of the MiKlip project showed significant skill (e.g. Mueller et al. (2012), Pohlmann et al. (2013)) in the MiKlip system based on the evaluation of decadal hindcast simulations initialized yearly from 1960-2010. In addition to this, the MiKlip system was also used for future decadal prediction running up to 2024, for 10 realizations and with an initialization in 2015. Module C of the first phase of the MiKlip project was devoted to the regionalisation of the MiKlip global model output, via dynamical downscaling. This was carried out over a European domain for the entire MiKlip period (1960-2024) using the CCLM at 0.44° resolution.

For the BINGO project, the FUB have further dynamically downscaled four realizations of the future decadal predictions (2015-2024) from 0.44° to 0.11° using the CCLM. To reduce computational expense, this has been carried out over two sub-domains (Fig. 2.2):

(1) **NW-EUR-11**: which contains the research sites at **Bergen**, **Veluwe** and **Wupper**.
(2) **IBERIA-11**: which contains the research sites at **Tagus** and **Badalona**.

The same variables and frequencies as listed in Table 2.2 are available and are downloadable via the online DECO plugin.

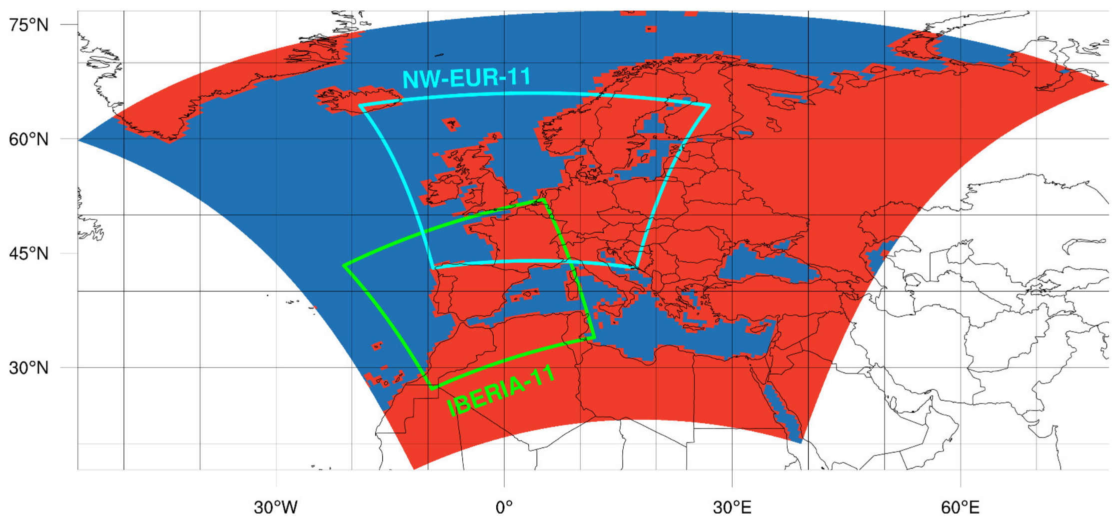

Figure 2.2: BINGO Domains for MiKlip decadal predictions (2015-2024). The area covered by the MiKlip 0.44° regionalisation is shown in colour, with red representing the 0.44° land mask and blue the sea. The IBERIA-11 and NW-EUR-11 domains of the BINGO project are then marked in green and cyan.

Chapter 3

Bias correction

Typically, systematic differences between the climate model simulation and observed data exist. The most prominent difference is a shift in the mean value. Climate model simulations are thus typically post-processed using a bias correction. This chapter aims at outlining this kind of post-processing. Two different bias correction methods are presented in this chapter: Seasonal Generalized Linear Model method and Cumulative Distribution Function Transform method.

The first Section 3.1 covers the reference datasets used. The Seasonal Generalized Linear Model method is describe in 3.2 followed by a description of Cumulative Distribution Function Transform method in Section 3.3.

3.1 Reference data

Depending on the Research Sites and hydrological models, the reference data for bias correction varies. In case the driving data is requested gauge based and gauges based reference data is available, we use this data for bias correction. For gridded products, we use the WATCH forcing data ERA-Interim (WFDEI, Weedon et al., 2014) as a reference in case no other gridded reference product was provided. As bias correction of gridded products based on gauge-based reference data is a lot less straightforward, this will not be included in this deliverable. The following list gives an overview over the reference data used at different Research Sites.

RS1 Bergen a gridded data product is requested and thus WFDEI dataset is used as reference.

RS2 Veluwe a gridded data product is requested and thus WFDEI dataset is used as reference.

RS3 Wupper a gridded data product is requested and thus WFDEI dataset is used as reference.

RS4 Badalona a gridded data product is requested and thus WFDEI dataset is used as reference.

RS5 Tagus-River (Portugal) a gauge-based data product for the the variables precipitation (mm/day), daily maximum and minimum near-surface air temperature (°C), surface downwelling shortwave flux in air (W/m^2), wind speed (m/s) and surface air pressure (kPa) at daily resolution is to be provided. Gauge locations have been provided but not all gauges record all the requested variables. Consequently, bias correction can only been made for those quantities where reference has been provided. That is

- Maximum and minimum temperature, wind speed (monthly) for Tapada da Ajuda, Salvaterra de Magos, Dois Portos, Santarem, Alvega, and
- Precipitation (daily) for Vila Nogueira, Moinhola, Canha, Barragem de Magos, Barragem de Montargil, Ota, Marianos, Santarem ESA, Tojeiras de Cima, Bemposta, and Pernes.

RS6 Troodos-Mountains no data products requested.

For most of the Research Sites a seasonally resolved climatology computed from WFDEI is the reference for the climate simulations. This forcing data set is been frequently used in the context of hydrological modeling (e.g., Gudmundsson et al., 2011; Koch et al., 2013; Prudhomme et al., 2014). However, for this data set, Rust et al. (2015) found that due to the way of merging the ERA-Interim reanalysis with a gridded observation-based data product from CRU implausible differences in daily temperatures across boundaries of calender month might arise for some regions. For Europe, however, these differences are insignificant.

For bias correction the following variables of the WFDEI are available: mean/min/max temperature, total precipitation, surface air pressure, near-surface wind speed, long-/shortwave incident radiation and near-surface specific humidity. A bias correction for mean/min/max relative humidity will be done in later stages. WFDEI does not provide vectorial information of the wind nor its directions, thus wind direction cannot be corrected. This is, however, not a problem for most of the cases relevant to BINGO. WFDEI comes on a coarser resolution than the COSMO-CLM simulation used for D2.1. it has thus been interpolated to the grid of the COSMO-CLM.

3.2 Seasonal Generalized Linear Model method

In this Section, the Seasonal Generalized Linear Model bias correction method is described. The underlying idea of the approach is first presented (Sect. 3.2) followed by the concrete application of the approach at the BINGO Research Sites in Sect. 3.2.2.

3.2.1 Underlying principle and modeling approach

The underlying idea of the bias correction applied here is the assumption that the climatological seasonal cycle of both, simulations and observations, is a smooth function of the day of the year. If the simulated seasonal cycle does not match the observed one, it needs to be adjusted. The smooth functions are modeled using a generalized linear model (GLM, McCullagh and Nelder, 1989) with harmonic function (sine and cosine) of the day of the year as predictors. Periodic functions, such as the seasonal cycle, can be always described with a series of harmonic functions(e.g., Priestley, 1992); the more features the cycle has, the higher the order of the series expansion must be. Generalized linear models are not restricted to Gaussian residuals as the standard linear regression is, Residuals can be from any distribution in the exponential family of probability distributions (McCullagh and Nelder, 1989), e.g. Exponential, Gamma, Binomial, Poisson. For our cases particularly interesting distributions are those with positive support for modeling precipitation and other non-negative quantities.

Once the two seasonal cycles have been obtained (modeling step), a difference (or ratio in case of precipitation) is obtained and this is used for adjusting the simulated data (adjustment step).

