CrossMark

Evaluation of the empirical–statistical downscaling method EPISODES

Frank Kreienkamp[1] · Andreas Paxian[2] · Barbara Früh[2] · Philip Lorenz[1] · Christoph Matulla[3]

Abstract

A comprehensive performance assessment of the empirical–statistical downscaling (ESD) technique named EPISODES is presented. Pertaining evaluation analyses consist of multifarious validation experiments as well as various comparisons of EPISODES' projections with those of three RCMs and two ESD methods based on the same GCM scenarios driven by two distinct representative concentration pathways (RCPs). EPISODES combines the downscaling of GCM simulations with a follow-up production of synthetic local time series. EPISODES is a comparably simple, computationally rather inexpensive technique, providing multi-variable and multi-site data that are suitable for being merged in an ensemble of RCM projections. This allows (e.g. for different RCPs) the compilation of large multi member ensembles derived from various GCM simulations via both main downscaling strategies (ESD and RCMs). Evaluation experiments reveal satisfying degrees of compliance between various results generated by EPISODES and observations. The grid cell bias for yearly values, for instance, is mostly less than 0.1 °C for temperature and 10% for precipitation totals. Recorded temperature values and precipitation totals corresponding to their 1st and the 99th percentiles are well represented by EPISODES too. Comparisons of various climate change signals derived by EPISODES and other downscaling approaches, present high levels of agreement as well. Many more findings referring to evaluation experiments and climate change projections are to be found in the paper as well as throughout the "Appendix".

1 Introduction

Climate change across the globe is driven by changing forcings (e.g. solar irradiance, chemical compositions of the atmosphere, volcanic outbreaks) and shaped by versatile processes in and in-between the spheres of the Earth system on various spatial-temporal scales.

Ever since it has been proven that humankind substantially affects the Earth's climate (Lockwood et al. 1991; IPCC 2007, 2013) the necessity of estimating impacts associated with potential future pathways of manhood (described

by so-called Representative Concentration Pathways RCPs, see e.g. Moss et al. 2008) has been recognized.

Global Climate Models (GCMs, see e.g. Edwards 2010; von Storch 2010; Müller 2010; Edwards 2011; Taylor et al. 2012), which simulate the evolution of climate states in dependence on given forcings, constitute the main tool to analyze potential changes in the climate system. Numerical resolutions of GCMs are given by their grid-scales (typically 100–200 km horizontally) on which equations describing atmospheric phenomena are delineated. Physically consistent specifications of processes within the climate system, however, require at least 3 times this grid-size in each direction, which is called 'skillful scale' (see e.g. von Storch et al. 1993; Jóhannesson et al. 1995). Hence, GCMs are capable of picturing climate states on global to continental scale, but not on regional scales. Regional scale information, however, is crucial in order to develop adaptation measures to protect socio-economic structures and ecosystems.

In order to generate consistent information on regional scales for impact studies, policy making and climate consultancy, etc., so-called downscaling strategies have to be

✉ Frank Kreienkamp
frank.kreienkamp@dwd.de

1 Deutscher Wetterdienst, Güterfelder Damm 87-91, 14532 Stahnsdorf, Germany

2 Deutscher Wetterdienst, Frankfurter Str. 135, 63067 Offenbach am Main, Germany

3 Zentralanstalt für Meteorologie und Geodynamik, Hohe Warte 38, 1190 Wien, Austria

applied to GCMs' output (see e.g. von Storch et al. 1993). There are two main branches of downscaling techniques: (1) empirical-statistical downscaling (ESD, Fowler et al. 2007; von Storch et al. 1993; Benestad et al. 2008), which relies on transfer-functions derived from observations on the coarse scale of GCMs and regional-scale records as well as (2) dynamical downscaling (DD) making use of Regional Climate Models (RCMs, Rummukainen 2010; Sánchez et al. 2015; Nikulin et al. 2012; Tang et al. 2016; Ozturk et al. 2017; Kotlarski et al. 2014), which are driven by GCM output at the borders of a limited area (e.g. Europe) and calculate atmospheric processes within this area at a much finer grid. Both techniques ultimately rely on the capability of GCMs to realistically reproduce processes within the climate system.

RCMs are (just as GCMs) based on physical laws and therefore are to be expected to correctly model atmospheric phenomena even under substantially changed conditions of the climate system. DD pictures processes on spatial scales of about one tenth of GCMs' scales. Processes, which RCMs cannot explicitly resolve, have to be parametrized (just as in the case of GCMs). This procedure involves past climate conditions and hence, may not comply with future conditions. Besides, DD requires extensive computational resources in terms of processor performance and storage capabilities.

Over the past 2 decades, much effort has been devoted to the development of RCMs and quite a number of DD projection runs have been derived (Sánchez et al. 2015; Nikulin et al. 2012; Tang et al. 2016; Ozturk et al. 2017; Kotlarski et al. 2014). Up to now more than 100 regional scale climate change projections have been generated within the EURO-CORDEX project, the European branch of the CORDEX initiative, which are based on different GCM-RCM combinations and several RCPs (http://www.euro-cordex.net, see e.g. Jacob et al. 2013; Kotlarski et al. 2014). About half of them exhibit a spatial resolution of 50 km, whilst the rest features a grid distance of 12.5 km (Sánchez et al. 2015; Nikulin et al. 2012; Tang et al. 2016; Ozturk et al. 2017; Kotlarski et al. 2014). This list is still extended by several projects like the German project ReKliEs-De (Hübener et al. 2017). Besides its high computational costs DD shows a lack of performance in complex terrain (e.g. across the Central European Alpine region) in respect of simulating temperature and precipitation distributions over climate periods. This is well known (see e.g. Haslinger et al. 2013; Rummukainen 2010). Thus, so-called statistical 'bias correction/bias adjustment' schemes (e.g. Flato et al. 2013; Nikulin et al. 2015; Maraun et al. 2017b) are applied to adjust DD results to observed temperature and precipitation distributions, for instance. Pertaining corrections are retained and subsequently applied to DD projections concerning the future.

ESD establishes transfer-functions between observations on the GCM scale and on regional scales and, by design, reproduces historical records with almost no bias. Since ESD is based on the observed coherence between the scales, they are not necessarily capable to simulate potential future processes outside the range of so far recorded atmospheric events. ESD techniques perform at comparably low computational costs and are therefore suitable to produce large numbers of climate change projections. A classification of ESD methods and a comprehensive literature overview can be found in Maraun et al. (2010), Gutiérrez et al. (2013a), Maraun et al. (2015) and Benestad et al. (2008).

During the last years a ESD specific CORDEX activity has been established (see http://cordex.org/domains/cordex-esd/). ESD activities are still in early stages albeit recommendations issued e.g. by the IPCC (2007) highlight the importance of joining DD and ESD projections to establish broad climate change ensembles. Hewitson et al. (2013) define this problem via a four dimensional matrix made up by feasible pathways of mankind, GCMs, ESD and DD projections and point out that this matrix is for no region on Earth sufficiently staffed. Therefore, even though DD and ESD products occasionally yield contradictory results, which possibly confuse practitioners, the extended size of associated ensembles is essential to understand various sources of uncertainties. Hewitson et al. (2013) emphasize the importance of comparing ESD results to those generated by DD in any case, i.e. even if they are in concert.

EPISODES is an ESD technique and hence capable of producing multi-scenario and multi-model ensembles at comparatively low computational costs. As such EPISODES addresses the above-mentioned demand and contributes to replenish the four dimensional matrix (see above), which is designed to assess the range of potential climate changes based on multiple GCM runs and different pathways of manhood (RCPs). Within the present study, we use the RCP scenarios RCP4.5 and RCP8.5; a description can be found in Moss et al. (2010) and van Vuuren et al. (2011).

One central goal of EPISODES is to supply different impact research areas with suitable regional scale climate change projections. Thereby EPISODES offers advantages over many other ESDs strategies providing single-site and single-variable projections, which means that meteorological variables are neither physically consistent in space nor amongst themselves (see, e.g., Maraun et al. 2017a; Jones et al. 2009; Wilks 2010). EPISODES projections are consistent in space and are thus actually useable for impact studies depending on geographical meaningful pattern and potentially on sets of variables (i.e. hydrological impact assessments).

EPISODES adopts a two-step approach generating consistent regional scale climate information. The first step is a daily-based, variable-specific downscaling, producing

single-site, single-variable data. These data are utilized in the second step for the selection of appropriate records assigned to single days within the past.

EPISODES is capable of generating physically consistent multi-GCM, multi-site and multi-variable ensembles for different pathways of humankind at relatively small computational costs.

This paper is structured as follows: Sect. 2 describes the data used in this study. Section 3 characterizes ESD methods employed within EPISODES. Section 4 introduces performance metrics used in this study. In Sect. 5 climate change projections produced by EPISODES are presented and compared to DD results. Finally, Sect. 6 summarizes major results and gives an outlook to future applications and further developments.

2 Data

This section presents ground based, regional-scale records as well as large-scale, atmospheric reanalysis and GCM data used in this study.

Regional-scale observations and large-scale reanalysis data are needed to establish statistical transfer functions between the scales, which are part of EPISODES.

GCMs' output, driven by different RCPs are entered into EPISODES in order to derive consistent regional-scale scenarios. These scenarios are to be compared to corresponding climate change projections generated via RCMs as well as by other ESD methods.

2.1 Near-surface observational data

Regional-scale records of daily mean temperature and precipitation totals applied in this study rest upon gridded observational data compiled by the Deutscher Wetterdienst (DWD) called HYRAS (5 km horizontal resolution; Rauthe et al. 2013; Frick et al. 2014), which covers Germany and parts of neighboring countries (see Fig. 1) from 1951 to 2006.

To enable straight forward comparisons to results generated by other downscaling strategies these data are regridded onto the EURO-CORDEX grid (see Sect. 2.4) via a conservative remapping procedure (CDO operator remapcon; Schulzweida 2017).

2.2 Atmospheric reanalysis data

From the range of available reanalysis products (e.g. Saha et al. 2010; Dee et al. 2011; Uppala et al. 2005; Kalnay et al. 1996; ONOGI et al. 2007; Rienecker et al. 2011; Mesinger et al. 2006), NCEP/NCAR reanalysis data (Kalnay et al. 1996) providing information from 1948 onwards,
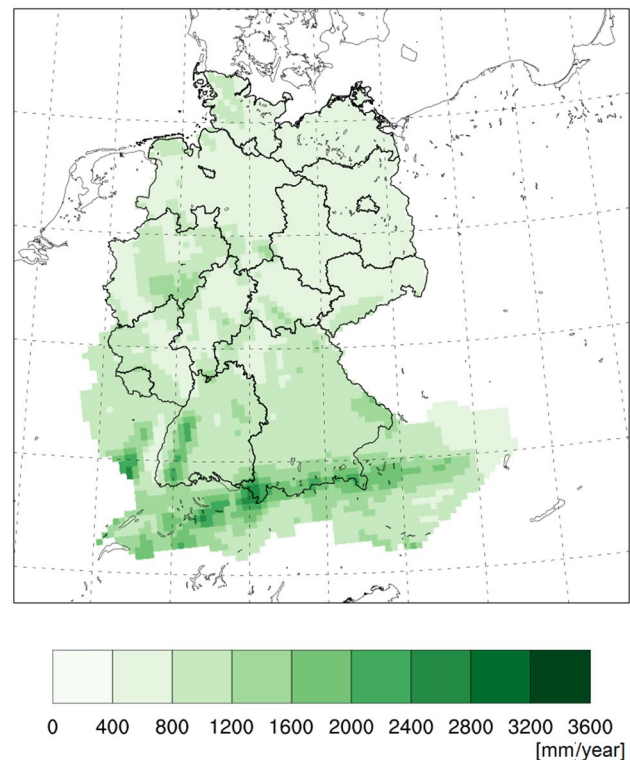
Fig. 1 Regional coverage of the used observational grid, here shown based on the 30-year (1971–2000) mean annual precipitation sums

are selected since they share the largest possible period of time with the regional-scale observations available; here the period from 1951 to 2006 is used.

In this study daily means of geopotential height, temperature and relative humidity at 1000, 850, 700 and 500 hPa, given at a 2.5° lon–lat grid, are extracted and entered into EPISODES. This applies to reanalysis- as well as to GCM-data (see below).

2.3 GCM data

GCM climate change scenarios are taken from CMIP5 (Taylor et al. 2012), whereby climate projections runs carried out with the German Earth System Model MPI-ESM-LR from the Max-Planck Institute for Meteorology (Hamburg, Germany) and the Canadian Earth System Model CanESM2 from the Canadian Centre for Climate Modelling and Analysis (Victoria, Canada) are used.

Concerning MPI-ESM-LR this selection comprises one historical run from 1951 to 2005 (Giorgetta et al. 2012a) and successive scenario runs (2006–2100) forced by RCP4.5 (Giorgetta et al. 2012b) and by RCP8.5 (Giorgetta et al. 2012c).

Considering CanESM2 one historical run (Canadian Centre 2015a) (1951–2005) and one successive RCP8.5

scenario run (Canadian Centre 2015b) (2006–2100) are used. The original model resolution of both models is T63 (∼ 1.9°).

RCP8.5 + CanESM2 is chosen since it refers to a strong radiative forcing scenario that is entered into a GCM, which is known to show high transient climate sensitivity (*TCR* = 2.4 °C; see table 9.5 in IPCC 2013). Hence, pertaining results are to reveal EPISODES' ability to cascade down substantial changes in atmospheric temperatures to regional scales.

For all runs (historical and scenario, all *r1i1p1*) atmospheric (variables and pressure levels match the data extracted from the NCEP/NCAR archives as stated in Sect. 2.2) and near-surface data (*temperature* and *precipitation*) are downloaded from the ESGF (Cinquini et al. 2014).

2.4 RCM data to be compared to EPISODES results

Within the EURO-CORDEX project MPI-ESM-LR results were downscaled by several RCMs, e.g. RCA4 (Strandberg et al. 2014) by the Swedish Meteorological and Hydrological Institute, REMO2009 (http://www.remo-rcm.de) by the Climate Service Center Germany and COSMO4.8-CLM17 (COSMO-CLM) by the CLM-Community (http://www.clm-community.eu/). All groups downscaled one historical, one RCP4.5, and one RCP8.5 run (realisation *r1i1p1* for all) to a grid with a resolution of 0.11° ≈ 12.5 km covering Europe (EURO-CORDEX grid); the observational data of this study (see Sect. 2.1) is remapped to exactly this grid. We use the results from cascade MPI-ESM-LR/SMHI-RCA4 version **v1a**.

The variables *temperature* and *precipitation* from COSMO-CLM and RCA4 are downloaded from the ESGF for the time period 1951 (1971 for RCA4) to 2100. Corresponding REMO data are taken from the results of the German project ReKliEs-De (Hübener et al. 2017; http://reklies.hlnug.de/), a project funded by the Federal Ministry of Education and Research. They are used for comparison with the climate change characteristics of EPISODES results.

2.5 ESD data to be compared to EPISODES results

Within ReKliEs-De (Hübener et al. 2017), GCM results were downscaled with the two ESD methods STARS3 by the Potsdam Institute for Climate Impact Research and WETTREG (Kreienkamp et al. 2013) by Climate and Environment Consulting Potsdam GmbH. Both groups downscaled historical and RCP8.5 results from the GCMs MPI-ESM-LR and CanESM2. The data are available from ReKliEs-De results for both methods on the EURO-CORDEX-grid (0.11° ≈ 12.5 *km*) for the time period 1951–2100.

3 The EPISODES method

Aside from pre- and post-processing modules EPISODES consists of two major parts:

1. A day-by-day downscaling technique preserving the phase relationship to large-scale processes prescribed by GCMs or reanalysis products;
2. A module that generates synthetic, regional-scale time series of the considered variables (e.g. daily temperature averages, precipitation totals).

The first part of EPISODES provides day-by-day meteorological information on regional scales (see Sect. 3.2) which are entered into the second part producing synthetic time series via the application of a weather generator (see Sect. 3.3). This strategy shall allow EPISODES' output to pursue climate evolutions imprinted by RCP driven GCMs. EPISODES' target mesh coincides with the EURO-CORDEX grid.

3.1 Preprocessing

Before the above-described data can be entered into EPISODES they have to be suitably adjusted. This is done in several steps combined in a preprocessing-module. Involved steps are characterized within a flow diagram shown in Fig. 2.

3.1.1 Regridding

Atmospheric (i.e. reanalysis and GCM) data are transferred via a cubic-spline interpolation scheme onto an equidistant grid of a mesh with of 100 km (see Fig. 3a). This grid is hereafter named 'AtmosGrid'.
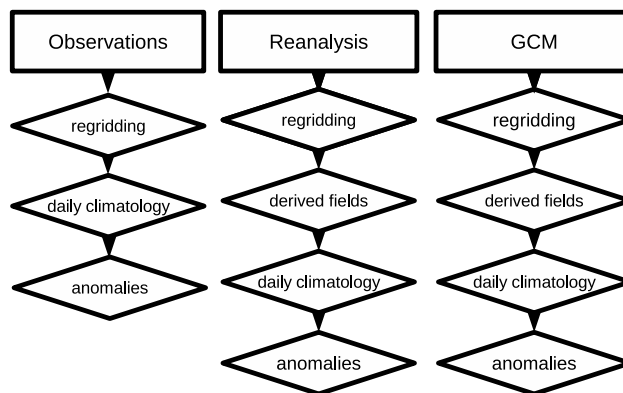

Fig. 2 Flow chart for pre-processing

(a) (AtmosGrid)
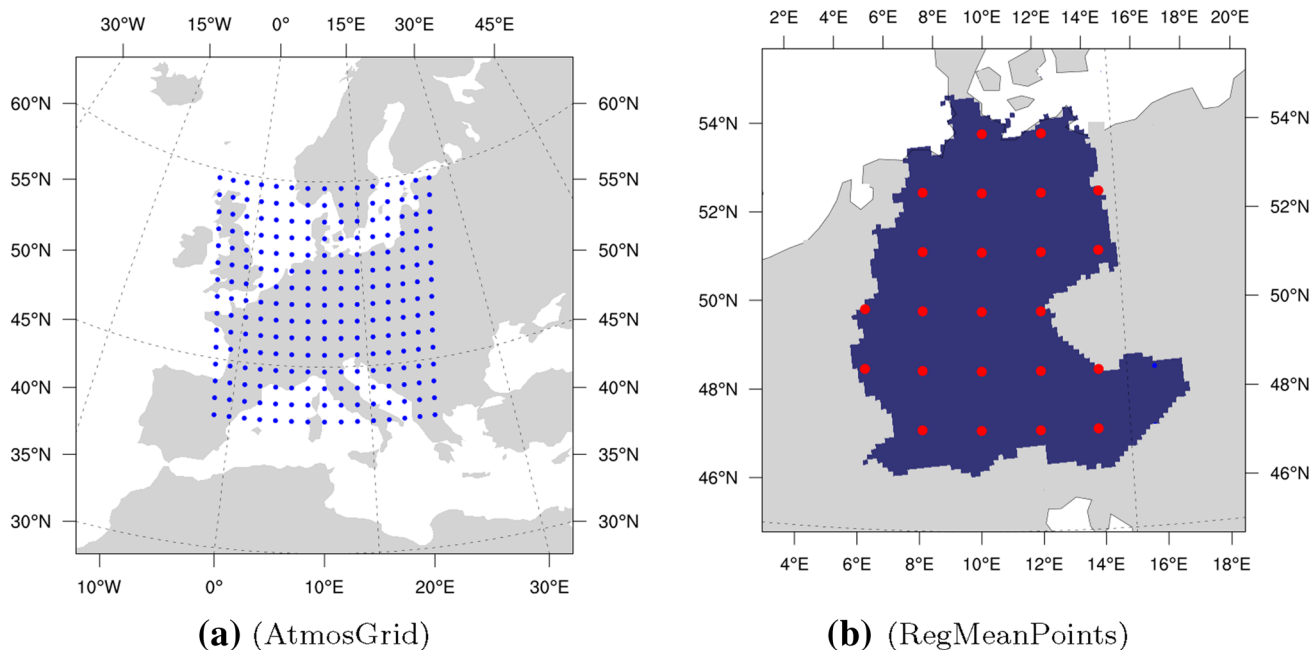

(b) (RegMeanPoints)

Fig. 3 **a** Equidistant 100 km grid (AtmosGrid). All atmospheric (reanalysis and GCM data) are interpolated towards this grid. **b** Red dots: position of regional mean points (RegMeanPoints), which are a subset of AtmosGrid. These points are used to calculate the regional mean values

The grid-points within Germany and in its vicinity are referred to as 'RegMeanPoints' (indicated as red dots in Fig. 3b).

