RESEARCH ARTICLE

# Bias correction of daily precipitation for ungauged locations using geostatistical approaches: A case study for the CORDEX-Africa ensemble

Manuel Lorenz[1] | Jan Bliefernicht[1] | Harald Kunstmann[1,2]

[1]Institute of Geography, University of Augsburg, Augsburg, Germany

[2]Institute of Meteorology and Climate Research, Campus Alpin, Karlsruhe Institute of Technology, Garmisch-Partenkirchen, Germany

**Correspondence**
Harald Kunstmann, Institute of Geography, University of Augsburg, Alter Postweg 118, 86165 Augsburg, Email: harald.kunstmann@uni-a.de

## Abstract

Climate model simulations typically exhibit a bias, which can be corrected using statistical approaches. In this study, a geostatistical approach for bias correction of daily precipitation at ungauged locations is presented. The method utilizes a double quantile mapping with dry day correction for future periods. The transfer function of the bias correction for the ungauged locations is established using distribution functions estimated by ordinary kriging with anisotropic variograms. The methodology was applied to the daily precipitation simulations of the entire CORDEX-Africa ensemble for a study region located in the West African Sudanian Savanna. This ensemble consists of 23 regional climate models (RCM) that were run for three different future scenarios (RCP 2.6, RCP 4.5, and RCP 8.5). The evaluation of the approach for a historical 50-year period (1950–2005) showed that the method can reduce the inherent strong precipitation bias of RCM simulations, thereby reproducing the main climatological features of the observed data. Moreover, the bias correction technique preserves the climate change signal of the uncorrected RCM simulations. However, the ensemble spread is increased due to an overestimation of the rainfall probability of uncorrected RCM simulations. The application of the bias correction method to the future period (2006–2100) revealed that annual precipitation increases for most models in the near (2020–2049) and far future (2070–2099) with a mean increase of up to 165 mm$\cdot$ a$^{-1}$ (18%). An analysis of the monthly and daily time series showed a slightly delayed onset and intensification of the rainy season.

**KEYWORDS**
bias correction, climate change, CORDEX-Africa, geostatistical approaches, precipitation, quantile mapping, West Africa

# 1 | INTRODUCTION

Knowledge of a region's climatology is indispensable for the management of its water bodies, agriculture, ecosystems, or technical systems. Reliable and long time series of meteorological variables in a sufficient spatiotemporal resolution are a prerequisite to analyse the climatology and event characteristics of the region or system at hand. Furthermore, such data is required to run impact models which simulate, for example, the discharge in a catchment or the potential crop yield. For such applications, daily or higher resolutions are typically required (e.g., Bruni et al., 2015). Long time series are required so that management decisions also take extreme events or accumulated events like dry spells into account. Since the climate is projected to change worldwide, decisions makers are also confronted with adapting the management strategies to the uncertain future climate.

Physically based climate model simulations constitute a source for future climate data. General circulation models (GCMs) simulate the mass and energy fluxes in the atmosphere in a spatial resolution of currently up to 0.25° (Buizza et al., 2017). However, decision making in water resources management and many other disciplines often requires a higher spatiotemporal resolution than what the GCMs can provide. Moreover, meteorological variables such as precipitation can be highly variable in space and time and are often not well reproduced. To overcome these shortcomings, regional climate models (RCMs) are applied, which use the GCM simulations as driving boundary conditions. They are set up for a confined region of interest by nesting these models into the numerical grid of the GCM with a higher spatiotemporal resolution (Rummukainen, 2009).

Nowadays, ensembles of GCM-RCM model combinations are used to provide a set of scenarios which can then be utilized to run impact models for different design studies, for instance in hydrology or agriculture. The advantage of ensemble simulations is that the impact studies do not rely on a single simulated time series and that the uncertainty can be quantified and considered in the planning. For Africa, an ensemble of daily RCM scenarios in a spatial resolution of 0.44° has been provided by the CORDEX-Africa project (Coordinated Regional Climate Downscaling Experiment; Nikulin et al., 2012) for a historical control period (1950–2005) and future period (2006–2100). For West Africa, a set of high resolution, ensemble-based regional climate change scenarios for a historical and two future periods was provided as part of the WASCAL (West African Science Service Centre on Climate Change and Adapted Land Use) program (Dieng et al., 2018; Heinzeller et al., 2018) using high-resolution RCMs (Klein et al., 2015; Dieng et al., 2017).