The model

As a simultaneous treatment of all data is advantageous over a separate treatment of data in different months, the seasonal variations of the variables can be captured by using harmonic functions, as mentioned above. For the generalized linear model, such a description is given in Eq. (3.1) for the expectation value $\mu(t)$.

$$\mu(t) = \mu_0 + \sum_{n=1}^{N} \mu_{n1} \sin(n\omega \cdot t) + \sum_{n=1}^{N} \mu_{n2} \cos(n\omega \cdot t) \tag{3.1}$$

with $\omega = \frac{2\pi}{365.25}$, $t = 1, \ldots, 366$ being the time variable running over all possible days of the year. For parameter estimation, t will be centered at the months of the year; a description of the seasonal cycle is, however, possible at a daily time resolution, thus $t = 1, \ldots, 366$. The choice of distribution for the residuals (anomalies) does vary with the meteorological parameter considered, the model for the expectation given in Eq. (3.1) remains basically the same.

Distributional assumptions

Precipitation Precipitation is a somewhat particular quantity. It shows a continuous probability distribution for strictly positive values but has a discontinuity at zero. For a statistical description, this is typically captured with a compound model consisting of a Binomial variable for describing dry and wet days and a strictly positive variable (Gamma or Exponential) for the quantity of precipitation on rainy days. Here, we adjust only the amount of precipitation on rainy days and not the distribution of dry and wet days. This is to avoid inconsistencies in the post-processing model simulations, such as precipitation on days with no clouds.

Other variables Table 3.1 gives an overview over all variables and the associated model distributions. The order of the harmonic series expansion (model selection) has

variable	long name	distribution
T_1	difference between T_{max} and T_{min}	log-Gaussian
T_2	sum of T_{max} and T_{min}	Gaussian
T_{mean}	mean surface temperature	Gaussian
RR	precipitation	gamma
v	near-surface wind speed	log-Gaussian
P	surface air pressure	Gaussian
LW_{down}	longwave incident radiation	log-Gaussian
SW_{down}	shortwave incident radiation	log-Gaussian
Q_{air}	near-surface specific humidity	log-Gaussian

Table 3.1: Bias-corrected variables and associated model distributions.

been chosen on the reference datasets. Harmonic series to 5th order have been considered and selected with the Bayesian Information Criterion (BIC). The model thus obtained has been used to describe the climate model simulations and the reference data.

Minimum and maximum temperature

Particular care needs to be taken when correction minimum and maximum temperature to avoid inconsistencies such as $T_{max} < T_{min}$. Here, a variable transformation given in Eq. (3.2) ensures physical consistency.

$$T_1 = T_{max} - T_{min} \tag{3.2}$$
$$T_2 = T_{max} + T_{min} \tag{3.3}$$

After correcting T_1 and T_2 based on the reference, the corrected values for T_{max} and T_{min} can be derived by back-transforming the variables.

3.2.2 At BINGO Research Sites

For reasons of data availability and for a robust fit, we use monthly means to estimate the model parameters (coefficients of the harmonic functions). For some cases, a vector generalized linear model (VGLM, Yee, 2015) has been used. Exemplarily, Fig. 3.1 shows the monthly mean precipitation for the reference stations Vila Nogueira de Azeitao, located on the Research Site Tagus-River (Portugal).

In the modeling step, the seasonality of the model output and the respective reference data are derived using the approach described above. Figure 3.2 shows an example of the two seasonal cycles for precipitation at the station Vila Nogueira de Azeitao on the Research Site Tagus-River (Portugal).

The adjustment step depends on the variable considered: For Gaussian distributions the difference (case 1) and for positive variables (e.g. Gamma or log-normal) the quotient (case 2) of the estimated seasonal cycle for the reference data set $\bar{r}(t)$ and the simulated

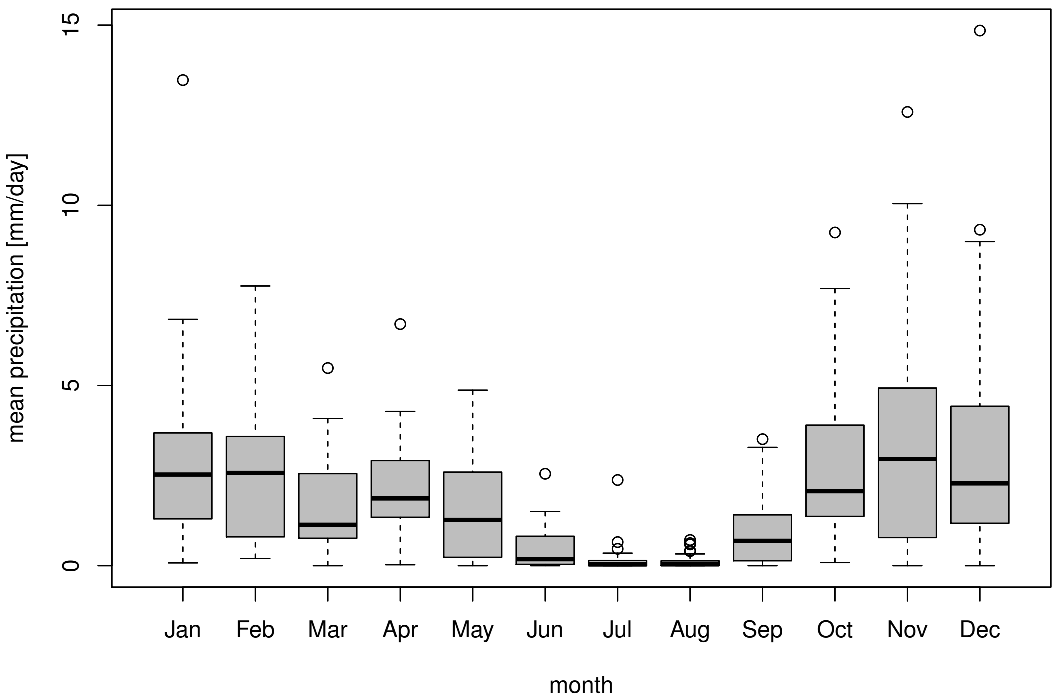

Figure 3.1: Monthly mean of daily precipitation sums at the reference station Vila Nogueira de Azeitao in the research site Tagus-River (Portugal). The data covers the period from 1980-10-01 to 2009-09-30.

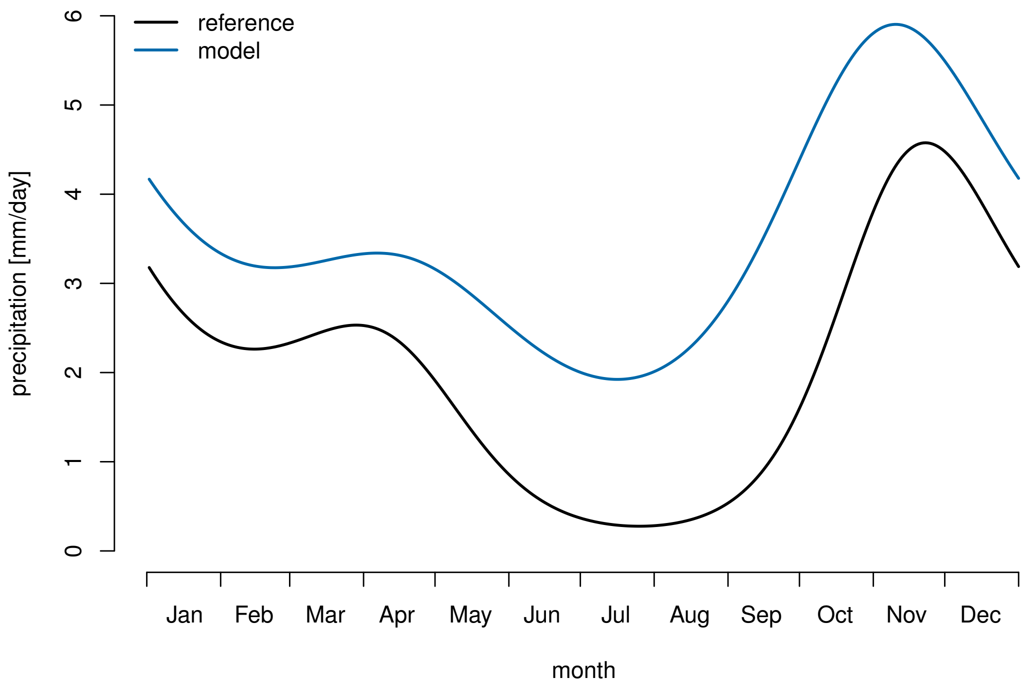

Figure 3.2: Seasonality of the climate simulation output (blue) and the reference observation (black) for precipitation at the station Vila Nogueira de Azeitao in the research site Tagus-River (Portugal) in a daily resolution derived by a VGLM trained with monthly mean values.