For each RegMeanPoint pertaining regional mean values are given by arithmetic averages over all observational EURO-CORDEX grid cells located within its 100 km × 100 km area. No height correction is applied.

In this manuscript the term 'grid cell' is used for units of gridded (regional-scale) surface variables and the term 'grid point' for spatially interpolated atmospheric variables (large-scale).

3.1.2 Derived fields

As is customary in physics and numerous natural sciences, quantities like temperature at 850 hPa, for instance, which may be extracted from reanalysis and GCM data (see Sects. 2.2 and 2.3), are referred to as fields. This applies to quantities derived from large-scale atmospheric datasets (e.g. geopotential) too, as indicated in the title above. Two such quantities are defined below. First, the horizontal, South-North difference δ of the geopotential:

$$\delta_{i,j} = \Phi_{i,j+1} - \Phi_{i,j-1}, \tag{1}$$

Φ indicates some geopotential point value on AtmosGrid; i and j are indices running from West to East and from South to North, respectively.

The second quantity refers to the geostrophic vorticity (ζ) and is determined by:

$$\zeta_{i,j} = -4 \cdot \Phi_{i,j} + \Phi_{i-1,j} + \Phi_{i,j-1} + \Phi_{i+1,j} + \Phi_{i,j+1} \tag{2}$$

Since these calculation procedures do not depend on height they apply equally to all pressure levels considered here (1000, 850, 700 and 500 hPa).

3.1.3 Daily climatology and anomaly data

Within several modules EPISODES processes anomalies, which are defined as daily deviations from appendant climatological values. Climatological values (reference period 1971–2000) for each day from January 1st until the end of December values are derived according to Eq. 3. The climatological value of Julian day 366 (leap years) is calculated as the mean of Julian day 365 and Julian day 1.

$$\bar{x}_d = \frac{1}{11} \sum_{m=d-5}^{d+5} \left(\frac{1}{30} \sum_{n=1971}^{2000} x_{m,n} \right), \tag{3}$$

d denotes the Julian day (1–366), n the year (1971–2000); m runs through 11-day periods having target day d in its center; $x_{m,n}$ indicates the value at day m in year n.

Anomalies $x'_{n,d}$ are given by the deviations of single day values from their associated daily climatology $\bar{x}_d$ (see Eq. 3):

$$x'_{n,d} = x_{n,d} - \bar{x}_d, \tag{4}$$

with d as the Julian day (1–366), n as year (all possible years), and $x_{n,d}$ as the daily value at day d and year n.

Daily climatology and anomaly data are calculated for all large-scale, atmospheric data and for derived quantities (see Sect. 3.1.2). Concerning regional-scale observations (see Sect. 2.1), daily climatological conditions and corresponding anomalies are calculated for temperature records only. Precipitation totals remain unaffected.

3.2 Regional day-by-day downscaling

The regional day-by-day downscaling part is done by combining a selection of analogue days with a follow-up regression (AFREG, abbreviation based on the German wording **A**naloge **F**älle und **Reg**ression).

The process of matching suitable analogues to atmospheric patterns generated by GCMs runs, which are driven by a particular RCP, imprints consistent regional weather developments according to this RCP. AFREG is used to provide a regional guiding value for the generation of a local synthetic time series. With AFREG each day in the GCM projection is downscaled for all variables and RegMean-Points separately. A flow chart of the AFREG procedure is shown in Fig. 4.

3.2.1 Analogue days (AF)

The main idea of the Analogue Method is to estimate regional scale weather conditions from large-scale atmospheric pattern. The selection of analogue days follows a *Perfect Prognosis*
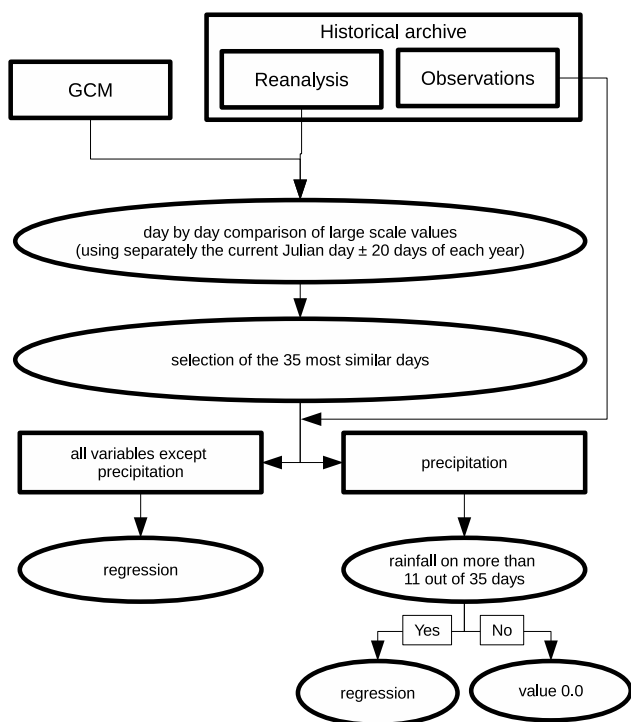
(PP) approach (see Klein et al. 1959; Gutiérrez et al. 2013a, b; Daoud et al. 2016; San-Martín et al. 2017). It is a well-established and robust statistical technique applied in weather forecasting for a long time already (e.g. Lorenz 1963, 1969; Zorita et al. 1995; Pätzhold and Balzer 1995; Wetterhall et al. 2005; Timbal and McAvaney 2001). In the analogue concept, each day's large-scale atmospheric condition in the GCM data set is compared with an historical (here reanalysis-based) archive. As reported in Zorita and von Storch (1999) the analogue concept produces the correct level of variability of the local variable. Daoud et al. (2016) provides a comprehensive overview over the methodology of the analogue concept. The approach used here falls into the category that uses target value specific predictors. Following Timbal et al. (2003) and San-Martín et al. (2017) a regionally focused search for analogue days is done. Using the most similar days in the reanalysis archive an assumption of the local weather conditions based on the observed values is performed (see Eq. 6 and the following description).

In EPISODES, the assignment of analogue days is based on two atmospheric fields X_1 and X_2 (referred to as 'selector-fields') for each target predictand. For temperature geopotential height at 500 hPa and its horizontal difference in N–S direction are applied. For precipitation vorticity at 850 hPa and the horizontal difference of geopotential height at 850 hPa in N–S direction are used (see Table 1). As shown in Raynaud et al. (2017) a predictand-specific method outperforms a multi-predictand method.

Since the analysis includes variables not defined in the same physical units, a normalization is necessary to maintain comparability and allow the linkage of both distances into a total distance (see Eq. 7). Therefore, the selector fields (X_1 and X_2) are a priori normalized by:

$$o'_{i,j,t} = \frac{o_{i,j,t} - \bar{o}_t}{max(o_{i,j,t}) - min(o_{i,j,t})}, \tag{5}$$

Fig. 4 Flow chart of the AFREG procedure

Table 1 Selector-fields (X_1, X_2) and predictor-fields used for analogue day search and regression analyses referring to each target day, respectively

Target predictand	Description	Source
Temperature		
X_1	Geopotential height 500 hPa	GCM
X_2	Horizontal diff. N–S 500 hPa	Equation 1
Predictor	Temperature 1000 hPa	GCM
Precipitation		
X_1	Vorticity 850 hPa	Equation 2
X_2	Horizontal diff. N–S 700 hPa	Equation 1
Predictor	Relative humidity 850 hPa	GCM

The rightmost column indicates associated sources (i.e. whether GCM output or derived fields are used)

with i and j indicating the two dimensions of the AtmosGrid, and t denoting time. Here, $\bar{o}_t$ is the spatial average over all grid point values $o_{i,j,t}$ at time t, and $max(o_{i,j,t}) - min(o_{i,j,t})$ the corresponding spatial spread.

The search algorithm for the analogue days uses the values of the surrounding grid points of the particular RegMeanPoint. In Fig. 5 an example of the used surrounding grid points ($P_1 \ldots P_{12}$) for one specific RegMeanPoint (P_0) is given.

For the assignment of analogue days, the differences D_{X_1} and D_{X_2} (for both selector fields, see Table 1) between the current day GCM field values and each day of the historical archive is calculated day by day using Eq. 6.

$$D_{X_i} = \sum_{n=0}^{12} (w_n \cdot |X_i(P_n, \text{GCM}) - X_i(P_n, \text{Reanalysis})|), \quad i = 1, 2,$$
(6)

with $X_i(P_n, \text{GCM})$ and $X_i(P_n, \text{Reanalysis})$ denoting the normalized (Eq. 5) grid values of the two selector fields (see Table 1) from the AtmosGrid of the GCM and the Reanalysis, respectively.

The weight (w_n) for grid points P_5, P_7, P_{10}, and P_{12} (green dots in Fig. 5) is one. For all other grid points the weight is three. The total distance D_{total} results from adding the distances of both selector fields with Eq. 7:

$$D_{total} = D_{X_1} + D_{X_2},$$
(7)

Eqs. 6 and 7 are computed for each day and RegMeanPoint separately.
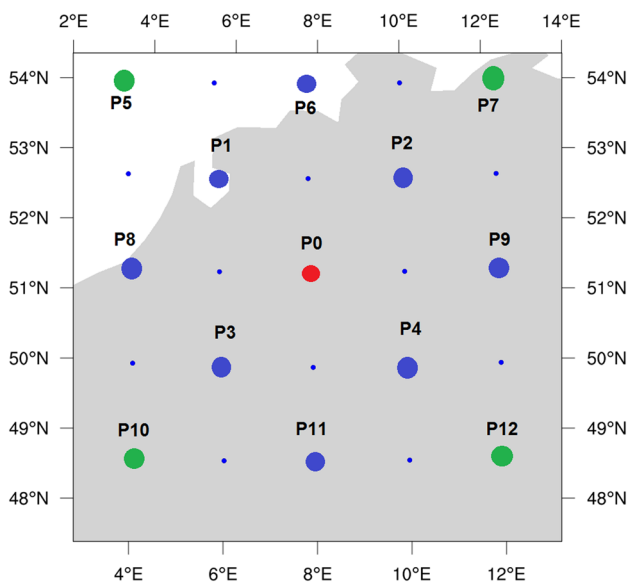
Fig. 5 Example for the grid points used for the search of analogue days (AfregGrid). The red dot (P_0) indicates the regional mean grid point

All days of all years of the historical archive with a Julian day close to the current Julian day ($\pm$ 20 Julian days) are analyzed separately to find analogue days. The distances D_{total} are sorted afterwards. The 35 days with the smallest values of D_{total} are used for the second step, i.e. the regression. The limitations of the use of more than one similar day have been extensively discussed in literature (see Young 1994; Yates et al. 2003; Beersma and Buishand 2003).

The historical archive contains all days from 1971-01-01 to 2006-12-31. To avoid problems with the bias between the reanalysis and the GCM data EPISODES uses daily anomalies (see Sect. 3.1.3). Biases in other moments of the distribution are not considered.

3.2.2 Regression (REG)

Based on the selected 35 days a regression analysis is performed between the regional mean value (RegMeanPoint, predictand) of the observation and one large-scale element of the derived reanalysis fields based on the AtmosGrid (predictor, see Table 1). Using the 35 P_0 values from the Atmos-Grid field and the 35 RegMeanPoint values in the historical archive the parameter of a linear regression are calculated. For precipitation the regression is done differently. In a first step, the number of days with precipitation (≥ 0.5 mm) inside the selected group of 35 days is determined. If the number of wet days is less than 35% (< 12 days), the predicted precipitation amount is set to zero. In a second step, only the wet days are used for regression. The precipitation amount is transformed using a fourth root ($\sqrt[4]{precipitation}$) transformation prior to regression (for details see Howell 1960; Woodley (1970); Fu et al. 2010).

The actual regional day value is calculated using the GCM large-scale predictor value and the linear regression coefficients. For each target variable and for each day in the GCM run a value is calculated for all regional mean points (red dots in Fig. 3b).

By now, neither inter-variable consistency nor spatial correlation are considered. AFREG step produces only specific single-site and single-variable results. Therefore AFREG alone would not provide an improvement over already existing ESD methods. Within EPISODES AFREG is used to provide data required in the next step—the production of synthetic local time series (see Sect. 3.3). In that step inter-variable consistency and spatial correlation are established.

3.3 Production of synthetic local time series

To close the gap between provided regional scale information via the above-described downscaling step (AFREG) and user needs for local information, an additional step that produces synthetic time series on the target grid is introduced. The result is a consistent multi-variable and multi-site data

set consisting of daily values. Associated synthetic time series are assembled by addition of three components: daily-based climatological values (C); the guiding mean climate change evolution (G), derived from the underlying GCM; and the short-term variability obtained from records (V):

$$Daily\ value = C + G + V_{obs}. \qquad (8)$$

A detailed description of the addends is given in the next subsections.

3.3.1 Daily climatology (C)

This *daily climatology* (C) is based on observational data. For each observed variable and grid cell a daily climatology is calculated as an average over the reference period 1971–2000. The calculation procedure is described in Sect. 3.1.3. The concept is shown in Fig. 6.

3.3.2 Mean climate change guidance (G)

The *mean climate change guidance* (G) is determined based on a low-pass filtering of the AFREG anomaly results. Each low-pass filter value is calculated with the information of ± 15 Julian days out of 11 years (actual year ± 5 years).

$$G_{d,y} = \frac{1}{31} \sum_{m=d-15}^{d+15} \left(\frac{1}{11} \sum_{n=y-5}^{y+5} x'_{m,n} \right). \qquad (9)$$

With x' denoting the anomaly value of the AFREG results with respect to climatology (1971–2000); m runs through 31-day periods having target day d in its center; n runs through 11-year periods having target year y in its center. The calculation is done for all days from 1956 to 2095. All days before 1956 are set to the constant values of 1956. The equivalent is done for all days after 2095. This climate change guidance is calculated for each RegMeanPoint (see Fig. 3b). The application of low-pass filtering maintains
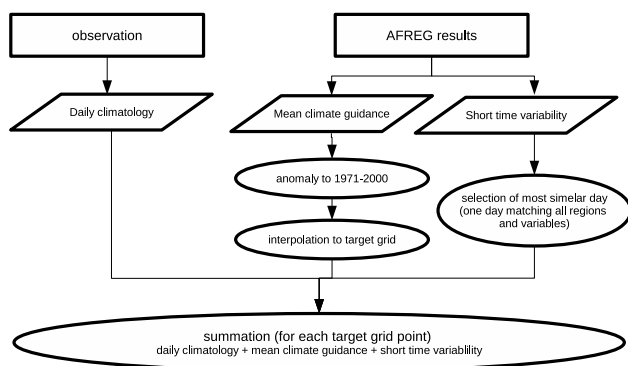
Fig. 6 Flow chart of the generation of synthetic time series

inter-variable consistency (provided by GCMs) for the *mean climate change guidance* (G).

To construct the first component of the synthetic time series the value of the daily climatology is used for each grid cell on the observational grid (see Fig. 1). The bi-linearly interpolated mean climate change guidance value is added to the target grid cell value. To complete the synthetic time series the inclusion of the short-term variability component is needed.

3.3.3 Short-term variability (V)

Impact modellers need spatially consistent data that also include regional details. All this information is already implicitly given by the observed data. To translate this into the synthetic time series the variability component is added by the use of observed days. The short-term anomaly values at the RegMeanPoints are compared with equivalent observational values for each day. An equivalent to the analogue days method is applied; however, this time the selection considers all RegMeanPoints and both target variables simultaneously.

AFREG grid point-specific short-term anomaly values are calculated as the difference between daily AFREG anomaly results and the *mean climate change guidance* (G) defined in the previous subsection (also based on the AFREG results). The *short-term variability* (V) is calculated for each Reg-MeanPoint (see Fig. 3b). In an analogue way, the short-term variability based on the observational data is calculated.

In order to select the most similar day from the observations in terms of the described *short-term variability* (V), a distance measure accounting for both variables—precipitation and temperature - is calculated. For precipitation a classification is used. Each AFREG and observed total precipitation value is transformed to a class number using the following class borders: ≤ 0.0, < 0.5, < 1.0, < 2.0, < 4.0, < 8.0, < 16.0, < 24.0 and < 1000.0 mm. Thereupon for each day of the GCM time series a comparison with observed days in the historical archive is done. The comparison is restricted to the ± 20 Julian days of the current date over for all years. In case the precipitation class of each RegMeanPoint is identical (for up to one RegMeanPoint an offset of one class is allowed), the observed day is added to a list of potential candidates used for the calculation of the total distance (see below).