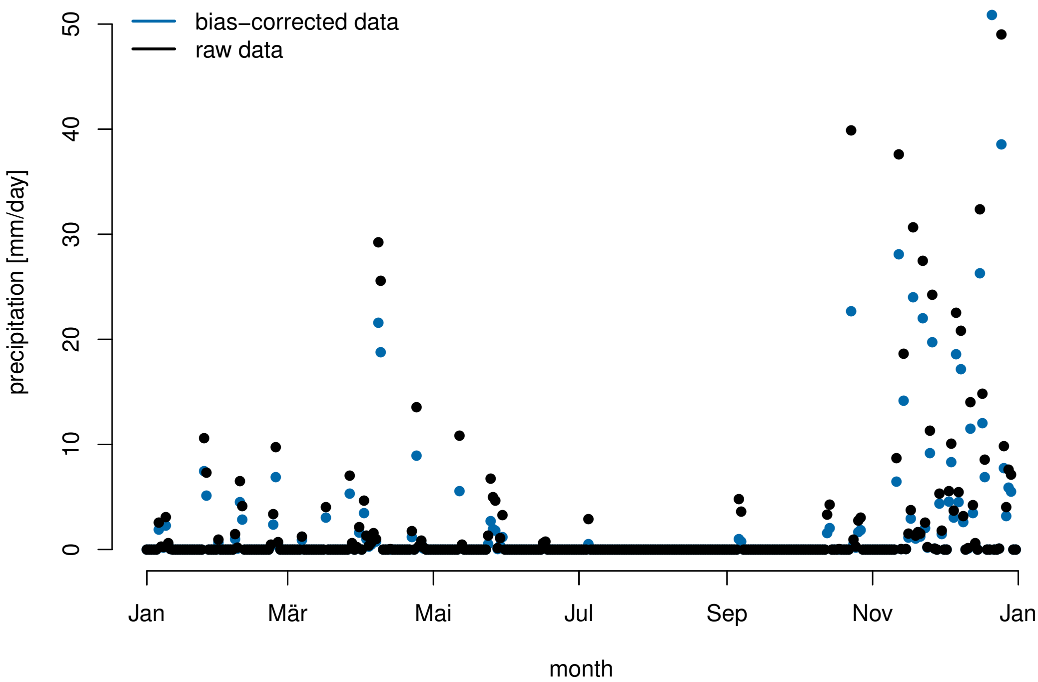

Figure 3.3: Precipitation of the climate simulation at the station Vila Nogueira de Azeitao in the research site Tagus-River (Portugal) for the year 1989 before the bias correction (black) and after (blue).

data set $\bar{x}(t)$ are obtained, see Eq. (3.4).

$$\Delta(t) = \begin{cases} \bar{r}(t) - \bar{x}(t) & \text{Gaussian} \\ \frac{\bar{r}(t)}{\bar{x}(t)} & \text{positive} \end{cases} \tag{3.4}$$

Finally, the adjusted (bias corrected) values are calculated by either adding (case 1) or multiplying (case 2) the thus obtained values to the climate model simulations, see Eq. (3.5).

$$x_{\text{corrected}}(t) = \begin{cases} x(t) + \Delta(t) & \text{Gaussian} \\ x(t) \cdot \Delta(t) & \text{positive} \end{cases} \tag{3.5}$$

Figure 3.3 and Figure 3.4 suggest that the precipitation dataset at the station Vila Nogueira de Azeitao in the research site Tagus-River (Portugal) was corrected in general towards lower values.

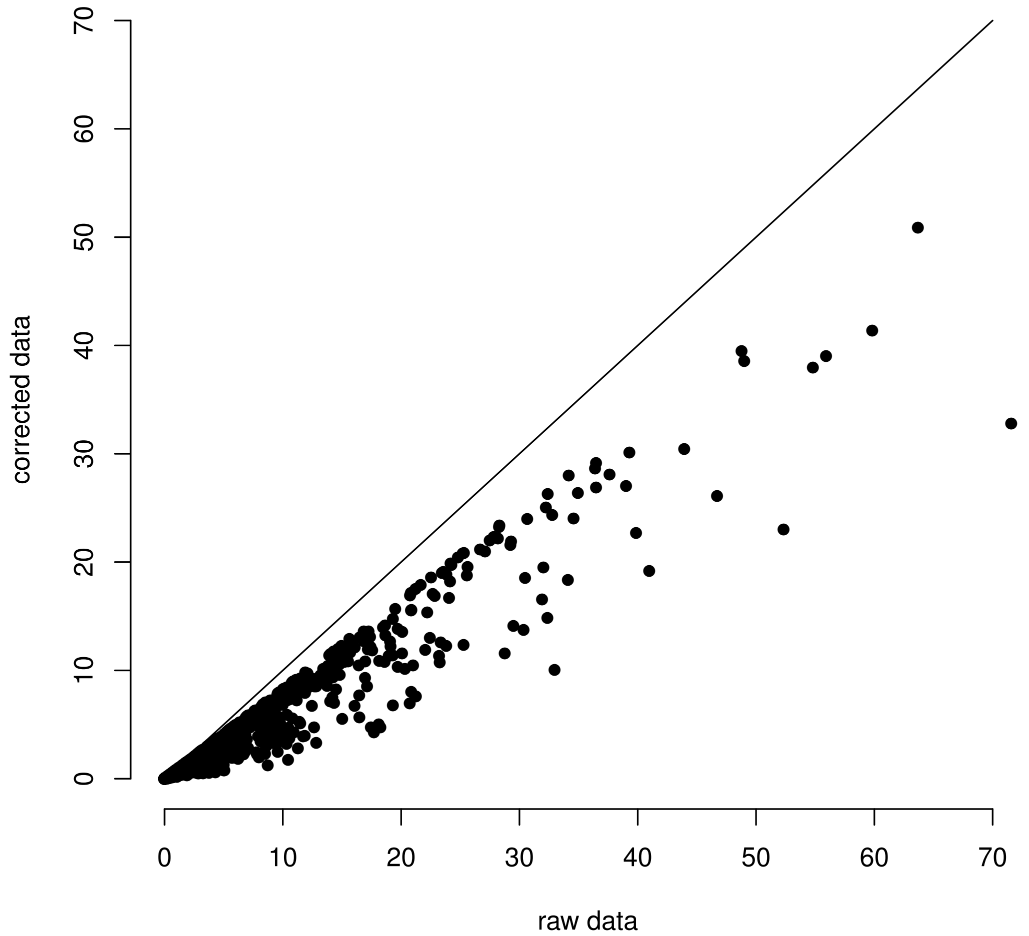

Figure 3.4: Precipitation of the climate simulation at the station Vila Nogueira de Azeitao in the research site Tagus-River (Portugal) before the bias correction (x-axis) and after (y-axis).

3.3 Cumulative Distribution Function Transform method

We present here a method, namely the CDF-Transform, which can be perceived as an extension of the classical quantile-mapping approach (Panofsky and Brier, 1968). This method has been developed by (Michelangeli et al., 2009) and applied in many climate-related studies (e.g. (Colette et al., 2012), (Tisseuil et al., 2012), (Vigaud et al., 2013), (Vrac and Friederichs, 2015)). In the following, we first recap the quantile-mapping method (Sect. 3.3.1) followed by a description of the CDF-Transform method (Sect. 3.3.2) and some concrete applications of the approach at the BINGO Research Sites (Sect. 3.3.3).

3.3.1 Quantile-Mapping Method

Let F_o stand for the CDF (Cumulative Distribution Function) of a climate random variable x_o (temperature, precipitation, wind, etc.) observed at a given weather station during the historical time period, and F_m for the CDF of the same variable x_m from the model, for the same time period. The idea of quantile-mapping is to correct the distribution function of the modelled climate variable to agree with the observed distribution function:

$$F_o(x_o) = F_m(x_m). \tag{3.6}$$

The corrected value x_{bc} can be obtained empirically from (see Figure 3.5),

$$x_{bc} = F_o^{-1}[F_m(x_m)]. \tag{3.7}$$

where F_o^{-1}, defined from [0,1], is the inverse function of F_o.