Unfortunately, after the downscaling step (AFREG) the daily distribution of the precipitation classes still differs from the observed distribution. This is due to the imperfect simulation of large-scale characteristics in the GCM and due to imperfect results of the downscaling method AFREG. To solve this problem the frequency of the precipitation classes needs to be changed after AFREG is performed. The 30-year class frequency is compared between observations and each

historical run (downscaled with AFREG), both covering the period 1971–2000. The percentage difference in each class is used to linearly scale the downscaled data towards observations. The distribution is changed based on a random selection and correction of single days separately per year.

The distance measure for temperature is the sum of the squared distances at all used RegMeanPoints.

$$D_{total} = \sum_{n=1}^{N} (V(P_n)_{AFREG} - V(P_n)_{obs})^2, \tag{10}$$

$V(P_n)$ denotes the variability component of the AFREG result and of the observational value for each of the N RegMeanPoints P_n.

Amongst those days having almost equal precipitation classes the one with the smallest total distance D_{total} is selected.

For each observational grid cell (see Fig. 1) the short-term variability value (V) of this selected observed day from the observational archive is added to the afore constructed synthetic time series ($C + T$ in Eq. 8). The use of one day from the observations at all sites and for all variables ensures inter-variable and spatial consistency. This final synthetic time series now includes the climatologic mean, the mean climate change guidance, the day-to-day variability and the regional specifics. The short-term and long-term persistence follows the GCM specifics.

4 Performance metrics

The performance of EPISODES is evaluated for each grid cell and the complete area of Germany. Maraun et al. (2015) and Kotlarski et al. (2017) present a framework for the evaluation of downscaling approaches. A selected number of indices have been used in this validation. In the following equations n denotes an individual grid cell, N the number of grid cells analyzed (e.g. all grid cells within the area of Germany) and O_n and M_n ($\bar{O}_n$ and $\bar{M}_n$) the daily (temporal mean) observational and climate model data at a particular grid cell n, respectively.

The performance of the climatological mean is evaluated by the bias given as:

$$\text{BIAS}_n = \bar{M}_n - \bar{O}_n. \tag{11}$$

Moderate extremes at the upper and lower end (temperature only) were evaluated by the 1st and 99th percentile P^x. For temperature the mean difference MD as the difference between model and observations is calculated.

$$\text{MD}_{x,n} = P^x(M_n) - P^x(O_n). \tag{12}$$

For precipitation the relative difference RD is determined. To calculate the percentiles only days with non-zero precipitation are used.

$$\text{RD}_{x,n} = 100. \times \frac{P^x(M_n) - P^x(O_n)}{P^x(O_n)}. \tag{13}$$

Furthermore, for precipitation an additional measure is used, i.e. the relative bias in the wet-day frequency. Here, the number of days with at least 1 mm precipitation is counted.

$$\text{WDFREQ}_n = 100 \times \frac{\text{wdfr}(M_n) - \text{wdfr}(O_n)}{\text{wdfr}(O_n)}. \tag{14}$$

The co-variability between observed and simulated spatial patterns of climatological means is assessed using pattern correlation as defined by the Pearson product-moment coefficient of linear correlation:

$$\text{PACO} = \frac{\text{cov}(\bar{M}_n, \bar{O}_n)}{\text{sd}(\bar{M}_n)\,\text{sd}(\bar{O}_n)}, \quad n = 1 \dots N, \tag{15}$$

with cov and sd representing the spatial covariance and standard deviation, respectively.

Based on the two-sample Kolmogorow–Smirnow-Test the d_n^{max} value is used. The d_n^{max} value is the greatest difference between the empirical distribution functions of daily observations $F(O_n)$ and climate model data $F(M_n)$ at a particular grid cell n:

$$d_n^{max} = \sup_x |F(M_n(x)) - F(O_n(x))|. \tag{16}$$

The last performance metric describes the lag auto-correlation (LAG). For temperature the metric is calculated for a lag of 2 and 5 days. For precipitation a lag of 1 and 3 days is evaluated. Here, the relative difference between the model and the observational value is presented.

$$\text{LAG}_n(\tau) = \frac{\sum_t \left[(X_n(t) - \bar{X}_n)) \times (X_n(t + \tau) - \bar{X}_n) \right]}{\sigma^2(X_n)}, \tag{17}$$

with the model or observational data X (i.e. M_n or O_n), day t, lag τ and variance $\sigma^2(X_n)$. For calculation we used the NCAR NCL routine *esacr* (NCAR 2017).

5 Results and discussions

Within this section the evaluation of EPISODES' performance is carried out. Therefore a number of climatological parameters, which are derived from EPISODES' output for the past, are validated against observations from 1971 to 2000. Corresponding analyses are here called 'validation experiments' and their setups are stated in the rows of Table 2. Moreover, projections towards the end of this

century (2071–2100) generated by EPISODES are compared to those produced via other downscaling strategies.

5.1 Modelling climate conditions (1971–2000) from reanalysis data and historical GCM runs

For brevity, analyses, which are based on GCMs refer generally to CanESM2. Exceptions are validation experiments, which are based on both GCMs, CanESM2 and MPI-ESM-LR (see line 2 and 15 of Table 2). Results corresponding to MPI-ESM-LR instead of CanESM2 are to be found in the "Appendix". Whenever EPISODES is driven by GCMs the entire calibration period is used for validation purposes. Apart from two validation experiments, which refer to a single grid cell encasing Potsdam (the so-called 'Potsdam grid cell'; row 3 and 4 of Table 2), all other experiments relate to the entire German territory.

(1) At first, EPISODES' ability to model recorded average temperature and precipitation developments from NCEP/NCAR reanalysis data is evaluated (first line in Table 2). Therefore a temporal cross-validation experiment (see e.g. Matulla et al. 2003)—simulating temperature and precipitation values for each year from 1971 to 2000 by providing EPISODES with all data available except from those associated with the particular year under investigation - is conducted. The outcome is compared to observations (HYRAS, Rauthe et al. 2013; Frick et al. 2014). Findings reveal that EPISODES' reconstructions exhibit yearly biases below 0.1 °C for temperature and of 10% for precipitation.

(2) The following validation analysis conducts a comparison of German wide averaged yearly temperature and precipitation values generated by different downscaling strategies. Pertaining results refer to three RCMs (CCLM, RCA4 and REMO2009) and to three ESDs (EPISODES, STARS3 and WETTREG). All results (except RCA4) are based on both GCMs—MPI-ESM-LR and CanESM2—whilst RCA4 bears on CanESM2 only. EPISODES and STARS3 produces results close to the observed values (1971–2000) and leaves the considered RCMs significantly behind in performance. Interestingly, EPISODES' findings are almost the same for both GCMs, which is why corresponding values overlay and only one symbol signifying EPISODES is visible in Fig. 7. Yearly means of temperature and precipitation show biases to the observations of less than 0.15 °C and 15%, respectively. Larger RCM biases are partly because no explicit calibration against the observed climate has been carried out for the RCMs.

(3) The following validation test pertains to the so-called 'Potsdam grid cell' (see row 3 in Table 2) and contrast EPISODES' daily based temperature and precipitation histograms with observed ones (see Fig. 8). Aside from small underestimations of extremes (in the very tails of the recorded distributions), associated histograms exhibit high levels of consistency. These results are encouraging and demonstrate EPISODES' versatile applicability to problems involving probability distributions—as, for instance, changes in damage events, which rely on particular temperature and/or precipitation occurrence frequencies across almost the entire range of underlying distributions (see e.g. Matulla et al. 2017).

(4) The experiment described in row four of Table 2 arranges the above derived histograms according to

Table 2 List of validation experiments carried out to assess EPISODES' performance in simulating various temperature and precipitation based quantities recorded in the past (1971–2000)

No.	Investigated quantity	Underlying dataset	Presentation of results
01	Mean yearly bias	NCAR reanalysis/EPISODES	Discussed in the text
02	Mean yearly bias	2 GCMs/EPISODES, 3 RCMs	Figure 7
03	Daily histograms	CanESM2/EPISODES	Figure 8
04	Seasonal cycle	CanESM2/EPISODES	Figure 9
05	Mean yearly bias	CanESM2/EPISODES	Figures 10a, 11a
06	Δ 1st percentile (tmp.)	CanESM2/EPISODES	Figure 10b
07	Δ 99th percentile	CanESM2/EPISODES	Figures 10c, 11b
08	Two-sample Kolmogorov–Smirnov test value	CanESM2/EPISODES	Figures 10d, 11d
09	$\Delta 2/\Delta 1$-day lag autocorrelations	CanESM2/EPISODES	$\Delta 2$: Fig. 10e/$\Delta 1$: Fig. 11e
10	$\Delta 5/\Delta 3$-day lag autocorrelations	CanESM2/EPISODES	$\Delta 5$: Fig. 10f/$\Delta 3$: Fig. 11f
11	Δ wet day frequencies	CanESM2/EPISODES	Figure 11c
12	Spatial pattern correlation	2 GCMs/EPISODES, 3 RCMs	Table 3

Analyses are performed for the entire German territory except for experiments described in row 3 and 4 below, which refer to the so-called 'Potsdam grid cell'

Bias to HYRAS (1971-2000)
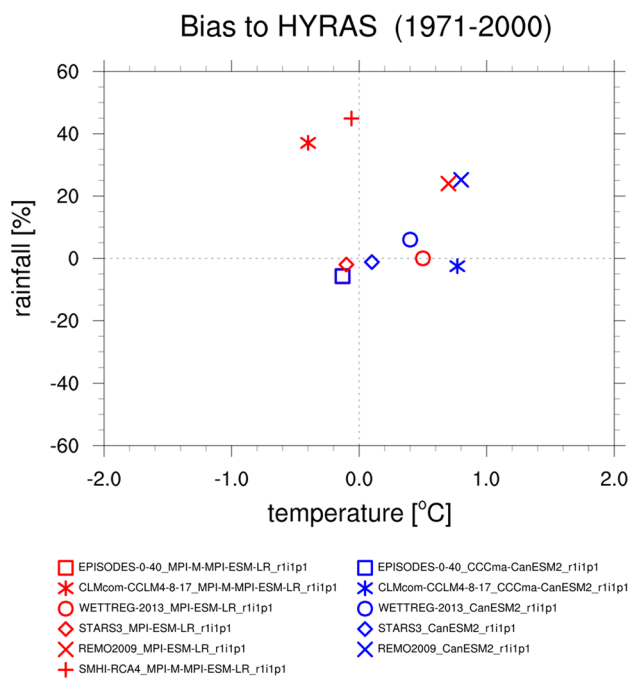
Fig. 7 Annual bias for temperature and precipitation compared to the observational data set (1971–2000). Downscaled results based on MPI-ESM-LR and CanESM2 are colored red and blue, respectively. Findings referring to EPISODES are shown as squares (since EPISODES' results are the same for both GCMs the squares overlay and only one square is visible); Findings of COSMO-CLM, RCA4, REMO2009, STARS3 and WETTREG are displayed as asterisk, upright cross, diagonal cross, diamond and circle respectively. Findings represent biases, which are averaged over all grid cells within the territory of Germany

months within the seasonal cycle. Given the case that the slight underestimation of the extremes (shown above) is due to minor faults in modelling monthly or seasonal distributions, this analysis is suited to identify affected sections of the year. Concerning temperature (see Fig. 9a) observed winter and summer values slightly exceed those modelled by EPISODES. No single month stands out. In fact, the entire seasonal cycle is rather satisfactorily simulated by EPISODES. As such, the minor deviations that can recognized in Fig. 8a towards minimum and maximum temperatures, accumulate over the winter and summer months, respectively. Figure 9b discloses a somewhat different behavior in case of precipitation. It shows that the slight underestimation of small totals may be addressed to EPISODES' simulation of winter totals (more exact: those of November, December and February), whilst the minor under-rating of large sums is caused by the modelling of June's totals, which fall behind recorded ones. Taking into account (1) that these analyses focus on just one grid cell and not on averages over the entire German

territory and (2) the well-known fact that downscaling on regional-scale precipitation is generally more complicated than downscaling on temperature (e.g. Matulla 2005; Haslinger et al. 2013), attained findings are rather encouraging. However, the main finding of the validation experiments, described in row four and five of Table 2, is EPISODES' capability to closely reproduce observed daily temperature and precipitation probability distributions as well as their seasonal cycles. This feature qualifies EPISODES (1) to address a broad range of problems dealing with probabilities and (2) to realistically picture the seasonal cycle of temperature and (to a somewhat less high degree) that of precipitation totals. The following validation experiments (from row 5 in Table 2 onwards), focus on Germany as a whole and on temperature as well as on precipitation - with the exception of row 6 and row 11 of Table 2, which refer solely to temperature and precipitation, respectively. As hitherto, findings displayed throughout the text are based on a historical run (1971–2000) conducted by CanESM2. Results corresponding to MPI-ESM-LR can be found in the "Appendix" if not stated otherwise. Findings associated with temperature are shown in Fig. 10 while those corresponding to precipitation are depicted in Fig. 11. In the following each validation experiment needs to be introduced and findings for temperature as well as for precipitation have to be discussed. For the sake of clarity each validation experiment, listed in Table 2 from row 5 onwards, is introduced and only findings corresponding to temperature are discussed. Once this has been achieved, all validation experiments are known and hence associated results for precipitation can be briefly presented afterwards.

(5)   This experiment evaluates the bias between yearly German temperature means as calculated from observations as well as from synthetic time series, which are generated via EPISODES from a CanESM2 representation of the historical period (1971–2000). Findings are displayed in Fig. 10a and yield a yearly mean bias of less than 0.15 °C.

(6) and (7)   Row 6 and 7 of Table 2 deal with EPISODES' ability to simulate extremes, whereby emphasis is placed on the modelling of the 1st and the 99th percentiles of German's temperature distribution. Figure 10b, reveals overestimations of the 1st percentile across the whole of Germany, which in total result in an area bias of about + 2.0 °C. The 99th percentile, on the other hand, is matched better and instead of overestimations, underestimations are to be detected. However, differences are small and hence an

(a) temperature
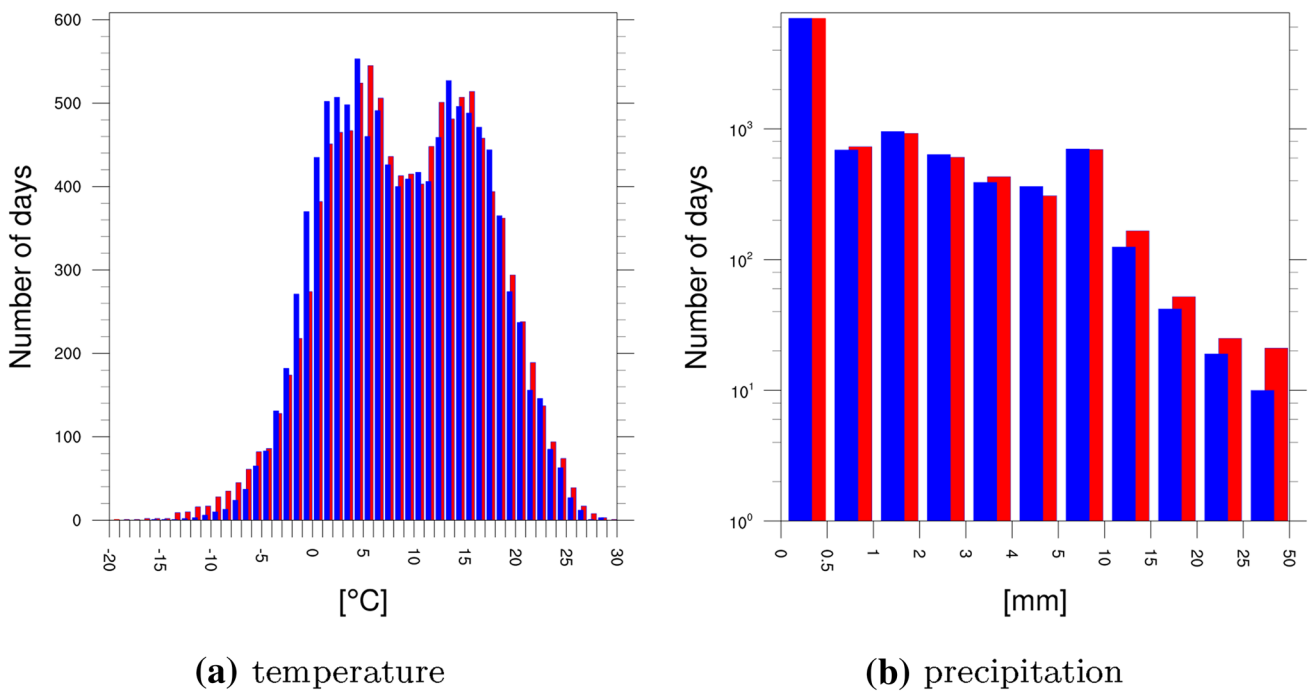
(b) precipitation

Fig. 8 Histogram of daily temperature and precipitation values comparing observational (red) with synthetic time series (blue) for one grid cell near Potsdam, Germany. The synthetic time series is based on downscaling the CanESM2 historical run 1 covering the period 1971–2000

(a) temperature
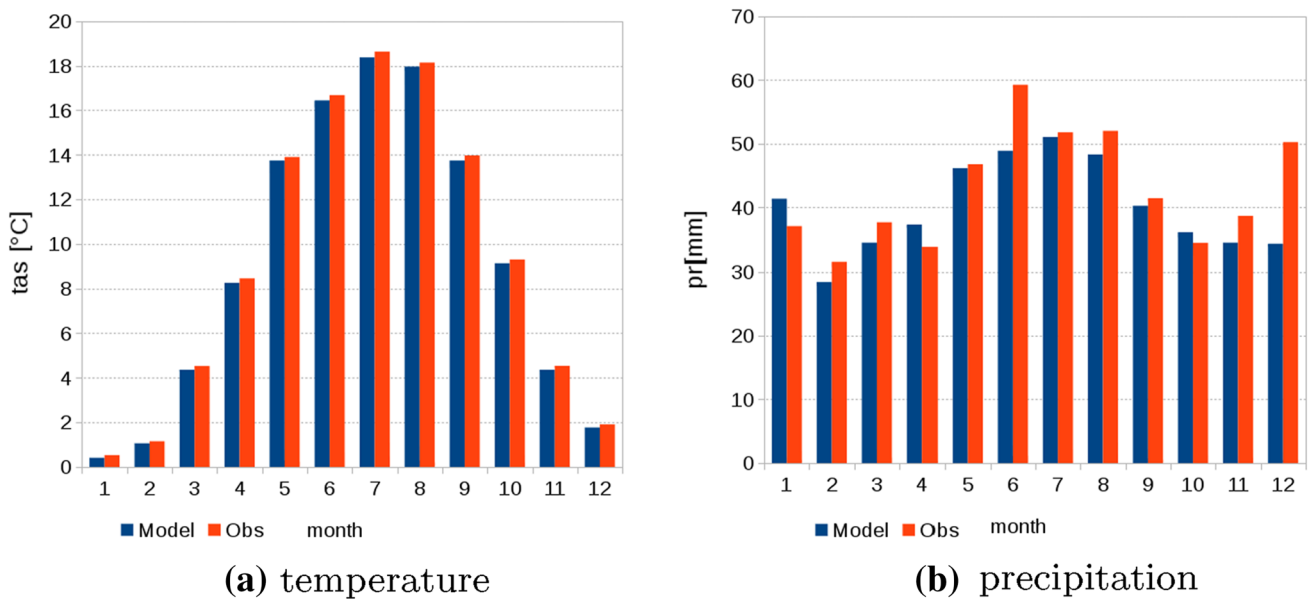
(b) precipitation

Fig. 9 Monthly temperature means and precipitation sums comparing observational (red) with synthetic time series (blue) for one grid cell near Potsdam, Germany. The synthetic time series is based on downscaling the CanESM2 historical run 1 covering the period 1971–2000

area average yields a bias of approximately $-0.4\,°C$ (see Fig. 10c).