The quantile-mapping method is only suitable when observations are available for the same time period as for the model output (Vrac and Friederichs, 2015). However, it often happens that one needs to correct model output that covers a time period longer than that of the observations. Another case where the classical quantile-mapping is not suited to bias correction is the correction of model future predictions (where observations are obviously not available). This method does not take into account the information on the distribution of the future modelled dataset. The CDF-Transform method is proposed to overcome this potential issue.

3.3.2 CDF-Transform Method

The CDF-Transform approach (hereafter "CDF-t") can be perceived as an extension of quantiles-mapping, directly dealing with and providing CDFs (Michelangeli et al., 2009).

Let $F_{o,h}$ stand for the CDF of a climate random variable observed at a given weather station during the historical time period (training period) and $F_{m,h}$ for the CDF of the

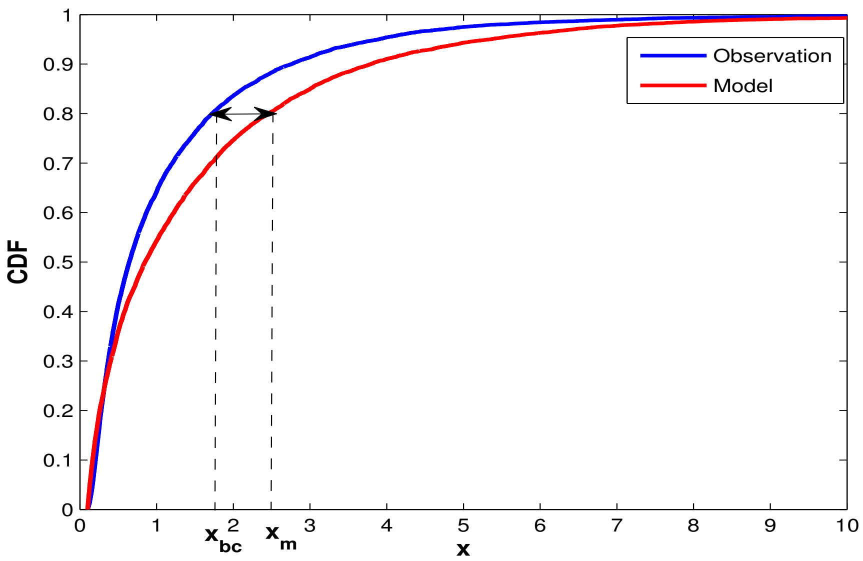

Figure 3.5: Illustration of the Quantile-Mapping Method

same variable from the model output during the same period. $F_{o,f}$ (unknown) and $F_{m,f}$ are the CDFs equivalent to $F_{o,h}$ and $F_{m,h}$ but for a future (or simply different) time period (see Table 3.2). The main goal of the CDF-t is to approximate the CDF of the observations in the future period ($F_{o,f}(x)$) based on historical information and then to apply quantile-mapping between $F_{o,f}(x)$ and $F_{m,f}(x)$.

	Historical period	Future period
Observation	$F_{o,h}(x)$	$F_{o,f}(x)$ (unknown)
Model	$F_{m,h}(x)$	$F_{m,f}(x)$

Table 3.2: Table summarizing CDF notations

Assuming that we know $F_{m,f}$ (which can be modelled via future model output), and that there exists a transformation T: $[0,1] \to [0,1]$ such that

$$T(F_{m,h}(x)) = F_{o,h}(x) \tag{3.8}$$

The CDT-t method is based on the assumption, which is made by most statistical bias correction approaches, that the transformation T is still valid in the future period

$$T(F_{m,f}(x)) = F_{o,f}(x). \tag{3.9}$$

Under this assumption, we can approximate $F_{o,f}$ by applying T to $F_{m,f}$.

The first step is to model T and the simple way to do so is to replace x by $F^{-1}_{m,h}(u)$ in (3.8), where u is any probability in $[0,1]$. We then obtain

$$T(u) = F_{o,h}(F^{-1}_{m,h}(u)), \tag{3.10}$$

corresponding to the simple definition of T. Inserting (3.10) in (3.8) leads to a modelling of $F_{m,f}$,

$$F_{o,f} = F_{o,h}(F^{-1}_{m,h}(F_{m,f}(x))). \tag{3.11}$$

From a technical/algorithmic point of view, the CDF transform approach is defined in three steps following (Michelangeli et al., 2009) :

1. The estimates of $F_{o,h}$, $F^{-1}_{m,h}$ and $F_{m,f}$, respectively $\hat{F}_{o,h}$, $\hat{F}^{-1}_{m,h}$ and $\hat{F}_{m,f}$, are empirically modelled respectively from the historical observations and the historical and future model output data.

2. Then, by combining them according to equation (3.11), we dispose of $\hat{F}_{0,f}$, an estimation of $F_{0,f}$. Note that it is also possible to use parametric CDFs.

3. Once $\hat{F}_{m,f}$ and $\hat{F}_{o,f}$ are estimated, quantile-mapping is applied as in Section 3.3.1

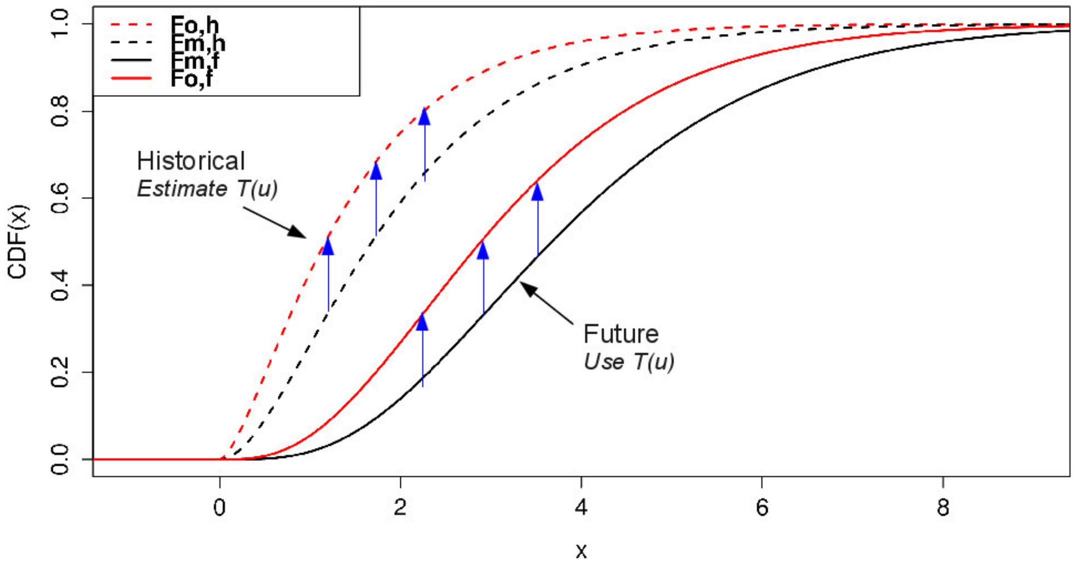

Figure 3.6: Illustration of the future observations CDF ($F_{o,f}$) estimation

Note that, in the equation (3.11), $F_{o,f}$ is only defined for $x \in [m_h, M_h]$, where m_h and M_h are respectively the minimum and the maximum of the model outputs in the historical period. Outside $[m_h, M_h]$, $F_{o,f}$ gives the same constant value. As in (Déqué, 2007) or (Michelangeli et al., 2009) a "constant correction" method is applied whenever x is outside $[m_h; M_h]$ e.g. if the maximum value of x in the range of $[m_h; M_h]$ is corrected by x_0, all x such as $x > M_f$ is corrected by x_0. It is important to mentioned that the portion of data for which the "constant correction" method is applied is very small.

3.3.3 Application of CDF-t at BINGO research sites

WATCH forcing data ERA-Interim (WFDEI) are used as reference data for all research sites. The available variables in WFDEI for bias correction are: temperature, total precipitation, surface air pressure, near-surface wind speed, long/shortwave incident radiation and near-surface specific humidity. To take care of the seasonality, the CDF-t is applied separately for each calender month. The calibration period is set to 1980-2013.

Special attention is given to precipitation since it shows a continuous probability distribution for strictly positive values but has a discontinuity at zero. The CDF-t is applied for daily precipitation amounts greater than a fixed threshold. The chosen threshold is 0.1mm for WFDEI and for the model, the threshold is adjusted so that the frequency of wet days is the same as in the WFDEI.

Figures 3.7 and 3.8 show some results obtained for precipitation at the grid point of coordinates (longitude=7.46 and latitude=51.11). In figure 3.7, as expected, a good agreement can be observed between the reference data quantiles and the quantiles of the model after bias correction. Although CDF-t is based on cumulative distribution function, it is able to correct also the mean and the variance. Indeed, in Figure 3.8 we compare the monthly mean (left) and standard deviation (right) of daily precipitation and there is a good agreement between the reference and the model after CDF-t bias correction.

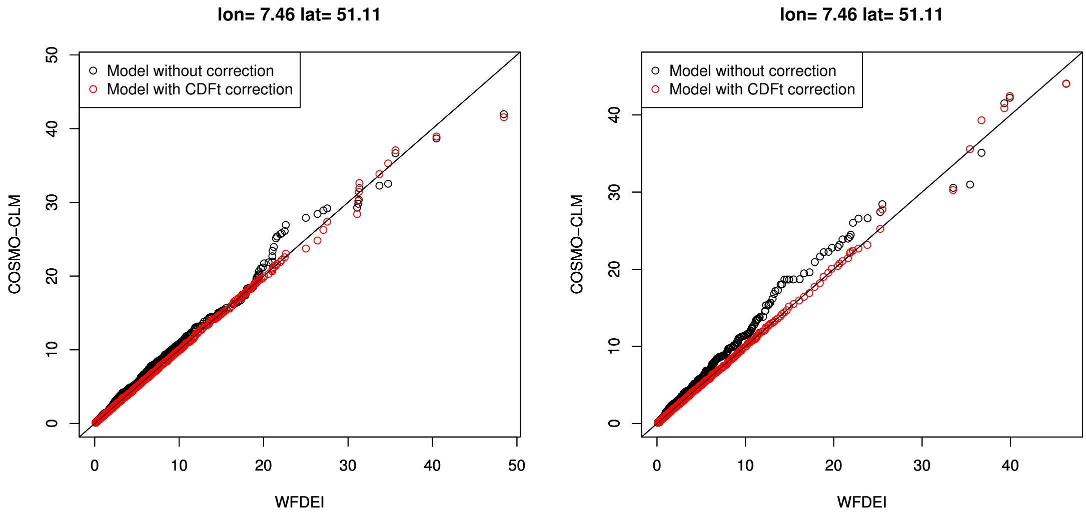

Figure 3.7: Quantile-quantile plot of WFDEI daily precipitation (mm/day) and COSMO-CLM daily precipitation (mm/day) without and with CDF-t bias correction at location defined by longitude=7.46 and latitude=51.11. The left plot corresponds to January and the right plot to the August

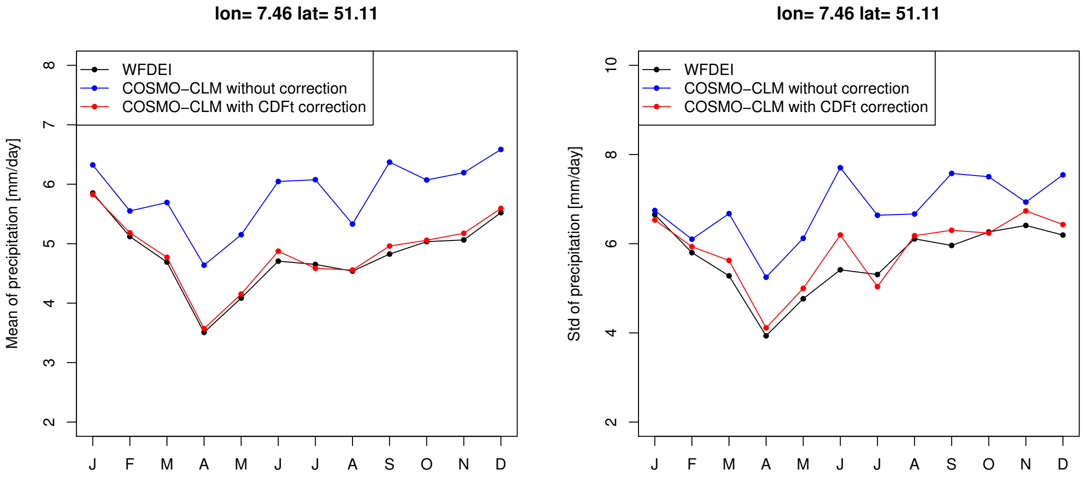

Figure 3.8: Monthly mean (left) and standard deviation (right) of daily precipitation at location defined by longitude=7.46 and latitude=51.11. The mean and the standard deviation are calculated only for wet days

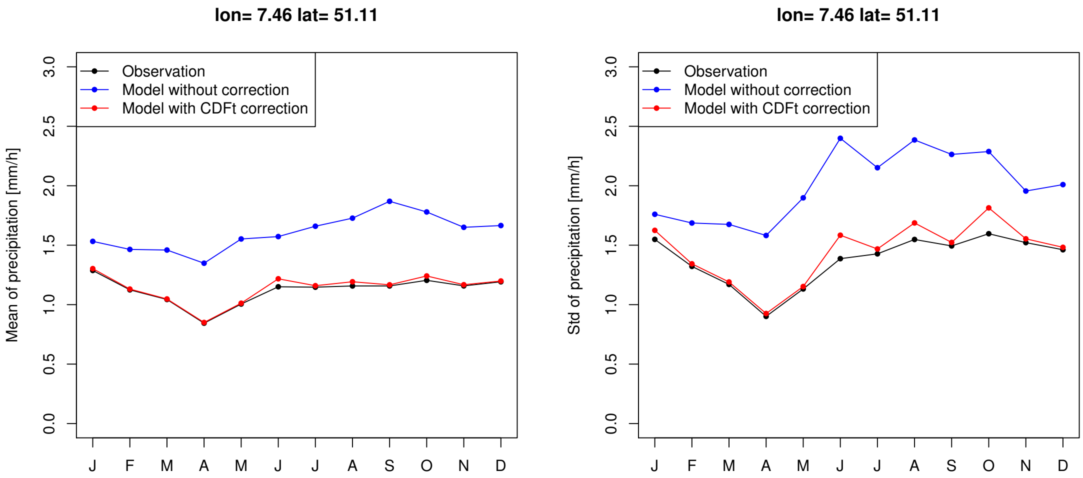

Figure 3.9: Monthly mean (left) and standard deviation (right) of 3 hourly precipitation at location defined by longitude=7.46 and latitude=51.11.

Chapter 4

DECO – A BINGO plug-in for FreVa

4.1 FreVa – Freie Universität Berlin evaluation system

Freva is the *Freie Universität Berlin Evaluation Framework for Earth System Science.* The fully operational hybrid features a HPC shell access and an user friendly web-interface. It employs one common system with a variety of verification tools and validation data from different projects in- and outside of the FUB. The evaluation system is located at the FUB, the DWD and German Climate Computing Centre (DKRZ), especially this has direct access to the bulk of its ESGF node including millions of climate model data sets, e.g. from CMIP5 and CORDEX. The database is organized by the international CMOR standard using the meta information of the self-describing model, reanalysis and observational data sets. Apache Solr is used for indexing the different data projects into one common search environment. This implemented meta data system with its advanced but easy to handle search tool supports users, developers and their tools to retrieve the required information. A generic application programming interface (API) allows scientific developers to connect their analysis tools with the evaluation system independently of the programming language used. Users of the evaluation techniques benefit from the common interface of the evaluation system without any need to understand the different scripting languages. Facilitating the provision and usage of tools and climate data increases automatically the number of scientists working with the data sets and identify discrepancies. Additionally, the history and configuration sub-system stores every analysis performed with the evaluation system in a MySQL database. Configurations and results of the tools can be shared among scientists via shell or web-system. Therefore, plugged-in tools gain automatically from transparency and reproducibility. Furthermore, when configurations match while starting a evaluation tool, the system suggests to use results already produced by other users–saving CPU time, I/O and disk space. website: freva.met.fu-berlin.de visitor-login: click on "Guest". A detailled description is currently being prepared for Geoscientific Model Development (Kadow et al., in preparation).