(8) Line 8 of Table 2 refers to a two-sample Kolmogorov–Smirnov-test through which potential deviations between observations and simulations can be assessed. Based on the test value d^{max} and an $\alpha = 0.05$ the two-sample Kolmogorov–Smirnov-test identifies EPISODES' results as being indistinguishable from the observations (see Fig. 10d).

Fig. 10 Temperature performance metrics (see Sect. 4) for the EPISODES run based on the GCM run CanESM2 run 1
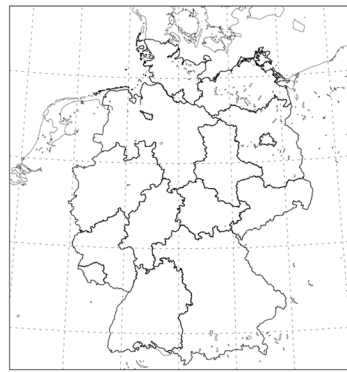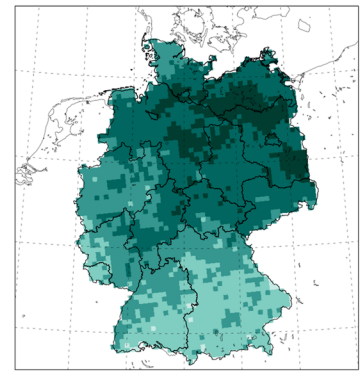
(a) Mean annual BIAS

(b) MD 1st percentile
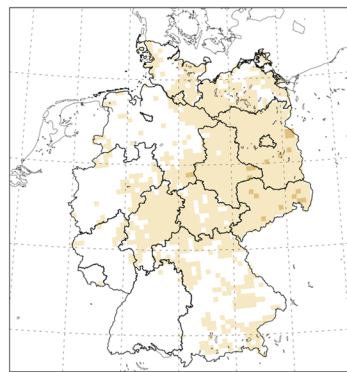
(c) MD 99th percentile
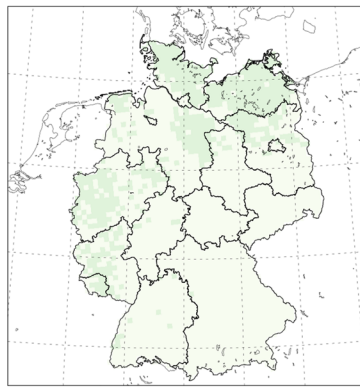
(d) d^{max}
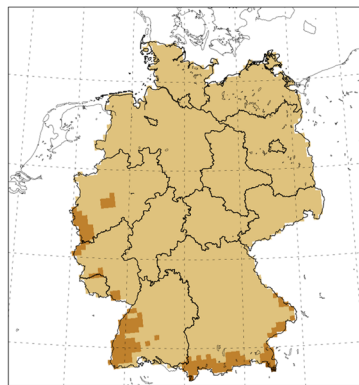
(e) Relative Difference LAG(2-day)
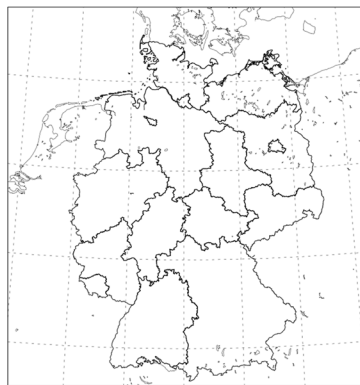
(f) Relative Difference LAG(5-day)

Fig. 11 Precipitation performance metrics (see Sect. 4) for the EPISODES run based on the GCM run CanESM2 run 1
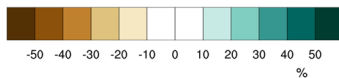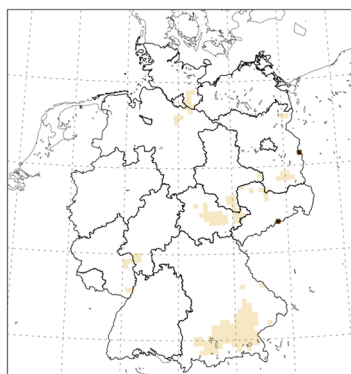
(a) Mean annual BIAS
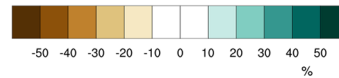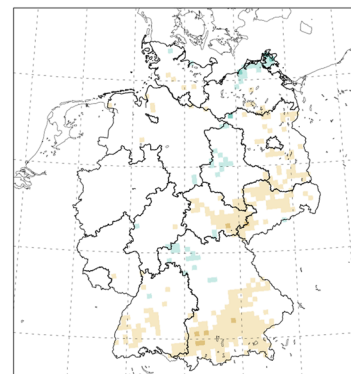
(b) RD 99th percentiles
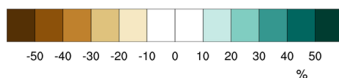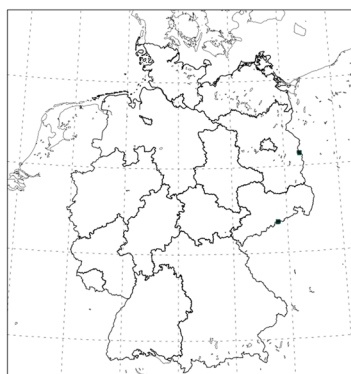
(c) WDFREQ
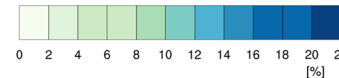
(d) d^{max}
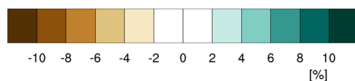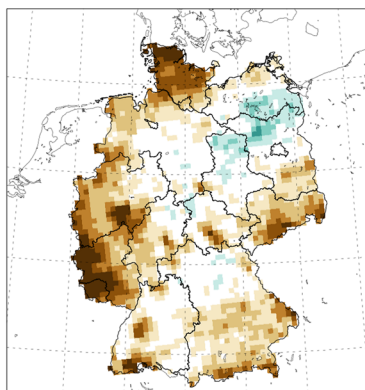
(e) Relative Difference LAG(1-day)
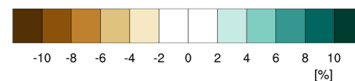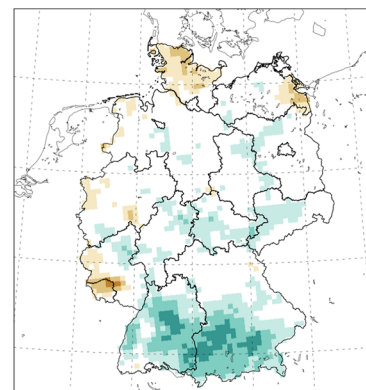
(f) Relative Difference LAG(3-day)

(9) and (10)   Rows 9 and 10 of Table 2 examine the temporal character of observed and modelled temperature auto-correlations. Findings are to be seen in Fig. 10e, f, whereby (e) and (f) depict 2- and 5-day lags, respectively. Values of 2-day lag auto-correlations are slightly underestimated, amounting up to 0.85 instead of 0.90. Associated results of simulated 5-day lag auto-correlations are almost equivalent to those observed. The final line of Table 2 comprises an analysis of the similarity between simulated spatial pattern for temperature and precipitation totals and those derived from records. Results are presented in Table 3). Here historical runs of both GCMs, CanESM2 and MPI-ESM-LR are entered into EPISODES and pertaining results are displayed in the first two rows of Table 3). Since all validation experiments concerned here are known by now, appendant findings for precipitation totals can be directly presented without extensive background descriptions. (5) The mean yearly bias for precipitation is for most German regions below $\pm 10\%$ (see Fig. 11a). (7) Differences between observed and simulated 99th percentiles are for extensive regions across Germany in-between $\pm 10\%$ (see Fig. 11b, yielding an area average of -4%).

(11)   Biases between the amount of simulated and observed wet days are very small. Almost no difference between observations and those modelled via CanESM2's historical run and EPISODES is present (see Fig. 11c). (8) The two-sample Kolmogorov–Smirnov-test, based on the test value d^{max} and $\alpha = 0.05$, detects no significant difference between simulations and records. Hence, on this level of uncertainty EPSIODES' output is to be considered equivalent to the observations (see Fig. 11d). (9) and (10) Differences in 1- and 3-day lag auto-correlations between EPISODES' results and observations are within $\pm 10\%$ (see Fig. 11e, f).

(12)   Results concerning correlations between modelled and observed spatial precipitation pattern are almost perfect for EPISODES (see Table 3). Overall, EPISODES' simulations of historical temperature and precipitation across the entire Germany territory are in high accordance with observations. Except for

Table 3 Results from the Pearson product-moment coefficient of spatial patterns (PACO)

Model cascade	Temperature	Precipitation
MPI-ESM/EPISODES	0.99	0.99
CanESM2/EPISODES	0.99	0.99
MPI-ESM/RCA4	0.95	0.76
MPI-ESM/CLMcom	0.95	0.93
CanESM2/CLMcom	0.94	0.93

the 2-day lag auto-correlation, EPISODE's output performs at least as well as those generated by both RCMs investigated in this study. concerning biases, d^{max} and PACO, EPISODES clearly outperforms them (further details can be found in the "Appendix").

5.2 Climate scenarios

Upon completion of various validation experiments involving several features of observed and simulated temperature and precipitation time series in the past (1971–2000; see Table 2), it seems worthwhile to compare climate change projections generated via EPISODES with those produced by other downscaling strategies. In order to ensure consistency all changes (2071–2100 in reference to 1971–2000) shown in this Section are derived from two GCMs (CanESM2 and MPI-ESM-LR) driven with the same two representative Concentration Pathways (RCP4.5, RCP8.5, Quotes).

Figure 12 shows projected seasonal and yearly temperature and precipitation changes (2071–2100 compared to 1971–2000) for RCP4.5 in Germany.

Presented values refer directly to MPI-ESM-LR output, averaged from 5°E/48°N to 15°E/51°N, as well as to CCLM, RCA4, REMO2009, WETTREG, STARS3 and EPISODES' projections, derived from MPI-ESM-LR scenarios and averaged over all EURO-CORDEX grid points within Germany.

Associated climate change signals are in rather close agreement amongst the above-mentioned downscaling approaches. They all agree on the general climate change signal in Germany: spring and summer come with largest changes in precipitation totals, reaching up to about $+15\%$ in spring and to approximately -20% in summer. In case of temperature most warming is found during summer and winter. Both seasons show mean temperature increases of a little less than $+2\,°C$. On a yearly base temperature increases amount up to somewhat less than $+2\,°C$ and changes in precipitation totals appear negligible.

Figure 13 is arranged as Fig. 12 and contains findings for RCP8.5. Apart from significantly larger climate change signals, the most obvious difference to Fig. 13 is the application of more than twice as many approaches (see Table 4). This is due to the use of both GCMs—CanESM2 and MPI-ESM-LR as well as to the application of two additional statistical downscaling techniques.

The spread amongst the approaches is significantly larger than in case of RCP4.5. This cannot be solely assigned to the inclusion of CanESM2 and two further ESDs. Even if only results based on MPI-ESM-LR are regarded (as in Fig. 12), the compliance for RCP8.5 is much less than for RCP4.5.

Temperature changes (1971–2000 to 2071–2100) averaged over all approaches are significantly more pronounced

(a) MAM

(b) JJA

(c) SON

(d) DJF

(e) YEAR

● MPI-ESM-LR_r1i1p1
□ EPISODES-0-40_MPI-ESM-LR_r1i1p1
✳ CCLM4-8-17_MPI-ESM-LR_r1i1p1
✛ RCA4_MPI-ESM-LR_r1i1p1
✕ REMO2009_MPI-ESM-LR_r1i1p1

◄**Fig. 12** Comparison of seasonal and yearly climate change signals (2071–2100 in reference to 1971–2000) of mean temperature and precipitation totals over all grid cells encased within the German territory. GCM projections and therefrom downscaled (statistically: EPISODES—as well as dynamically: CCLM, RCA4, REMO) regional scale scenarios are driven by RCP4.5. Displayed findings refer to GCM output averaged from 5° to 15° East and from 48° to 51° North and the just mentioned downscaling strategies (see legend)

than in case of RCP4.5. This applies to all seasons (see Fig. 13a–d) and consequently to the entire year (Fig. 13e). Pertaining values are about $+3.5\,°C$ (spring), $+5.0\,°C$ (summer), $+4.0\,°C$ (fall), $+4.0\,°C$ (winter) and approximately $+4.0\,°C$ (year).

Mean seasonal differences over all approaches for precipitation totals for RCP8.5 are larger too, especially during summer (-20%, range from no change to -60%) and winter ($+20\%$, range from no change to $+45\%$). Averaged over the annual cycle (see Fig. 13e), however, mean changes in totals are mostly negligible again.

Aside from these findings some features amongst the investigated approaches appear noteworthy. The distinct impact of the underlying GCMs on attained results appears to be most prominent. Temperature changes based on MPI-ESM-LR are systematically less pronounced than those derived from CanESM2. Considering the seasonal climate change signal of precipitation this dependence is most obvious summer and winter.

Results produced by statistical and dynamical downscaling techniques based on either CanESM2 or MPI-ESM-LR appear to be consistent amongst each other. However, along the seasonal cycle, there is no systematic arrangement of downscaling methods relative to the GCM based averages visible. The only exception are the results from STARS3, which are usually related to the strongest drying. In this study, climate change signals simulated by CanESM2 tend to differ for dynamical and statistical downscaling approaches. This applies to precipitation changes in spring and winter as well as temperature signals during summer and somewhat less during fall. Related features have been detected in other studies too (e.g. Teichmann et al. 2013; Heinrich et al. 2014; Keuler et al. 2016). Compared to both GCM averages EPISODES yields less warming throughout the entire seasonal cycle whereas the other ESD techniques generate more warming in spring and winter and less warming in summer and fall. Dynamically downscaled precipitation sums from CanESM2 tend towards smaller increases than those derived by CanESM2 directly. Such rather ambiguous statements concerning climate change driven precipitation signals are in line with many other studies (e.g. Teichmann et al. 2013; Keuler et al. 2016).

The analysis of EPISODES' projections based on two GCMs (CanESM2 and MPI-ESM-LR) as well as two Representative Concentration Pathways (RCP4.5 and RCP8.5) showed, together with a thorough comparison of EPISODES' temperature and precipitation climate change signals (2071–2100 in reference to 1971–2000) with several other downscaling strategies, a satisfactory quality of EPISODES.

However, after the rather pleasant completion of seasonal and yearly analyses of EPISODES' projections (see Figs. 12 and 13) a daily based evaluation (comparable to Figs. 9 and 8) appears worthwhile, and is shown in Fig. 14. Here, however, no validation experiment is conducted. Both distributions relate to the 'Potsdam grid cell' again.

Changes in occurrence frequencies (downscaled from a CanESM2 realization driven by RCP8.5) of precipitation totals are small. However, reductions in light to medium events are visible, just as increasing frequencies of medium to high totals. This is in line with pertaining statements of the IPCC (2013) and various studies published (e.g. Jacob et al. 2013; Keuler et al. 2016) and means that (1) yearly precipitation across Germany appear to be left unchanged (2071–2100 in reference to 1971–2100), and (2) yearly precipitation totals are projected to be caused by less events producing light to medium sums while strong to heavy precipitation events tend to increase by the end of this century.

EPISODES' regionalization of day-to-day CanESM2/RCP8.5 temperature change signal in the vicinity of Potsdam is shown in Fig. 14a. For the historical period they are in very good agreement with the observations (see Fig. 8a and respective discussions above). Therefore displayed changes are trustworthy in case (1) manhood follows closely the RCP8.5 pathway through the 8 decades ahead, and (2) the climate system of the Earth is well represented by CanESM2 and shares its climate sensitivity in particular, and (3) the climate sensitivity of CanESM2/RCP8.5 is well captured.

From the various transformations, necessary to generate the Potsdam probability distribution corresponding to 2071–2100 from the one related to 1971–2000, merely the most pronounced shifts and consequences will be discussed here.

Largest displacements of temperature values can be found in the very tails (i.e. the extremes) of the past Potsdam distribution. These are of about the same order for both tails, amounting up to approximately $7\,°C$. Hence, very cold daily winter temperatures below approximately $-7\,°C$ will occur extremely seldom (if at all) towards the end of the twenty-first century.

Daily temperatures from about -6 to $+5\,°C$ and from approximately $+10$ to $+17\,°C$ will occur considerably less frequent.

Temperatures within the ranges: $+5$ to $+10\,°C$ and $+20$ to $+25\,°C$ on the other hand will be observed more and more frequently towards the end of this century.
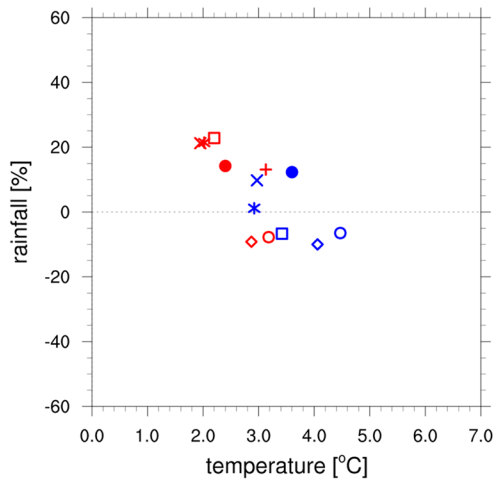
(a) MAM
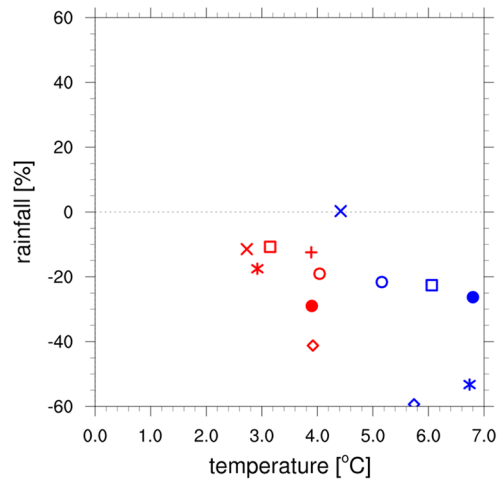
(b) JJA
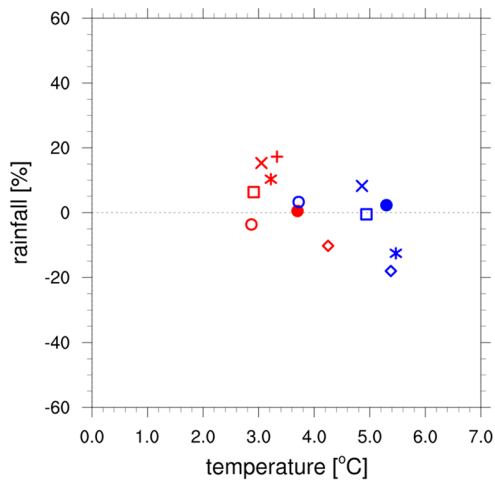
(c) SON
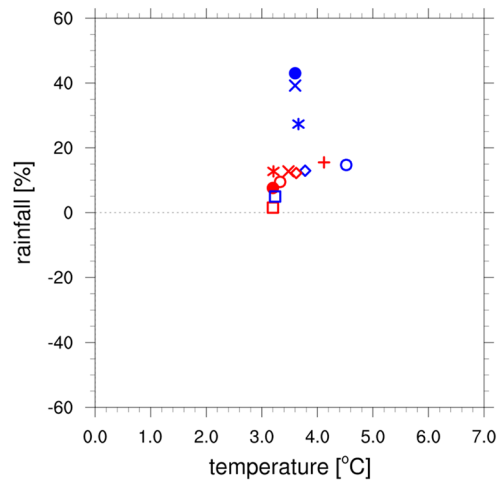
(d) DJF
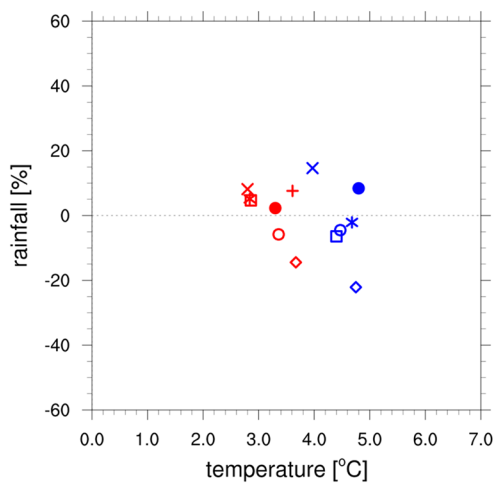
(e) YEAR

- ● MPI-ESM-LR_r1i1p1
- □ EPISODES-1-00_MPI-ESM-LR_r1i1p1
- ◇ STARS3_MPI-ESM-LR_r1i1p1
- ○ WETTREG-2013_MPI-ESM-LR_r1i1p1
- ✳ CCLM4-8-17_MPI-ESM-LR_r1i1p1
- ✕ REMO2009_MPI-ESM-LR_r1i1p1
- ＋ RCA4_MPI-ESM-LR_r1i1p1

- ● CanESM2_r1i1p1
- □ EPISODES-1-00_CanESM2_r1i1p1
- ◇ STARS3_CanESM2_r1i1p1
- ○ WETTREG-2013_CanESM2_r1i1p1
- ✳ CCLM4-8-17_CanESM2_r1i1p1
- ✕ REMO2009_CanESM2_r1i1p1

◄**Fig. 13** Structure as in Fig. 12. Differences to Fig. 12 are: (1) findings shown here are based on two GCMs: CanESM2 and MPI-ESM-LR, which are (2) driven by RCP8.5; (3) next to EPISODES two other statistical downscaling techniques, called STARS3 and WETTREG were applied. This yields to more than twice as many results. As for the labeling of various model approaches: averaging of MPI-ESM-LR and CanESM2 output (5°E to 15°E and 48°N 51°N); CCLM, RCA4, REMO, EPISODES, STARS3 and WETTREG, see the legend

Daily temperatures significantly larger than $+28\,°C$ are very seldom reached in Potsdam or have not been observed so far. Thus, new temperature records up to $+35\,°C$ can be expected from 2071 to 2100.

6 Summary and outlook

This study presents a thorough assessment of the empirical–statistical downscaling (ESD) technique called EPISODES. The evaluation experiments, carried out on yearly, seasonal and daily time scales, focused on temperature and precipitation time series across the entire German territory as well as a single grid cell close to Potsdam.

Pertaining analyses are based on (via EPISODES) downscaled reanalysis data and historical runs of two GCMs (MPI-ESM-LR and CanESM2), which together with a thorough comparison to observations and results achieved by other downscaling strategies, result in a rather successful assessment of EPISODES' performance.

Extending beyond the above described evaluation analyses, EPISODES' projected climate change signals (2071–2100 in reference to 1971–2000), driven by the aforementioned GCMs related to two Representative Concentration Pathways (RCP4.5 and RCP8.5), are compared to other projections that are generated by (1) three RCMs (CCLM, RCA4 and REMO), (2) two statistical methods (STARS3 and WETTREG), and (3) by the spatial average from 5° E/48°N to 15°E/51°N of the two GCMs (CanESM2 and MPI-ESM-LR). Assessments involve yearly, seasonal and daily time scales as well as Germany and the 'Potsdam grid cell'.

EPISODES' projections on yearly and seasonal time scales yield results rather close to those derived from CanESM2 and MPI-ESM-LR and are well within the bandwidth of results associated with all other downscaling techniques. This applies equally to all seasons, to the entire year and to both driving Representative Concentration Pathways (RCP4.5 and RCP8.5). As such, projections generated by EPISODES perform as well as those produced by all other strategies investigated in this study.

Comparisons of EPISODES' day-to-day projections concerning (2071–2100, and the 'Potsdam grid cell') with the historical period (1971–2000) reveal changes in probability distributions, which are pronounced in case of temperature and minor for precipitation totals.

Changes in the very tails (i.e. the extremes) of the temperature distribution amount up to about $+7\,°C$, whilst alterations in occurrence frequencies inside its bimodal distribution are less dominant.

Albeit differences between past and future distributions of precipitation totals are small, decreasing occurrence frequencies of light to medium events and increasing frequencies of strong to heavy precipitation events are visible.

These findings with regard to temperature and precipitation are in line with IPCC assessment reports (e.g. IPCC 2013) and studies published (e.g. Jacob et al. 2013; Keuler et al. 2016; Hübener et al. 2017).

Overall, results confirm that EPISODES is a robust and high performing downscaling technique, which consistently links large-scale processes to regional and local scale phenomena.

Thus, EPISODES can be successfully applied to provide consistent multi-variable and multi-site data sets for the past and ensembles of climate change projections for impact studies and follow-up users for the territory of Germany.

As such EPISODES is suitable to significantly contribute to attaining an urgent goal of central importance: Complementing existing sets of dynamically produced scenarios with empirically generated projections in order to establish ensembles of sufficient sample sizes pertaining to various potential future pathways of mankind (e.g. Hewitson et al. 2013; Landgraf et al. 2015).

This goal his generally pursued, e.g. by IPCC, several climate research centers, impact research communities and stakeholders from federations of tourism, transport, ecology, economy and industry, etc.

However, even though a particular model yields results, concerning the climate system's future, which are in good agreement with findings of other approaches, its climate change signal still remains a projection.

Hence, it is important not to overrate the accordance of findings generated by various models amongst each other. As stated by Hewitson et al. (2013) and Webber and Donner (2016), the closeness of simulations, generated by different models ('precision'), is in fact no measure how well real processes are approximated ('accuracy').

This caveat can be directly transferred to the spatial-temporal resolution. EPISODES' and other downscaling models' ability to produce highly resolved output implies by no mean the reduction of uncertainty (Hewitson et al. 2013).

Regarding the outlook of this work, the perhaps most obvious challenge refers to the generation of sets of regional scale climate change projections derived via EPISODES from various GCMs, forced by different RCPs. The resulting sets of projections need to be merged with those produced by RCMs. In this way the current ensemble-size can be substantially enhanced and pertaining ensembles are characterized by a methodologically sound basis, because they do not

Table 4 List of climate change experiments (2071–2100 compared to 1971–2000) calculated by EPISODES and other downscaling methods from GCM projections driven by RCP4.5 and RCP8.5

No.	Projected quantity	Downscaling setup	Presentation of results
01	Seasonal, yearly climate change signals	MPI-ESM-LR, MPI-ESM-LR/EPISODES, 3 RCMs, RCP4.5	Figure 12
02	Seasonal, yearly climate change signals	2 GCMs, 2 GCMs/EPISODES, 3 RCMs, 2 ESDs, RCP8.5	Figure 13
03	Daily histogram	CanESM2/EPISODES, RCP8.5	Figure 14

Analyses are performed for the entire Germany territory except for the experiment described in no. 3, which refers to the so-called 'Potsdam grid cell'

(a) temperature
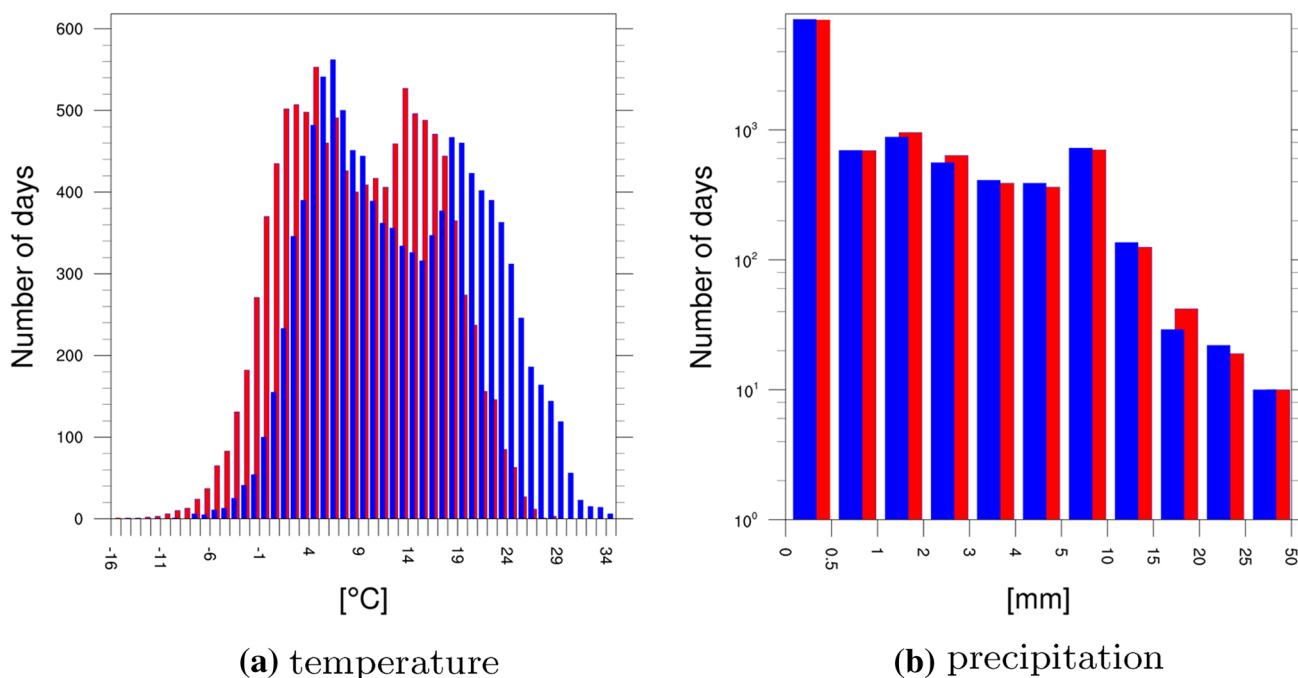
(b) precipitation

Fig. 14 Histograms of (1) past (1971–2000; red bars) daily mean temperatures (left panel) and precipitation totals and (2) potential future (2071–2100; blue bars) daily means of temperature and precipitation sums. Future values are downscaled to the 'Potsdam grid cell' via EPISODES from a CanESM2 projection driven by RCP8.5

just provide various GCM scenarios downscaled by either dynamical or statistical techniques, but encompass projections produced by both strategies at the same time.

The simulation of 'problem specific events' (PSEs) is another goal to be tackled with EPISODES. Europe's transport infrastructure may serve as an example. One prominent PSE refers to certain long-term precipitation episodes, triggering landslides (e.g. Matulla et al. 2017). The risk of damages initiated by PSEs varies with occurrence frequency that is expected to increase with climate change. Through an optimization procedure (variation of e.g. particularly suited large scale predictor combinations and factoring in appropriate weighting sequences taking account of the evolution of weather states in reference to PSEs, see e.g. Matulla 2005; Matulla et al. 2008) that involves the conduct of series of

validation experiments, EPISODES can be adjusted to simulate occurrence frequencies of PSEs with high-performance. Subsequent projections are sought-for input for risk assessment strategies.

So far EPISODES has been applied to the rather low level lands of Germany. Hence, one important target is the evaluation of EPISODES' performance over complex terrain and, if necessary, its adaption to structured orography. A very well suited region for the implementation of related comprehensive adaptation- and validation-analyses, are the European Alps, across which extensive and high-quality data sets are on hand. Current results from Horton et al. (2017) provide a solid basis.

Further topics of future research will focus on:

(1) Investigating precipitation events induced by small scale convection, which are not explicitly resolved by the GCMs. In Addition the lessons learned from the experiments done by San-Martín et al. (2017) needs to be included.

(2) Using EPISODES to simulate sub-daily phenomena. This aspect is mainly limited by the poor temporal coverage in terms climatic periods and their availability as datasets in space comparable to HYRAS. However, circumstances may cease in the short to medium term future as increasingly more and more sub-daily observations for steadily growing periods of time across Germany become available or in case modelling approaches applied to already existing data derive products useable by EPISODES.

(3) Investigating and potentially enhancing EPISODES' skill on sub-climatological timescales down to decades.

(4) Using season specific selection of predictors (see for instance Enke et al. 2005a, b; Wetterhall et al. 2007), which might improve the skill.

Author contributions FK developed the method, carried out projection runs evaluated within this study and prepared the first version of this paper. AP, BF, PL and CM contributed via fruitful discussions and editing a lot to the evaluation and the establishment of the paper in its present form.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

Fig. 15 Mean annual temperature bias (Model−Observation; 1971–2000)
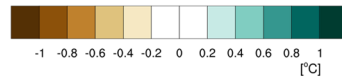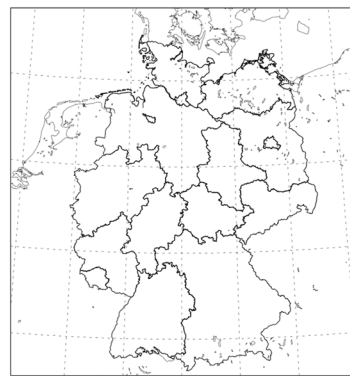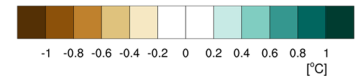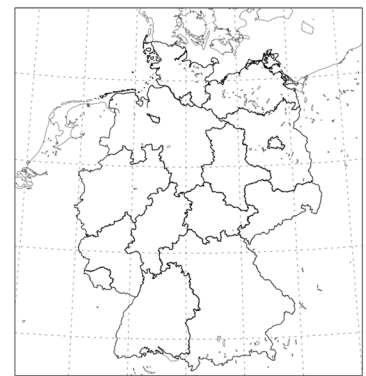
(a) EPISODES (CanESM2)
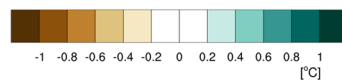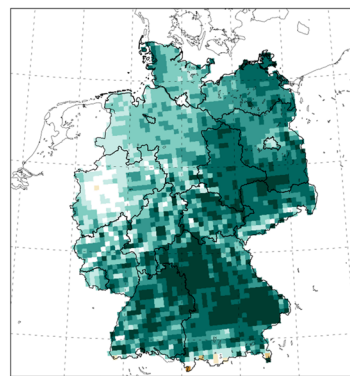
(b) EPISODES (MPI-ESM)
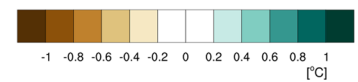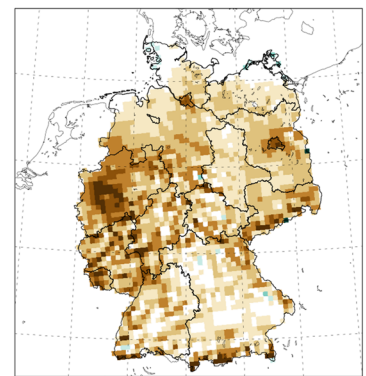
(c) CLMcom (CanESM2)
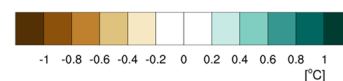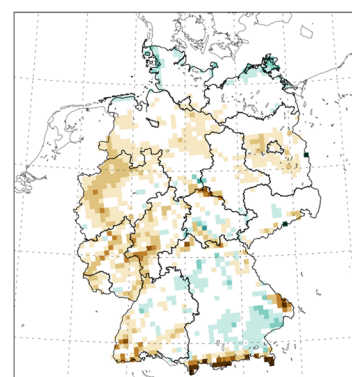
(d) CLMcom (MPI-ESM)

(e) RCA4 (MPI-ESM)