4.2 Documentation of DECO

4.2.1 Introduction

The BINGO-DECO FreVa plug-in produces meteorological and climatological input data for hydrological models within the BINGO project. For a list of models see Tab. 4.1 in Sect. 4.2.2. The plug-in is part of an general evaluation system (Sect. 4.1); a sketch of this system with the BINGO-DECO plug-in highlighted is given in Fig. 4.1. FreVa is

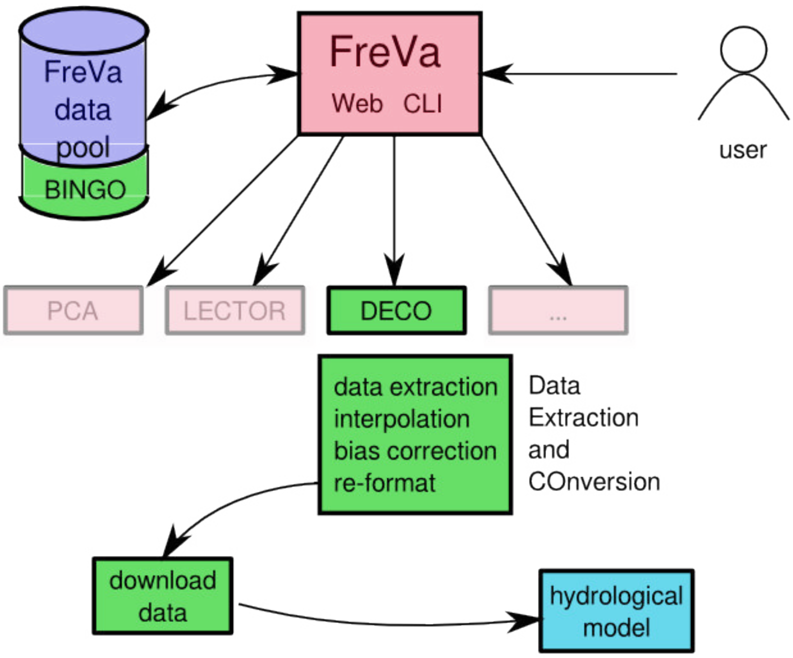

Figure 4.1: Sketch of the FreVa evaluation system with the BINGO-DECO plugin highlighted.

a framework hosting several plug-ins (applications) which have all access to a common data pool. Data in this pool is standardized such that plug-ins can instantly access new data which comes into this pool. The climate data produced for BINGO will be part of this pool. To use the framework, users must register via the FreVa web-page https://freva.met.fu-berlin.de/ and access is granted by the FreVa team. The team has currently a list of BINGO hydrological modelers provided by WP3.

Regardless of the chosen meteorological data set to be processed, the plug-in prepares a specific downloadable standard output according to the selected hydrological model and Research Site respectively (Tab. 4.3). Besides preparing the original meteorological data for hydrological models, the plug-in has an option to bias-correct the data based on different methods beforehand, see Chap. 3 .

This section is structured as follows: Section 4.2.2 describes the preprocessing, Sections 4.2.3 and 4.2.4 give an overview over the input and the output of the BINGO-DECO plug-in, respectively.

4.2.2 Preprocessing

The preprocessing consists basically of a spatial selection of the region/stations, as well as a temporal selection of dates plus a conversion of variables and their associated units. To avoid unnecessary grid remapping, in the first version of the plug-in the spatial selection is applied on the native grid of the chosen meteorological input data. For station data, a nearest neighbor remapping (see `cdo -remapnn` in the CDO User's Guide (Schulzweida, 2015)) is applied to get the nearest grid cell of a given longitude/latitude location. For the region selection, a box will be selected (`cdo -sellonlatbox`) regarding to a defined longitude/latitude rectangle. This definition of rectangle is based on the replies to the hydrological-model-based questionnaires, sent around by BINGO WP3. Results of these questionnaires are given in Tab. 4.1. Note, that due to the use of native grid, the final selected grid boundary is larger and does not exactly match the defined rectangle. Nevertheless, all grid cell centers of the selected native grid are definitively inside the defined rectangle range by users. With finer grids in later project stages, this effect becomes more and more negligible.

Table 4.1: Naming convention of BINGO Research Sites and hydrological models which are used in the BINGO-DECO plug-in. Additionally, the defined lon/lat rectangle based on the user-questionnaire is shown. Note, in case we did not get feedback related to the region boundaries, we defined a rectangle matching the Research Site.

#	Country	Research Site	Hydrol. Model	lon/lat rectangle
1	NOR	Bergen	ENKI	5.10°E - 5.76°E 60.13°N - 60.64°N
2	NOR	Bergen	HBV	5.10°E - 5.76°E 60.13°N - 60.64°N
3	NED	Veluwe	AZURE	3.039615°E - 7.569199°E 50.580175°N - 53.737922°N
4	GER	Wupper	NASIM	6.914°E - 7.620°E 50.995°N - 51.370°N
5	GER	Wupper	TALSIM	6.914°E - 7.620°E 50.995°N - 51.370°N
6	ESP	Badalona	Infoworks-ICM	1.935000°E - 2.401875°E 41.2400°N - 41.60446°N
7	ESP	Badalona	Mohid	0.06°E - 3.60°E 40.1°N - 43.1°N

Table 4.2 continued

8	POR	Tagus-River	CE-QUAL-W2	NA (station data)
9	POR	Tagus-River	ECO-SELFE	9.6°W - 8.6°W 38.4°N - 39.3°N
10	POR	Tagus-River	Feflow-BALSEQ	NA (station data)
11	POR	Tagus-River	HECHMS	9.30°W - 9.05°W 38.7500°N - 38.9833°N
12	POR	Tagus-River	QUAL-2K	NA (station data)
13	POR	Tagus-River	SCHISM	9.6°W - 8.6°W 38.4°N - 39.3°N
14	CYP	Troodos-Mountains	WRF-Hydro	32.0°W - 35.0°W 34.0°N - 36.0°N

4.2.3 Input parameters

Using the first option of the plug-in lets the user choose among the Research Sites and the associated hydrological models combination (**Research site and hydrological model**). The option **Date range** specifies the time period to be processed. The parameter **Bias correction** optionally leads to a subsequent application of a bias correction scheme (Sect. 3 to the simulated data; choose *None* for no bias correction and output of the native, uncorrected data instead. Bias correction is applied with reference to station data in case these data was made accessible for us and with reference to the WATCH Forcing Data ERA-Interim (WFDEI)(Weedon et al., 2014) otherwise.

The meteorological input data to be chosen for hydrological model can be uniquely addressed by specifying seven parameters. These parameters result from the standardized storage of climate model data according to the CMOR convention also used in the Coupled Model Inter-comparison Project 5 (CMIP5, Taylor et al., 2012). For the deliverable D2.1 these are set by default. For completeness, these options are **Dataproject**, **Dataproduct**, **Institute** and **Model providing input data**. Further, the **Experiment**, **Ensemble member** and **Time frequency of provided input data** must be specified. Many of these parameters will make sense at a later stage of BINGO.

With the following parameters various technical options can be set: A specific output (**Outputdir**) and cache directory (**Cachedir**) can be defined, plus the **Output type**. By selecting the option *Basic* only one compressed zip-file containing the hydrological-model-specific-formatted data files will be produced, while *Additional* will result in an additional NetCDF-file containing all variables, time steps and locations. The subsequent operator **Cacheclear** allows the cache directory to be cleared and the intermediate data used for processing to be deleted or not, in case one wants to guard these data. In **Ntask** the number of parallel-processing task can be selected, to optimize the usage of computer resources. Setting the option **Dryrun** to *True* gives only a list of the meteorological data chosen. Finally, a **caption** can be associated with the results produced by this run. The last parameter option **Unique output id** will prevent you from overwriting an existing result. A full list of all available parameters and their descriptions can be found in Tab. 4.2

Table 4.2: Input parameters for the BINGO-DECO plug-in.

Research site and hydrological model *mandatory*	Combination of Research Site and Hydrological Model for which you want to produce the input data.
Date range *mandatory*	Time period for which you want to process the data (comma-separated format: YYYY-MM-DD,YYYY-MM-DD).
Bias correction *mandatory*	Bias-correction method.
Dataproject providing input data *mandatory*	The data project providing the meteorological data, e.g. bingo.
Dataproduct providing input data *mandatory*	The data product providing the meteorological input data, e.g. eur-11.
Institute providing input data *mandatory*	The institute providing the meteorological input data, e.g. clmcom.
Model providing input data *mandatory*	The model providing the meteorological input data, e.g. ecmwf-eraint-clmcom-cclm4-8-17-v1.
Experiment of provided input data *mandatory*	The experiment name of the provided meteorological input data, e.g. evaluation.
Ensemble member of provided input data *mandatory*	The ensemble member of the provided meteorological input data, e.g. r1i1p1.
Time frequency of provided input data *mandatory*	The time frequency of the provided meteorological input data, e.g. day.
Outputdir	Output directory

Table 4.2 continued

mandatory	*default*: /net/scratch/*user*/evaluation_system/ output/bingo_deco/*timestemp*
Cachedir	Cache directory
mandatory	*default*: /net/scratch/*user*/evaluation_system/ cache/bingo_deco/*timestemp*
Output type	Type of ouput. Basic: One compressed zip-file (stored in **Outputdir**/ZIP) containing the hydrological-model-specified-formatted data files. Additional: Additionally, one NetCDF-file (stored in **Outputdir**/NETCDF) containing all variables, time steps and locations.
mandatory	*default*: Basic
Cacheclear	Option to clear the cache directory.
mandatory	*default*: True
Ntask	Number of parallel-processing tasks.
mandatory	*default*: 6
Dryrun	Set "True" for just showing the list of found files of your chosen meteorological data. Set "False" to process data.
mandatory	*default*: False
Caption	An additional caption to be displayed with the results.
optional	
Unique output id	If true append the freva run id to every output folder.
mandatory	*default*: True

4.2.4 Output

The resulting data files are stored in the chosen **Outputdir** and can be accessed either directly via the web download from the FreVa system or by using secure copy (scp) or secure shell (ssh)[1]. The downloadable compressed zip-file which contains the data files formatted as specified for the hydrological model is stored in the **Outputdir**/ZIP directory. An additionally produced NetCDF files (if **Output type** is set *Additional*) are located in the **Outputdir**/NETCDF directory. Note that, depending on the chosen research site and hydrological model, the files in the two directories could be the same. Some information about the output can be found in Tab. 4.3. The specific output content and format is based on the questionnaire send out by WP3. In case we did not get any feedback a standard output (NetCDF file) will be provided.

[1]Please contact the developers if you like to use scp and ssh.

Table 4.3: File formats, variables with units and their description given for each Research Site and hydrological model combination.

#	Type	Format	Variable [unit]	Variable Description
1	grid curvilinear	nc	pr [mm/day]	Precipitation (solid + liquid)
			tas [°C]	Near Surface Air Temperature
			uas [m/s]	Eastward Near-Surface Wind
			vas [m/s]	Northward Near-Surface Wind
			rsds [W/m^2]	Surface Downwelling SW Radiation
			rlds [W/m2]	Surface Downwelling LW Radiation
			huss [1]	Near-Surface Specific Humidity
2	grid curvilinear	nc	pr [mm/day]	Precipitation (solid/liquid)
			tasmin [°C]	Min. Near Surface Air Temperature
			tasmax [°C]	Max. Near Surface Air Temperature
3	grid curvilinear	nc	pr [mm/day]	Precipitation (solid + liquid)
			tas [°C]	Near Surface Air Temperature
			ps [Pa]	Surface Air Pressure
			hurs [%]	Near Surface Relative Humidity
			sfcWind [m/s]	Near-Surface Wind Speed
			rsds [W/m^2]	Surface Downwelling SW Radiation
			rlds [W/m2]	Surface Downwelling LW Radiation
4	grid curvilinear	nc	pr [mm/day]	Precipitation (solid + liquid)
			tas [K]	Near Surface Air Temperature
			ps [Pa]	Surface Air Pressure
			uas [m/s]	Eastward Near-Surface Wind
			vas [m/s]	Northward Near-Surface Wind
			rsds [W/m^2]	Surface Downwelling SW Radiation
			rlds [W/m2]	Surface Downwelling LW Radiation
			huss [1]	Near-Surface Specific Humidity
5	grid curvilinear	nc	pr [mm/day]	Precipitation (solid + liquid)
			tas [°C]	Near Surface Air Temperature
			uas [m/s]	Eastward Near-Surface Wind
			vas [m/s]	Northward Near-Surface Wind
			ps [Pa]	Surface Air Pressure

Table 4.3 continued

			rsds [W/m^2]	Surface Downwelling SW Radiation
			rlds [W/m2]	Surface Downwelling LW Radiation
			huss [1]	Near-Surface Specific Humidity
6	grid curvilinear	nc4	pr [mm/day]	Precipitation (solid + liquid)
7	grid curvilinear	nc4	uas [m/s]	Eastward Near-Surface Wind
			vas [m/s]	Northward Near-Surface Wind
			tas [°C]	Near Surface Air Temperature
			rsds [W/m^2]	Surface Downwelling SW Radiation
			hurs [%]	Near Surface Relative Humidity
			clt [%]	Total Cloud Fraction
8	station	csv	Prec [mm/day]	Precipitation (solid + liquid)
			Tmin [°C]	Min. Near Surface Air Temperature
			Tmax [°C]	Max. Near Surface Air Temperature
			RHmin [%]	Min. Near Surface Relative Humidity
			RHmax [%]	Max. Near Surface Relative Humidity
			vv [m/s]	Near-Surface Wind Speed
			P [kPa]	Surface Air Pressure
			Rs [W/m2]	Surface Downwelling SW Radiation
9	grid curvilinear	nc	stmp [K]	Near Surface Air Temperature
			uwind [m/s]	Eastward Near-Surface Wind
			vwind [m/s]	Northward Near-Surface Wind
			prmsl [Pa]	Sea Level Pressure
			dswrf [W/m^2]	Surface Downwelling SW Radiation
			dlwrf [W/m2]	Surface Downwelling LW Radiation
			spfh [1]	Near-Surface Specific Humidity
10	station	csv	Prec [mm/day]	Precipitation (solid + liquid)
			Tmin [°C]	Min. Near Surface Air Temperature
			Tmax [°C]	Max. Near Surface Air Temperature
			RHmin [%]	Min. Near Surface Relative Humidity
			RHmax [%]	Max. Near Surface Relative Humidity
			vv [m/s]	Near-Surface Wind Speed

Table 4.3 continued

			P [kPa]	Surface Air Pressure
			Rs [kPa]	Surface Downwelling SW Radiation
11	grid curvilinear	csv	pr [mm/day]	Precipitation (solid + liquid)
			tas [°C]	Near Surface Air Temperature
			psl [Pa]	Sea Level Pressure
			hurs [%]	Near Surface Relative Humidity
			sfcWind [km/h]	Near-Surface Wind Speed
			rsds [W/m^2]	Surface Downwelling SW Radiation
			rlds [W/m2]	Surface Downwelling LW Radiation
12	station	csv	Prec [mm/day]	Precipitation (solid + liquid)
			Tmin [°C]	Min. Near Surface Air Temperature
			Tmax [°C]	Max. Near Surface Air Temperature
			RHmin [%]	Min. Near Surface Relative Humidity
			RHmax [%]	Max. Near Surface Relative Humidity
			vv [m/s]	Near-Surface Wind Speed
			P [kPa]	Surface Air Pressure
			Rs [W/m2]	Surface Downwelling SW Radiation
13	grid curvilinear	nc	uwind [m/s]	Eastward Near-Surface Wind
			vwind [m/s]	Northward Near-Surface Wind
			prmsl [Pa]	Sea Level Pressure
14	grid curvilinear	nc	tas [K]	Near Surface Air Temperature
			uas [m/s]	Eastward Near-Surface Wind
			vas [m/s]	Northward Near-Surface Wind
			ps [Pa]	Surface Air Pressure
			rsds [W/m^2]	Surface Downwelling SW Radiation
			rlds [W/m2]	Surface Downwelling LW Radiation
			huss [1]	Near-Surface Specific Humidity