Fig. 16 Temperature 1st percentile difference (Model−Observation; 1971–2000)
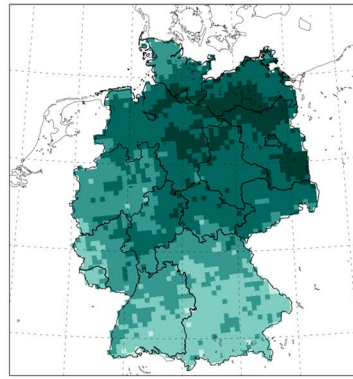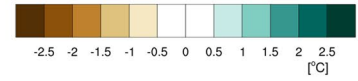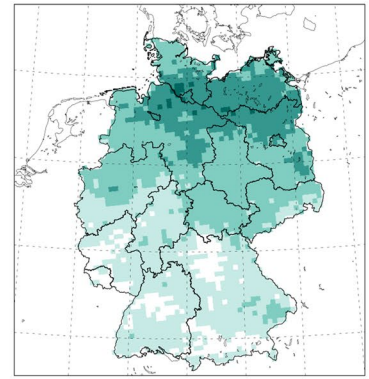
(a) EPISODES (CanESM2)
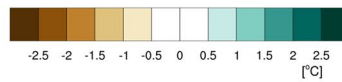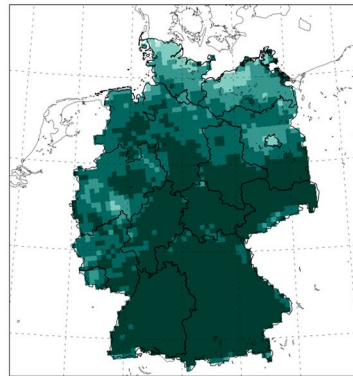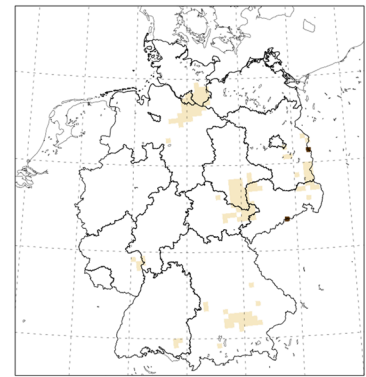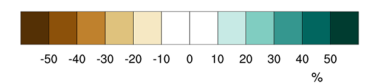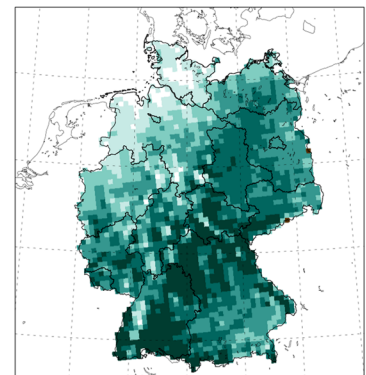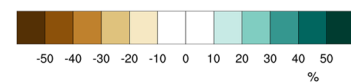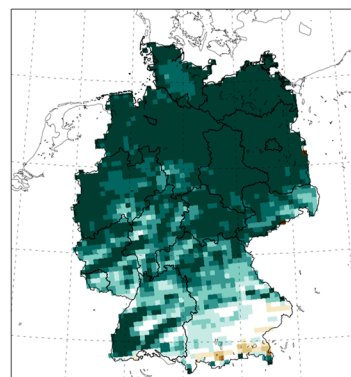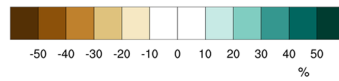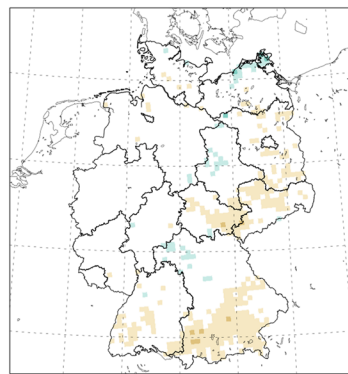
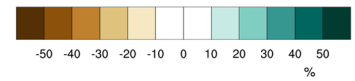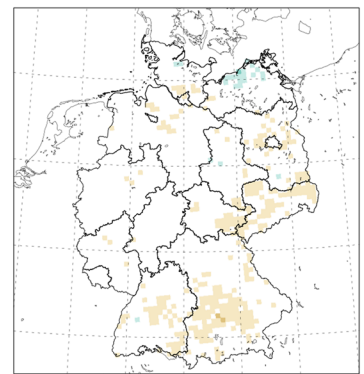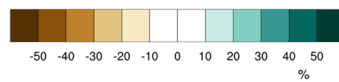(b) EPISODES (MPI-ESM)
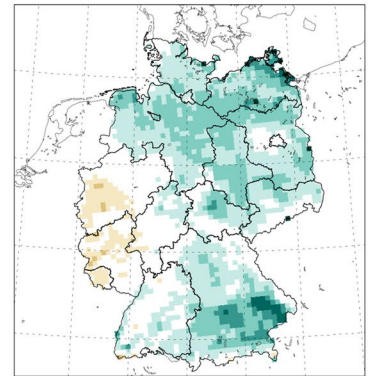
(c) CLMcom (CanESM2)
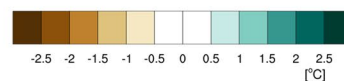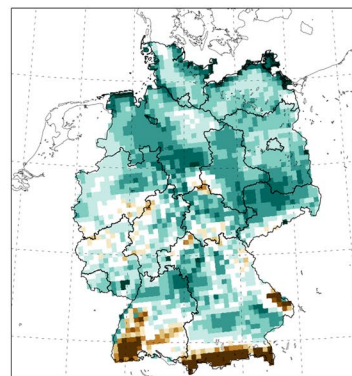
(d) CLMcom (MPI-ESM)

(e) RCA4 (MPI-ESM)

Fig. 17 Temperature 99th percentile difference (Model−Observation; 1971–2000)
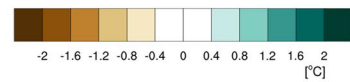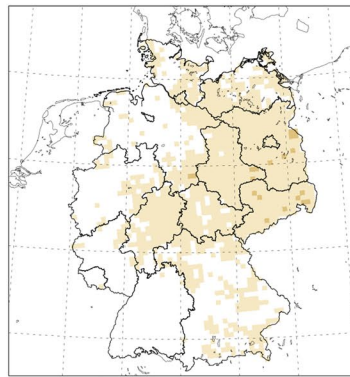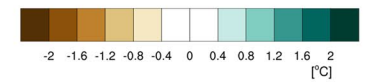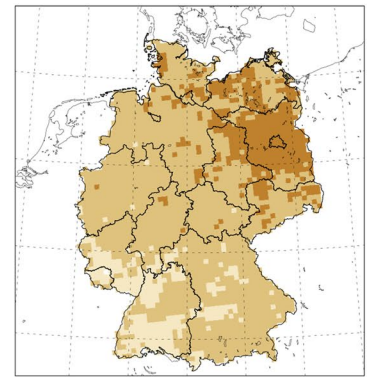
(a) EPISODES (CanESM2)
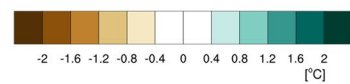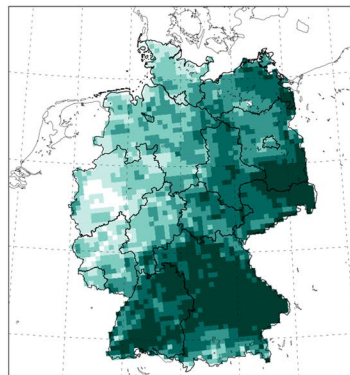
(b) EPISODES (MPI-ESM)
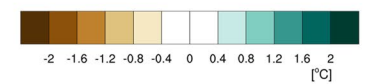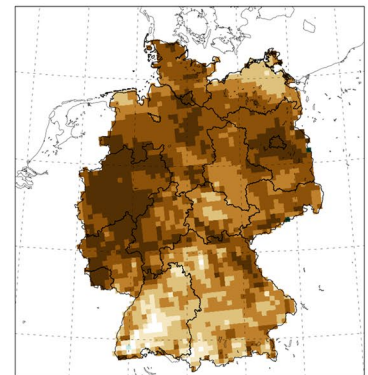
(c) CLMcom (CanESM2)
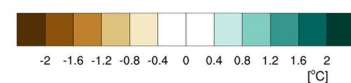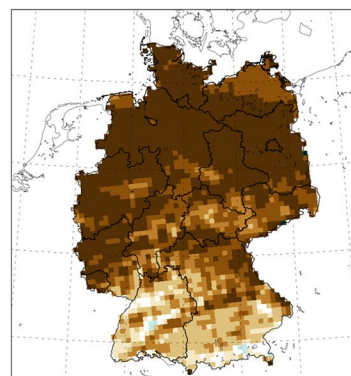
(d) CLMcom (MPI-ESM)

(e) RCA4 (MPI-ESM)

Fig. 18 Temperature 2-day lag auto-correlation difference (Model−Observation; 1971–2000)
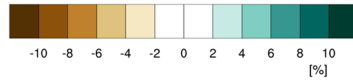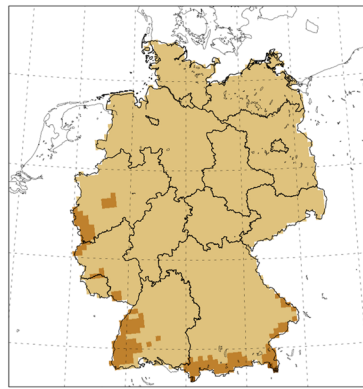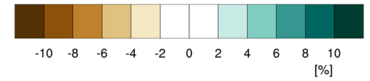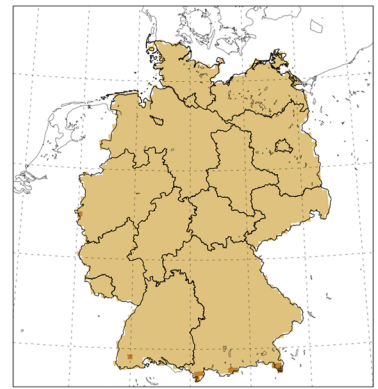
(a) EPISODES (CanESM2)
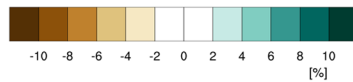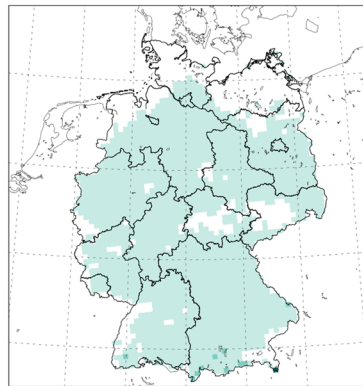
(b) EPISODES (MPI-ESM)
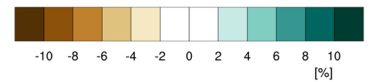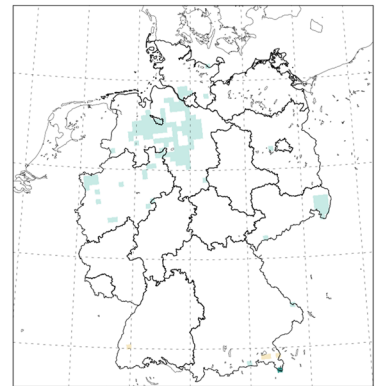
(c) CLMcom (CanESM2)
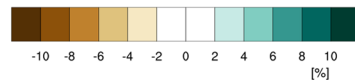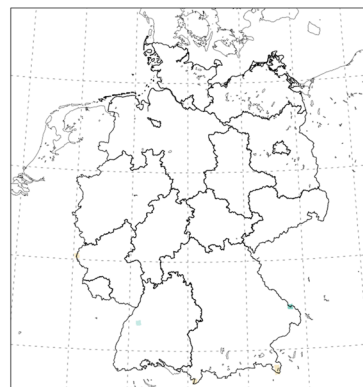
(d) CLMcom (MPI-ESM)

(e) RCA4 (MPI-ESM)

Fig. 19 Temperature 5-day lag auto-correlation difference (Model−Observation; 1971–2000)
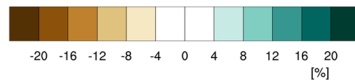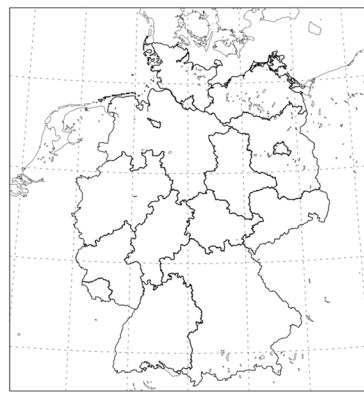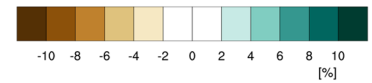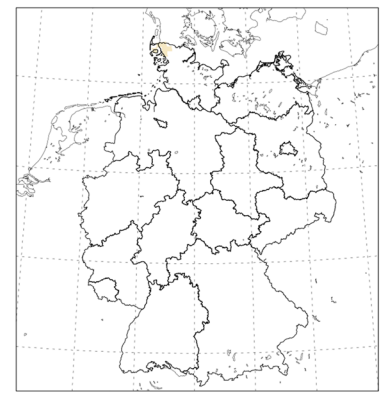
(a) EPISODES (CanESM2)
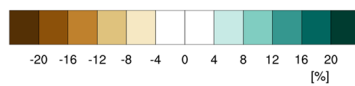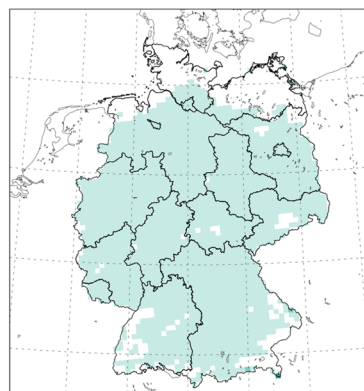
(b) EPISODES (MPI-ESM)
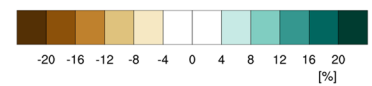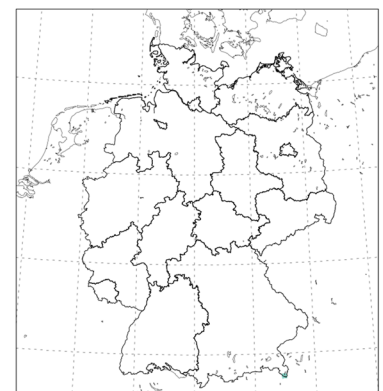
(c) CLMcom (CanESM2)
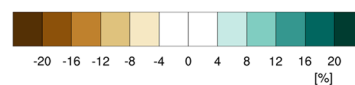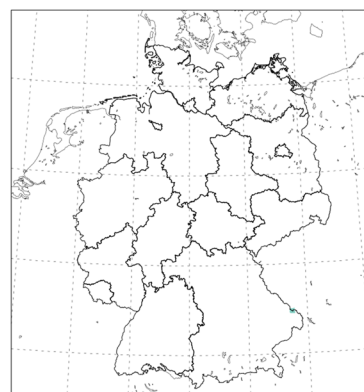
(d) CLMcom (MPI-ESM)

(e) RCA4 (MPI-ESM)

**Fig. 20** Maximum distance d^{max} for temperature between the model and the HYRAS-data empirical distribution function (1971–2000)
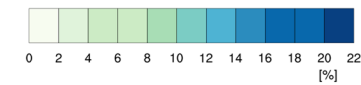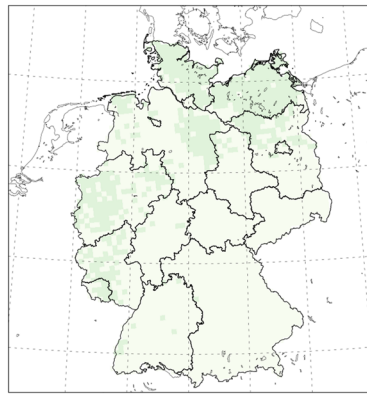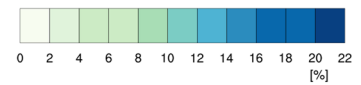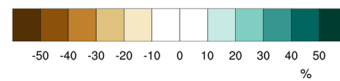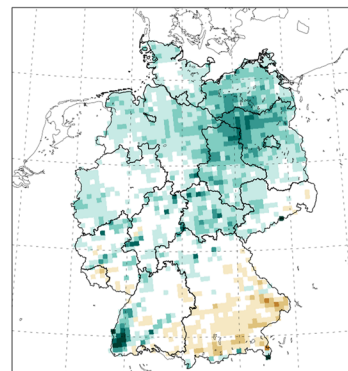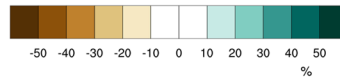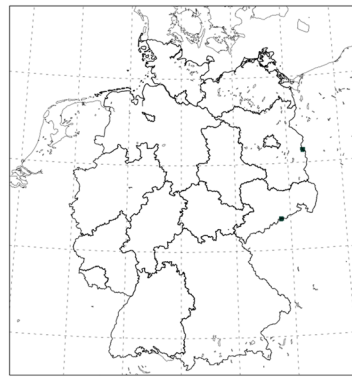

(a) EPISODES (CanESM2)
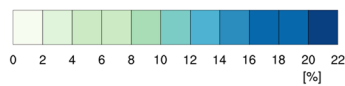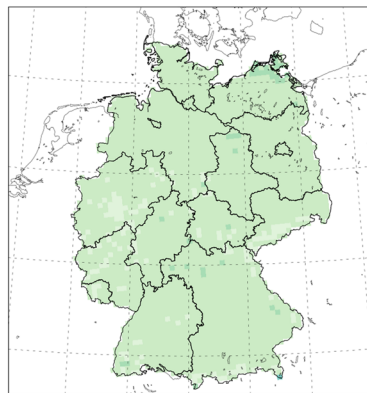
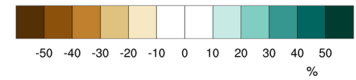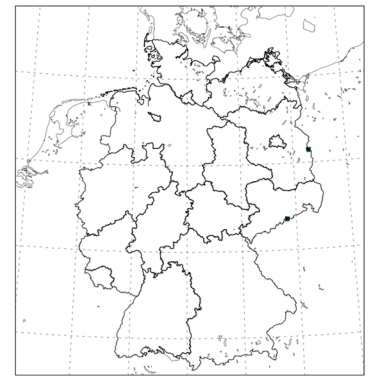(b) EPISODES (MPI-ESM)
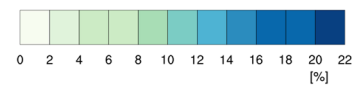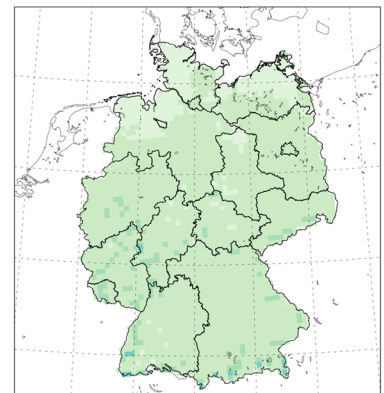
(c) CLMcom (CanESM2)
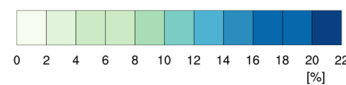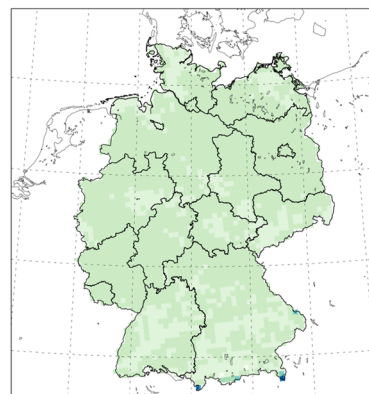
(d) CLMcom (MPI-ESM)