Chapter 5

Summary

This document describes the development of a web based application for the extraction and conversion of climate model simulations. Climate model data is extracted from a central data pool, post-processed (bias-corrected, regridded) and converted to a set of meteorological driving data directly usable for hydrological models. This application has been realized as a plug-in to the *Freie Universität Berlin Evaluation Framework for Earth System Science* (FreVa) which can hold various types of evaluation workflows having access to an indexed data pool. Workflows can be accessed via a web-platform or the command line interface. The BINGO-DECO plug-in has been specifically designed for the hydrological models at the six BINGO Research Sites but can in principle be extended to other models. A major advantage is the on-demand post-processing and conversion of the driving data from a standardized climate model data source. Instead of storing and keeping the very same meteorological driving information in different formats, the system holds the conversion routines and generates the data on demand. This is storage efficient and ensures reproducible and transparent results.

Bibliography

Nikolina Ban, Juerg Schmidli, and Christoph Schär. Heavy precipitation in a changing climate: Does short-term summer precipitation increase faster? *Geophysical Research Letters*, 42(4):1165–1172, 2015.

Erwan Brisson, Matthias Demuzere, and Nicole PM van Lipzig. Modelling strategies for performing convection-permitting climate simulations. *Meteorologische Zeitschrift*, 2015.

Augustin Colette, Robert Vautard, and M Vrac. Regional climate downscaling with prior statistical correction of the global climate forcing. *Geophysical Research Letters*, 39 (13), 2012.

D. P. Dee et al. The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteor. Soc.*, 137(656):553–597, 2011. ISSN 1477-870X. doi: 10.1002/qj.828.

Michel Déqué. Frequency of precipitation and temperature extremes over france in an anthropogenic scenario: model results and statistical correction according to observed values. *Global and Planetary Change*, 57(1):16–26, 2007.

L. Gudmundsson, L. M. Tallaksen, K. Stahl, and A. K. Fleig. Low-frequency variability of european runoff. *Hydrol. Earth System Sci.*, 15(9):2853–2869, 2011.

U. Heikkilä., A. Sandvik., and A. Sorteberg. Dynamical downscaling of ERA-40 in complex terrain using the wrf regional climate model. *Clim. Dyn.*, 37(7-8):1551–1564, 2011.

C. Kadow, S. Illing, O. Kunst, T. Schartner, I. Kirchner, H.W. Rust, U. Cubasch, and U Ulbrich. Freva - freie univ evaluation framework for scientific infrastructures in earth system modeling. *Geoscientif. Model Develop.*, in preparation.

Elizabeth J Kendon, Nigel M Roberts, Hayley J Fowler, Malcolm J Roberts, Steven C Chan, and Catherine A Senior. Heavier summer downpours with climate change revealed by weather forecast resolution model. *Nature Climate Change*, 4(7):570–576, 2014.

H. Koch, S. Liersch, and F. Hattermann. Integrating water resources management in eco-hydrological modelling. *Integrative Water Resource Management in a Changing World: Lessons Learnt and Innovative Perspectives*, page 13, 2013.

Jochem Marotzke, Wolfgang A Müller, Freja SE Vamborg, Paul Becker, Ulrich Cubasch, Hendrik Feldmann, Frank Kaspar, Christoph Kottmeier, Camille Marini, Iuliia Polkova, et al. Miklip-a national research project on decadal climate prediction. *Bulletin of the American Meteorological Society*, (2016), 2016.

P. McCullagh and J. Nelder. *Generalized Linear Models.* CRC Press, Boca Raton, Fla, 2 edition, 1989.

Edmund P Meredith, Vladimir A Semenov, Douglas Maraun, Wonsun Park, and Alexander V Chernokulsky. Crucial role of black sea warming in amplifying the 2012 krymsk precipitation extreme. *Nature Geoscience*, 2015.

P-A Michelangeli, Matthieu Vrac, and H Loukos. Probabilistic downscaling approaches: Application to wind cumulative distribution functions. *Geophysical Research Letters*, 36(11), 2009.

Wolfgang A Mueller, Johanna Baehr, Helmuth Haak, Johann H Jungclaus, Jürgen Kröger, Daniela Matei, Dirk Notz, Holger Pohlmann, JS Storch, and Jochem Marotzke. Forecast skill of multi-year seasonal means in the decadal prediction system of the max planck institute for meteorology. *Geophysical Research Letters*, 39(22), 2012.

Hans A Panofsky and Glenn Wilson Brier. *Some applications of statistics to meteorology.* University Park : Penn. State University, College of Earth and Mineral Sciences, 1968.

Holger Pohlmann, Wolfgang A Mueller, K Kulkarni, M Kameswarrao, Daniela Matei, FSE Vamborg, C Kadow, S Illing, and Jochem Marotzke. Improved forecast skill in the tropics in the new miklip decadal climate predictions. *Geophysical Research Letters*, 40(21):5798–5802, 2013.

Andreas F Prein, Wolfgang Langhans, Giorgia Fosser, Andrew Ferrone, Nikolina Ban, Klaus Goergen, Michael Keller, Merja Tölle, Oliver Gutjahr, Frauke Feser, et al. A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of geophysics*, 53(2):323–361, 2015.

M. B. Priestley. *Spectral Analysis and Time Series.* Academic Press, London, 1992.

C. Prudhomme, I. Giuntoli, E. L. Robinson, D. Clark B, N. W. Arnell, R. Dankers, B. M. Fekete, W. Franssen, D. Gerten, S. N. Gosling, et al. Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. *Proc. Nat. Acad. Sci.*, 111(9):3262–3267, 2014.

B. Rockel, A. Will, and A. Hense. The regional climate model COSMO-CLM (CCLM). *Meteorol. Z.*, 17(4):347–348, 2008.

H. W. Rust, T. Kruschke, A. Dobler, M. Fischer, and U. Ulbrich. Discontinuous daily temperatures in the watch forcing datasets. *J. Hydrometeor.*, 16(1):465–472, 2015.

U. Schulzweida. *Climate Data Operators: CDO User's Guide*. MPI for Meteorology, October 2015. URL https://code.zmaw.de/projects/cdo/embedded/cdo.pdf. Version 1.7.0.

Doug M Smith, Stephen Cusack, Andrew W Colman, Chris K Folland, Glen R Harris, and James M Murphy. Improved surface temperature prediction for the coming decade from a global climate model. *science*, 317(5839):796–799, 2007.

K. E. Taylor, R. J. Stouffer, and G. A. Meehl. An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, 93(4):485–498, 2012.

Clément Tisseuil, M Vrac, G Grenouillet, AJ Wade, M Gevrey, Thierry Oberdorff, J-B Grodwohl, and S Lek. Strengthening the link between climate, hydrological and species distribution modeling to assess the impacts of climate change on freshwater biodiversity. *Science of the total environment*, 424:193–201, 2012.

C. Torma, F. Giorgi, and E. Coppola. Added value of regional climate modeling over areas characterized by complex terrain—precipitation over the Alps. *J. Geophys. Res.: Atmospheres*, 120(9):3957–3972, 2015. 2014JD022781.

N Vigaud, M Vrac, and Y Caballero. Probabilistic downscaling of gcm scenarios over southern india. *International Journal of Climatology*, 33(5):1248–1263, 2013.

Mathieu Vrac and Petra Friederichs. Multivariate-intervariable, spatial, and temporal-bias correction. *Journal of Climate*, 28(1):218–237, 2015.

G. P. Weedon, G. Balsamo, N. Bellouin, S. Gomes, M. J. Best, and P. Viterbo. The wfdei meteorological forcing data set: Watch forcing data methodology applied to era-interim reanalysis data. *Water Resour. Res.*, 50(9):7505–7514, 2014.

T. W. Yee. *Vector generalized linear and additive models*. Springer, 2015.