(e) RCA4 (MPI-ESM)

Fig. 21 Mean annual precipitation bias (1971–2000)
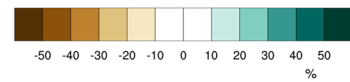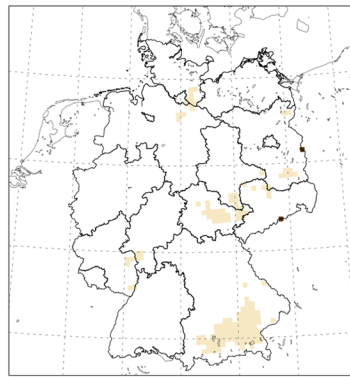
(a) EPISODES (CanESM2)
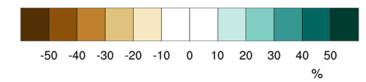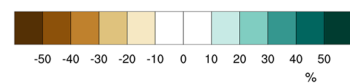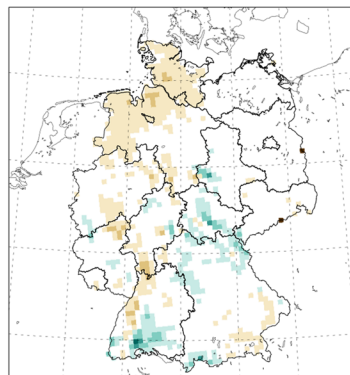
(b) EPISODES (MPI-ESM)

(c) CLMcom (CanESM2)

(d) CLMcom (MPI-ESM)

(e) RCA4 (MPI-ESM)

Fig. 22 Precipitation 99th percentile relative difference (Model−Observation; 1971–2000)

(a) EPISODES (CanESM2)
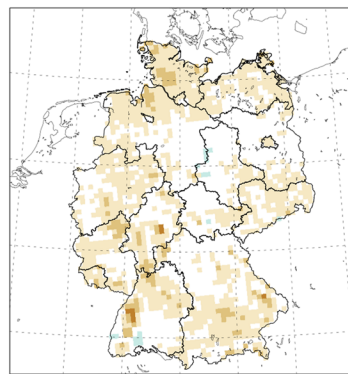
(b) EPISODES (MPI-ESM)
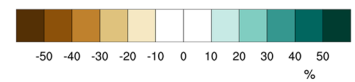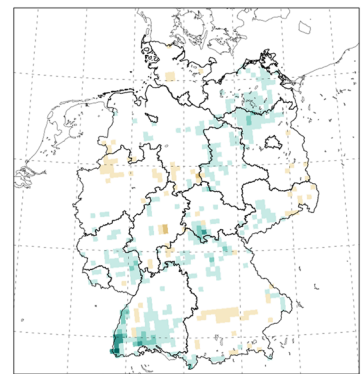
(c) CLMcom (CanESM2)

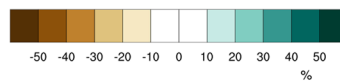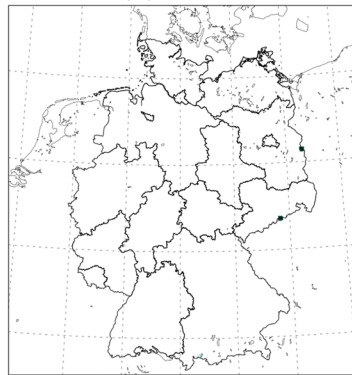**(d)** CLMcom (MPI-ESM)

(e) RCA4 (MPI-ESM)

Fig. 23 Relative difference
between number of wet days
observation and model. The
values are based on the time
period 1971–2000

(a) EPISODES (CanESM2)

(b) EPISODES (MPI-ESM)

(c) CLMcom (CanESM2)

(d) CLMcom (MPI-ESM)

(e) RCA4 (MPI-ESM)

Fig. 24 Precipitation 1-day lag auto-correlation difference (Model−Observation; 1971–2000)

(a) EPISODES (CanESM2)
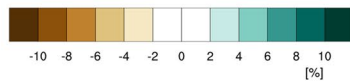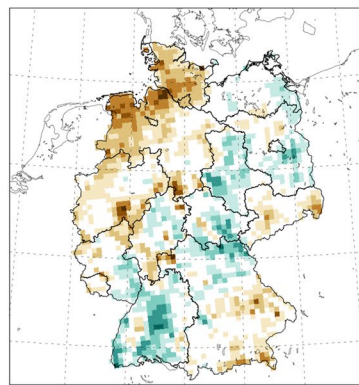
(b) EPISODES (MPI-ESM)
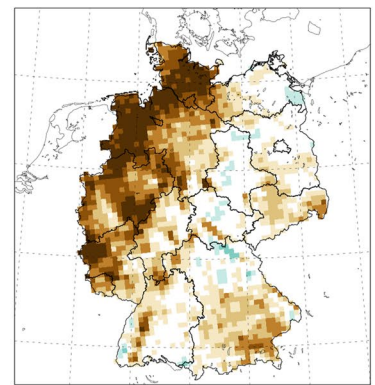
(c) CLMcom (CanESM2)
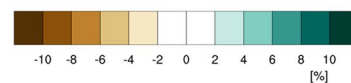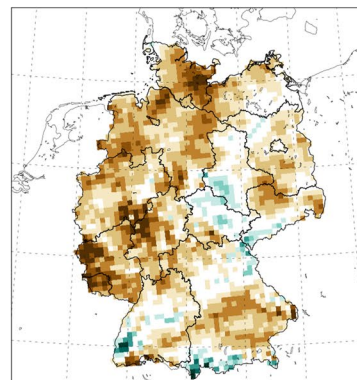
(d) CLMcom (MPI-ESM)

(e) RCA4 (MPI-ESM)

Fig. 25 Precipitation 3-day
lag auto-correlation differ-
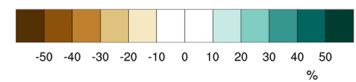ence (Model−Observation;
1971–2000)
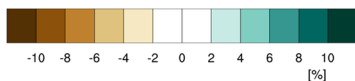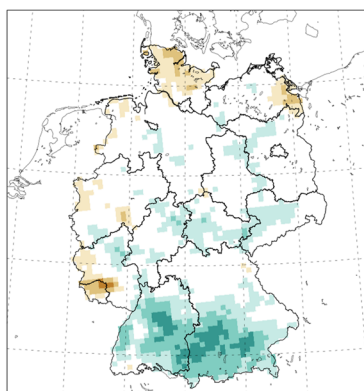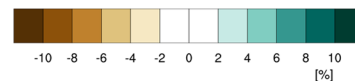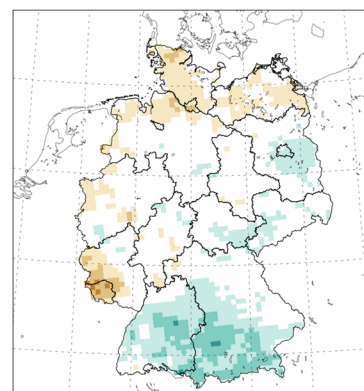
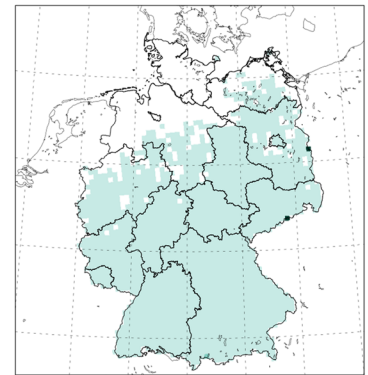**(a)** EPISODES (CanESM2)
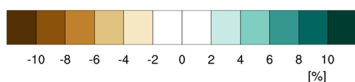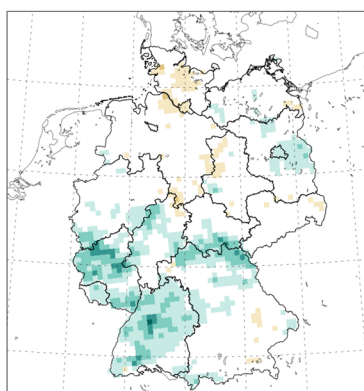
(b) EPISODES (MPI-ESM)
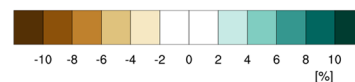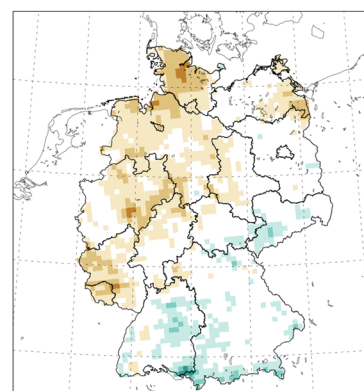
(c) CLMcom (CanESM2)
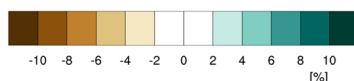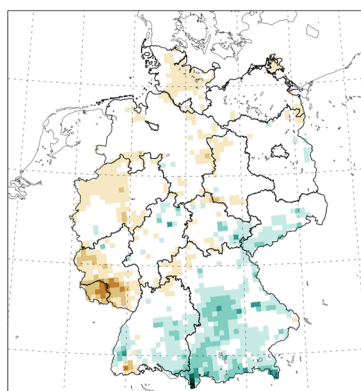
(d) CLMcom (MPI-ESM)

(e) RCA4 (MPI-ESM)

Fig. 26 Maximum distance d^{max} for precipitation between the model and the HYRAS-data empirical distribution function (1971–2000)
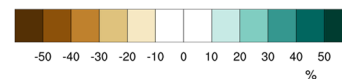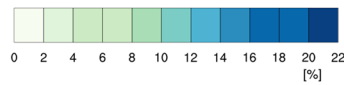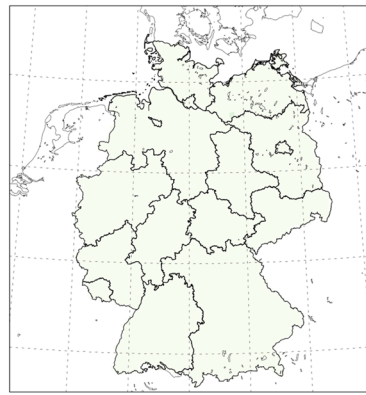

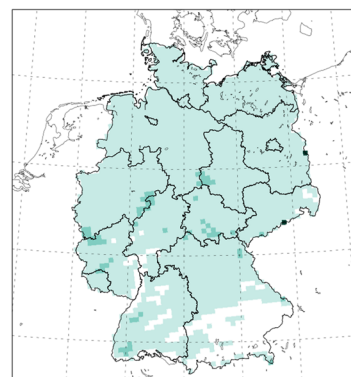**(a)** EPISODES (CanESM2)
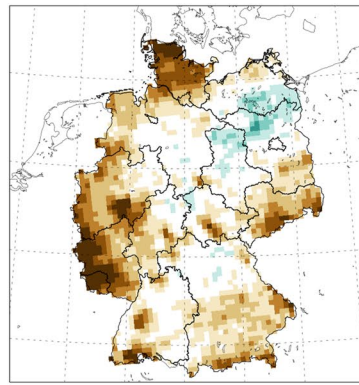

(b) EPISODES (MPI-ESM)
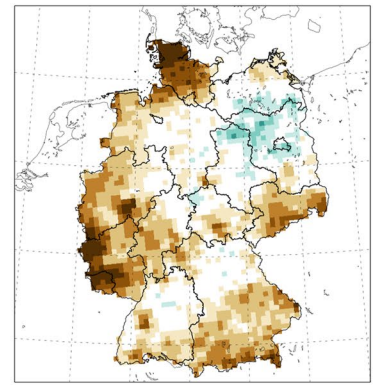

(c) CLMcom (CanESM2)

(d) CLMcom (MPI-ESM)

(e) RCA4 (MPI-ESM)

Appendix

See Figs. 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 and 26.

References

Beersma JJ, Buishand T (2003) Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation. Clim Res 25:121–133

Benestad R, Hanssen-Bauer I, Cheng D (2008) Empirical-statistical downscaling. World Scientific Publishing, Singapore

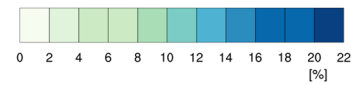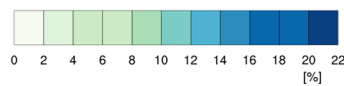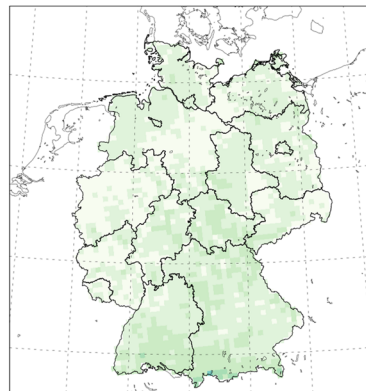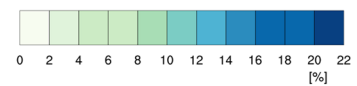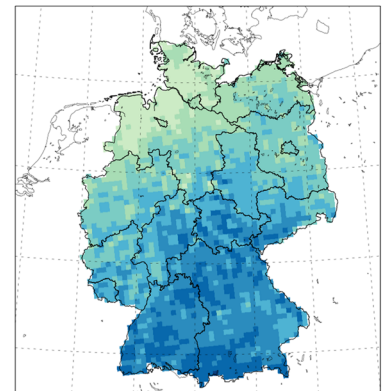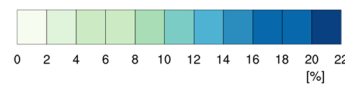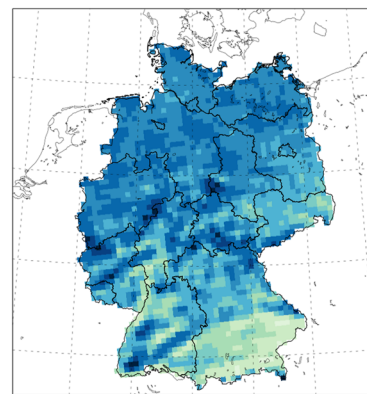Canadian Centre for Climate Modelling and Analysis (CCCma) (2015a) CanESM2 model output prepared for CMIP5 historical, served by ESGF

Canadian Centre for Climate Modelling and Analysis (CCCma) (2015b) CanESM2 model output prepared for CMIP5 rcp85, served by ESGF

Cinquini L, Crichton D, Mattmann C, Harney J, Shipman G, Wang F, Ananthakrishnan R, Miller N, Denvil S, Morgan M, Pobre Z, Bell GM, Doutriaux C, Drach R, Williams D, Kershaw P, Pascoe S, Gonzalez E, Fiore S, Schweitzer R (2014) The earth system grid federation: an open infrastructure for access to distributed geospatial data. Futur Gener Comput Syst 36:400–417

Daoud AB, Sauquet E, Bontron G, Obled C, Lang M (2016) Daily quantitative precipitation forecasts based on the analogue method: improvements and application to a French large river basin. Atmos Res Part A 169:147–159

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J-J, Park B-K, Peubey C, de Rosnay P, Tavolato C, Thépaut J-N, Vitart F (2011) The era-interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137(656):553–597

Edwards P (2010) A vast machine—computer models, climate data, and the politics of global warming. MIT Press, Cambridge

Edwards P (2011) History of climate modeling. WIREs Clim Change 2:128–139. https://doi.org/10.1002/wcc.95

Enke W, Deutschländer T, Schneider F, Küchler W (2005a) Results of five regional climate studies applying a weather pattern based downscaling method to ECHAM4 climate simulations. Meteorol Z 14:247–257. https://doi.org/10.1127/0941-2948/2005/0028

Enke W, Schneider F, Deutschländer T (2005b) A novel scheme to derive optimized circulation pattern classifications for downscaling and forecast purposes. Theor Appl Climatol 82:51–63. https://doi.org/10.1007/s00704-004-0116-x, http://www.springerlink.com/content/m737r30600222273/fulltext.pdf

Flato G, Marotzke J, Abiodun B, Braconnot P, Chou S, Collins W, Cox P, Driouech F, Emori S, Eyring V, Forest C, Gleckler P, Guilyardi E, Jakob C, Kattsov V, Reason C, Rummukainen M (2013)

Evaluation of Climate Models, book section 9. Cambridge University Press, Cambridge, pp 741–866

Fowler HJ, Blenkinsop S, Tebaldi C (2007) Review: Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. Int J Climatol 27:1547–1578

Frick C, Steiner H, Mazurkiewicz A, Riediger U, Rauthe M, Reich T, Gratzki A (2014) Central European high-resolution gridded daily data sets (HYRAS): mean temperature and relative humidity. Meteorol Z 23(1):15–32. https://doi.org/10.1127/0941-2948/2014/0560

Fu G, Viney NR, Charles SP (2010) Evaluation of various root transformations of daily precipitation amounts fitted with a normal distribution for australia. Theor Appl Climatol 99(1):229–238

Giorgetta M, Jungclaus J, Reick C, Legutke S, Brovkin V, Crueger T, Esch M, Fieg K, Glushak K, Gayler V, Haak H, Hollweg H-D, Kinne S, Kornblueh L, Matei D, Mauritsen T, Mikolajewicz U, Müller W, Notz D, Raddatz T, Rast S, Roeckner E, Salzmann M, Schmidt H, Schnur R, Segschneider J, Six K, Stockhause M, Wegner J, Widmann H, Wieners K-H, Claussen M, Marotzke J, Stevens BA (2012a) CMIP5 simulations of the Max Planck Institute for Meteorology (MPI-M) based on the MPI-ESM-LR model: the historical experiment, served by ESGF

Giorgetta M, Jungclaus J, Reick C, Legutke S, Brovkin V, Crueger T, Esch M, Fieg K, Glushak K, Gayler V, Haak H, Hollweg H-D, Kinne S, Kornblueh L, Matei D, Mauritsen T, Mikolajewicz U, Müller W, Notz D, Raddatz T, Rast S, Roeckner E, Salzmann M, Schmidt H, Schnur R, Segschneider J, Six K, Stockhause M, Wegner J, Widmann H, Wieners K-H, Claussen M, Marotzke J, Stevens BA (2012b) CMIP5 simulations of the Max Planck Institute for Meteorology (MPI-M) based on the MPI-ESM-LR model: the rcp45 experiment, served by ESGF

Giorgetta M, Jungclaus J, Reick C, Legutke S, Brovkin V, Crueger T, Esch M, Fieg K, Glushak K, Gayler V, Haak H, Hollweg H-D, Kinne S, Kornblueh L, Matei D, Mauritsen T, Mikolajewicz U, Müller W, Notz D, Raddatz T, Rast S, Roeckner E, Salzmann M, Schmidt H, Schnur R, Segschneider J, Six K, Stockhause M, Wegner J, Widmann H, Wieners K-H, Claussen M, Marotzke J, Stevens BA (2012c) CMIP5 simulations of the Max Planck Institute for Meteorology (MPI-M) based on the MPI-ESM-LR model: the rcp85 experiment, served by ESGF

Gutiérrez J, Bedia J, Benestad R, Page C (2013a) SPECS D52.1. Review of the different statistical downscaling methods for S2D prediction. Tech. Notes Santander Meteorology Group (CSIC-UC): GMS, vol 3, pp 1–12

Gutiérrez J, San-Martín D, Brands S, Manzanas R, Herrera S (2013b) Reassessing statistical downscaling techniques for their robust application under climate change conditions. J Clim 26:171–188. https://doi.org/10.1175/JCLI-D-11-00687.1

Haslinger K, Anders I, Hofstätter M (2013) Regional climate modelling over complex terrain: an evaluation study of COCMO-CLM hindcast model runs for the greater alpine region. Clim Dyn 40(1):511–529

Heinrich G, Gobiet A, Mendlik T (2014) Extended regional climate model projections for Europe until the mid-twentyfirst century: combining ensembles and cmip3. Clim Dyn 42(1):521–535. https://doi.org/10.1007/s00382-013-1840-7

Hewitson BC, Daron J, Crane RG, Zermoglio MF, Jack C (2013) Interrogating empirical-statistical downscaling. Clim Change 122(4):539–554. https://doi.org/10.1007/s10584-013-1021-z

Horton P, Jaboyedoff M, Obled C (2017) Using genetic algorithms to optimize the analogue method for precipitation prediction in the Swiss Alps. J Hydrol. https://doi.org/10.1016/j.jhydrol.2017.04.017

Howell WE (1960) A comparison between two transformations used in normalizing meteorological data. J Meteorol 17(6):684–684

Hübener H, Bülow K, Fooken C, Früh B, Hoffmann P, Höpp S, Keuler K, Menz C, Mohr V, Radtke K, Ramthun H, Spekat A, Steger C, Toussaint F, Warrach-Sagi K, Woldt M (2017) Reklies-de ergebnisbericht. Technical report, Hessisches Landesamt für Naturschutz, Umwelt und Geologie (HLNUG), Germany

IPCC (2007) Climate Change 2007: the physical science basis. In: Solomon S, Qin D, Manning M, Chen Z, Maquis M, Averyt KB, Tignor M, Miller HL (eds) Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge

IPCC (2013) Climate Change 2013: the physical science basis. In: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Technical report, Intergovernmental Panel On Climate Change, Cambridge. http://www.climatechange2013.org/images/report/WG1AR5_ALL_FINAL.pdf (Stand: 10.02.2014)

Jacob D, Petersen J, Eggert B, Alias A, Christensen OB, Bouwer L, Braun A, Colette A, Déqué M, Georgievski G, Georgopoulou E, Gobiet A, Menut L, Nikulin G, Haensler A, Hempelmann N, Jones C, Keuler K, Kovats S, Kröner N, Kotlarski S, Kriegsmann A, Martin E, Meijgaard E, Moseley C, Pfeifer S, Preuschmann S, Radermacher C, Radtke K, Rechid D, Rounsevell M, Samuelsson P, Somot S, Soussana J-F, Teichmann C, Valentini R, Vautard R, Weber B, Yiou P (2013) Euro-cordex: new high-resolution climate change projections for European impact research. Regional Environmental Change, pp 1–16

Jóhannesson T, Jonsson T, Källén E, Kaas E (1995) Climate change scenarios for the nordic countries. Clim Res 5:181–195

Jones P, Kilsby C, Harpham C, Glenis V, Burton A (2009) UK climate projections science report: projections of future daily climate for the UK from the weather generator. Technical report, University of Newcastle, UK

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Leetmaa A, Reynolds R, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo K, Ropelewski C, Wang J, Jenne R, Joseph D (1996) The NCEP/NCAR 40-Year Reanalysis Project. Bull Am Meteorol Soc 77:437–471

Keuler K, Radtke K, Kotlarski S, Lüthi D (2016) Regional climate change over Europe in cosmo-clm: influence of emission scenario and driving global model. Meteorol Z 25(2):121–136. https://doi.org/10.1127/metz/2016/0662

Klein W, Lewis B, Enger I (1959) Objective prediction of five-day mean temperatures during winter. J Meteorol 16:672–682

Kotlarski S, Keuler K, Christensen OB, Colette A, Déqué M, Gobiet A, Goergen K, Jacob D, Lüthi D, van Meijgaard E, Nikulin G, Schär C, Teichmann C, Vautard R, Warrach-Sagi K, Wulfmeyer V (2014) Regional climate modeling on european scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. Geosci Model Dev 7(4):1297–1333

Kotlarski S, Szabó P, Herrera S, Räty O, Keuler K, Soares PM, Cardoso RM, Bosshard T, Pagé C, Boberg F, Gutiérrez JM, Isotta FA, Jaczewski A, Kreienkamp F, Liniger MA, Lussana C, Pianko-Kluczyíska K (2017) Observational uncertainty and regional climate model evaluation: a pan-European perspective. Int J Climatol. https://doi.org/10.1002/joc.5249

Kreienkamp F, Spekat A, Enke W (2013) The weather generator used in the empirical statistical downscaling method WETTREG. Atmosphere 4:169–197. http://www.mdpi.com/2073-4433/4/2/169/

Landgraf M, Matulla C, Haimberger L (2015) Statistically downscaled projections of local scale temperature in the topographically complex terrain of austria up to the end of the 21st century. Meteorol Z 24(4):425–440. https://www.schweizerbart.de/content/papers/download/84736

Lockwood JG, Gregory S, Scorer RS (1991) Climate change: the IPCC scientific assessment. Report prepared by working group Houghton IJT, Jenkins GJ, Ephraums JJ (eds). Int J Climatol 11(4):457–458

Lorenz E (1963) Deterministic nonperiodic flow. J Atmos Sci 20:130–141

Lorenz E (1969) Atmospheric predictability as revealed by naturally occurring analogues. J Atmos Sci 26:636–646

Maraun D, Huth R, Gutiéérrez J, San Martín D, Dubrovský M, Fischer A, Hertig E, Soares P, Bartholy J, Pongracz R, Widmann M, Casado JM, Ramos P, Bedia J (2017a) The value perfect predictor experiment: evaluation of temporal variability. Int J Climatol. https://doi.org/10.1002/joc.5222

Maraun D, Shepherd TG, Widmann M, Zappa G, Walton D, Gutiérrez José M, Hagemann S, Richter I, Soares PMM, Hall A, Mearns LO (2017b) Towards process-informed bias correction of climate change simulations. Nature, Climate Change, pp 764–773

Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienen S, Rust HW, Sauter T, Themeßl M, Venema VKC, Chun KP, Goodess CM, Jones RG, Onof C, Vrac M, Thiele-Eich I (2010) Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. Rev Geophys. https://doi.org/10.1029/2009RG000314

Maraun D, Widmann M, Gutiérrez JM, Kotlarski S, Chandler RE, Hertig E, Wibig J, Huth R, Wilcke RA (2015) Value: a framework to validate downscaling approaches for climate change studies. Earth's Future 3(1):1–14

Matulla C (2005) Regional, seasonal and predictor-optimized downscaling to provide groups of local scale scenarios in the complex structured terrain of austria. Meteorol Z 14(1):31–45

Matulla C, Hollósi B, Andre K, Gringinger J, Chimani B, Namyslo J, Fuchs T, Auerbach M, Herrmann C, Sladek B, Berghold H, Gschier R, Eichinger-Vill E (2017) Climate change driven evolution of hazards to Europe's transport infrastructure throughout the twenty-first century. Theor Appl Climatol. https://doi.org/10.1007/s00704-017-2127-4

Matulla C, Scheifinger H, Menzel A, Koch E (2003) Exploring two methods for statistical downscaling of central European phenological time series. Int J Biometeorol 48(2):56–64

Matulla C, Zhang X, Wang XL, Wang J, Zorita E, Wagner S, von Storch H (2008) Influence of similarity measures on the performance of the analog method for downscaling daily precipitation. Clim Dyn 30(2):133–144

Mesinger F, DiMego G, Kalnay E, Mitchell K, Shafran PC, Ebisuzaki W, Jović D, Woollen J, Rogers E, Berbery EH, Ek MB, Fan Y, Grumbine R, Higgins W, Li H, Lin Y, Manikin G, Parrish D, Shi W (2006) North American regional reanalysis. Bull Am Meteorol Soc 87(3):343–360

Moss R, Babiker M, Brinkman S, Calvo E, Carter T, Edmonds J, Elgizouli I, Emori S, Erda L, Hibbard K, Jones R, Kainuma M, Kelleher J, Lamarque J, Manning M, Matthews B, Meehl J, Meyer L, Mitchell J, Nakicenovic N, O'Neill B, Pichs R, Riahi K, Rose S, Runci P, Stouffer R, van Vuuren D, Weyant J, Wilbanks T, van Ypersele J, Zurek M (2008) Technical summary: towards new scenarios for analysis of emissions, climate change, impacts and response strategies, Noordwijkerhout. Technical report, Intergovernmental Panel on Climate Change

Moss R, Edmonds J, Hibbard K, Manning M, Rose S, van Vuuren D, Carter T, Emori S, Kainuma M, Kram T, Meehl G, Mitchell J, Nakicenovic N, Riahi K, Smith S, Stouffer R, Thomson A, Weyant J, Wilbanks T (2010) The next generation of scenarios for climate change research and assessment. Nature 463:747–756

Müller P (2010) Constructing climate knowledge with computer models. Wiley Interdiscip Rev Clim Change 1(4):565–580

NCAR (2017) The NCAR Command Language (Version 6.4.0) [Software]

Nikulin G, Jones C, Giorgi F, Asrar G, Büchner M, Cerezo-Mota R, Christensen OB, Déqué M, Fernandez J, Hänsler A, van Meijgaard E, Samuelsson P, Sylla MB, Sushama L (2012) Precipitation climatology in an ensemble of CORDEX-Africa regional climate simulations. J Clim 25(18):6057–6078

Nikulin G, Bosshard T, Yang W, Bärring L, Wilcke R, Vrac M, Vautard R, Noel T, Gutiérrez JM, Herrera S, Fernández J, Haugen JE, Benestad R, Landgren OA, Grillakis M, Ioannis T, Koutroulis A, Dosio A, Ferrone A, Switanek M (2015) Bias Correction Intercomparison Project (BCIP): an introduction and the first results. In: EGU General Assembly Conference Abstracts, EGU General Assembly Conference Abstracts, vol 17, p 2250

Onogi K, Tsutsui J, Koide H, Sakamoto M, Kobayashi S, Hatsushika H, Matsumoto T, Yamazaki N, Kamahori H, Takahashi K, Kadokura S, Wada K, Kato K, Oyama R, Ose T, Mannoji N, Taira R (2007) The JRA-25 reanalysis. J Meteorol Soc Jpn Ser II 85(3):369–432

Ozturk T, Turp MT, Türkes M, Kurnaz ML (2017) Projected changes in temperature and precipitation climatology of central asia CORDEX region 8 by using REGCM4.3.5. Atmos Res 183:296–307

Pätzhold S, Balzer K (1995) Dokumentation afreg-routine. Technical report, Deutscher Wetterdienst

Rauthe M, Steiner H, Riediger U, Mazurkiewicz A, Gratzki A (2013) A Central European precipitation climatology? Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS). Meteorol Z 22(3):235–256. https://doi.org/10.1127/0941-2948/2013/0436

Raynaud D, Hingray B, Zin I, Anquetin S, Debionne S, Vautard R (2017) Atmospheric analogues for physically consistent scenarios of surface weather in Europe and Maghreb. Int J Climatol 37(4):2160–2176

Rienecker MM, Suarez MJ, Gelaro R, Todling R, Bacmeister J, Liu E, Bosilovich MG, Schubert SD, Takacs L, Kim G-K, Bloom S, Chen J, Collins D, Conaty A, da Silva A, Gu W, Joiner J, Koster RD, Lucchesi R, Molod A, Owens T, Pawson S, Pegion P, Redder CR, Reichle R, Robertson FR, Ruddick AG, Sienkiewicz M, Woollen J (2011) Merra: Nasa's modern-era retrospective analysis for research and applications. J Clim 24(14):3624–3648

Rummukainen M (2010) State-of-the-art with regional climate models. WIREs Clim Change 1:82–96

Saha S, Moorthi S, Pan H-L, Wu X, Wang J, Nadiga S, Tripp P, Kistler R, Woollen J, Behringer D, Liu H, Stokes D, Grumbine R, Gayno G, Wang J, Hou Y-T, Chuang H-Y, Juang H-MH, Sela J, Iredell M, Treadon R, Kleist D, Delst PV, Keyser D, Derber J, Ek M, Meng J, Wei H, Yang R, Lord S, Dool HVD, Kumar A, Wang W, Long C, Chelliah M, Xue Y, Huang B, Schemm J-K, Ebisuzaki W, Lin R, Xie P, Chen M, Zhou S, Higgins W, Zou C-Z, Liu Q, Chen Y, Han Y, Cucurull L, Reynolds RW, Rutledge G, Goldberg M (2010) The ncep climate forecast system reanalysis. Bull Am Meteorol Soc 91(8):1015–1057

San-Martín D, Manzanas R, Brands S, Herrera S, Gutiérrez JM (2017) Reassessing model uncertainty for regional projections of precipitation with an ensemble of statistical downscaling methods. J Clim 30(1):203–223

Sánchez E, Solman S, Remedio ARC, Berbery H, Samuelsson P, Da Rocha RP, Mourão C, Li L, Marengo J, de Castro M, Jacob D (2015) Regional climate modelling in CLARIS-LPB: a concerted approach towards twentyfirst century projections of regional temperature and precipitation over South America. Clim Dyn 45(7):2193–2212

Schulzweida U (2017) Climate Data Operators (CDO) User Guide

Strandberg G, Bärring L, Hansson U, Jansson C, Jones C, Kjellström E, Kolax M, Kupiainen M, Nikulin G, Samuelsson P, Ullerstig A, Wang S (2014) CORDEX scenarios for Europe from the Rossby Centre regional climate model RCA4. Technical Report Report Meteorology and Climatology 116, Swedish Meteorological and Hydrological Institute, Norrköping

Tang J, Li Q, Wang S, Lee D-K, Hui P, Niu X, Gutowski WJ, Dairaku K, McGregor J, Katzfey J, Gao X, Wu J, Hong S-Y, Wang Y, Sasaki H (2016) Building asian climate change scenario by multi-regional climate models ensemble. Part I: surface air temperature. Int J Climatol 36(13):4241–4252

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. Bull Am Meteorol Soc 93(4):485–498

Teichmann C, Eggert B, Elizalde A, Haensler A, Jacob D, Kumar P, Moseley C, Pfeifer S, Rechid D, Remedio AR, Ries H, Petersen J, Preuschmann S, Raub T, Saeed F, Sieck K, Weber T (2013) How does a regional climate model modify the projected climate change signal of the driving GCM: a study over different cordex regions using remo. Atmosphere 4(2):214–236

Timbal B, McAvaney BJ (2001) An analogue-based method to downscale surface air temperature: application for Australia. Clim Dyn 17(12):947–963

Timbal B, Dufour A, McAvaney B (2003) An estimate of future climate change for western France using a statistical downscaling technique. Clim Dyn 20(7):807–823

Uppala S, Kållberg P, Simmons A, Andrae U, da Costa Bechtold V, Fiorino MJ, Haseler JG, Hernandez A, Kelly G, Li X, Onogi K, Saarinen S, Sokka N, Allan R, Andersson E, Arpe K, Balmaseda M, Beljaars A, van de Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Hólm E, Isaksen BHL, Janssen P, Jenne R, McNally A, Mahfouf J-F, Morcrette J-J, Rayner N, Saunders R, Simon P, Sterl A, Trenberth K, Untch A, Vasiljevic D, Viterbo P, Woollen J (2005) The ERA-40 re-analysis. Q J R Meteorol Soc 131:2961–3012

van Vuuren DP, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, Hurtt GC, Kram T, Krey V, Lamarque J-F, Masui T, Meinshausen M, Nakicenovic N, Smith SJ, Rose SK (2011) The representative concentration pathways: an overview. Clim Change 109(1):5–31. https://doi.org/10.1007/s10584-011-0148-z

von Storch H (2010) Climate models and modeling: an editorial essay. Wiley Interdiscip Rev Clim Change 1(3):305–310

von Storch H, Zorita E, Cubasch U (1993) Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. J Clim 6:1161–1171

Webber S, Donner SD (2016) Climate service warnings: cautions about commercializing climate science for adaptation in the developing world. Wiley Interdiscip Rev Clim Change. https://doi.org/10.1002/wcc.424

Wetterhall F, Halldin S, Xu C (2005) Statistical precipitation downscaling in central Sweden with the analogue method. J Hydrol 306(1):174–190

Wetterhall F, Halldin S, Xu C-Y (2007) Seasonality properties of four statistical-downscaling methods in central Sweden. Theor Appl Climatol 87(1):123–137

Wilks D (2010) Use of stochastic weather generators for precipitation downscaling. WIREs Clim Change 1:898–907. https://doi.org/10.1002/wcc.85

Woodley WL (1970) Precipitation results from a pyrotechnic cumulus seeding experiment. J Appl Meteorol 9(2):242–257

Yates D, Gangopadhyay S, Rajagopalan B, Strzepek K (2003) A technique for generating regional climate scenarios using a nearest-neighbor algorithm. Water Resour Res 39(7):1199

Young KC (1994) A multivariate chain model for simulating climatic parameters from daily data. J Appl Meteorol 33(6):661–671

Zorita E, von Storch H (1999) The analog method as a simple statistical downscaling technique: comparison with more complicated methods. J Clim 12(8):2474–2489

Zorita E, Hughes J, Lettenmaier D, Storch HV (1995) Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation. J Clim 8:1023–1042