Many investigations showed that key atmospheric drivers of the West African Monsoon (WAM) such as jet streams (the Tropical and African Easterly Jet) and the southwest monsoon fluxes can be reproduced by state-of-the art climate models for this challenging region (Paeth et al., 2011; Sylla et al., 2013; Klein et al., 2015). Gbobaniyi et al. (2014) also showed that specific WAM features such as the occurrence of the WAM jump, the intensification and northward shift of the Saharan Heat Low can be simulated by an ensemble of CORDEX RCM simulations. Moreover, Sylla et al. (2015) and Nikiema et al. (2017) showed that the CORDEX RCMs can improve the precipitation simulations for different climatological zones in West Africa in comparison to GCMs.

Nevertheless, an intercomparison of 10 CORDEX-Africa RCMs by Nikulin et al. (2012) showed that all models exhibit a significant systematic differences between observations and simulations (bias) in the rainy season for West Africa when compared with observations. Remarkably, ERA-Interim showed a dry bias for West Africa in this season, but the ERA-Interim driven RCMs can lead to positive and negative biases. Moreover, some of the models simulated the onset of the rainy season too early, while others have problems regarding the northward extension of the monsoon rain belt. Mascaro et al. (2015) reported similar findings for the annual precipitation amount in the Niger River basin using 18 GCM-RCM combinations of the CORDEX-Africa ensemble. Klutse et al. (2016) illustrated substantial differences between CORDEX-Africa RCMs and observations for daily rainfall characteristics such as intensity, frequency, and extreme indices.

Bias correction is often applied to climate model simulations to reduce systematic differences to the real climatology. To this end, a meteorological variable simulated by an RCM is transformed to a bias corrected value via a transfer function. There is a strong debate about the applicability of bias correction in general (Maraun, 2016) because the variables of the uncorrected RCM are physically consistent. After the bias correction, this may no longer be the case and higher aggregated variables can exhibit a stronger bias than before the correction (Ehret et al., 2012). In practice, however, bias correction is still widely applied since a biased meteorological input variable is regarded as very detrimental to the performance of subsequent impact models. The available bias correction techniques differ in how transfer function are built (Maraun, 2016). A frequently used univariate technique that reproduces the observed distribution functions is quantile mapping (e.g., Chen et al., 2013) which is closely related to histogram equalization and local intensity scaling (e.g., Berg et al., 2012). Another problem is that many bias correction approaches

provide information for observed sites (gauges) or rely on gridded observations. However, bias correction that provide point information for ungauged sites are still very limited but are needed for local impact studies. In some case studies, only a single output variable of the RCM is of interest, for example, the daily temperature or precipitation. If several variables need to be bias corrected, the correction is mostly performed individually for each variable. In recent years, more complex bias correction methods have been developed such as copula-based bias correction scheme (Laux et al., 2011; Mao et al., 2015) to generate an ensemble of values from the conditional distribution. In addition, multivariate bias correction have been proposed (Piani and Haerter, 2012; Cannon, 2016; Vrac, 2018) for a joint correction of meteorological variables.

The use of bias correction methods for data-scarce regions such as West Africa are still limited, although these datasets are in high demand by impact modellers and practitioners in agriculture and other disciplines. For West Africa, most of these techniques were mainly applied to correct global seasonal forecasts (Batté and Deéqué, 2011; Feudale and Tompkins, 2011; Oettli et al., 2011; Siegmund et al., 2015; Rauch et al., 2019) or global climate scenarios (Sultan et al., 2014). Mbaye et al. (2016) is one of the earliest studies who bias corrected regional climate scenarios driven by CMIP3 models for West Africa. Recently, Laux et al. (2021) used univariate bias correction methods for precipitation and temperature from selected CORDEX-Africa RCMs.

The objectives of this study are the development of a bias correction method for providing point information at ungauged locations and its application using a set of regional climate scenarios for a data-scarce region. The double quantile mapping (Bárdossy and Pegram, 2011) is used and extended by geostatistical approaches to estimate the distribution of daily precipitation at ungauged locations. The approach is similar to Mamalakis et al. (2017), but in addition to their study, anisotropic variograms are applied and point scale statistics are regionalized and not spatially averaged. In contrast to many other bias correction methods, this enables the provision of bias-corrected precipitation time series for ungauged locations, depending on the regionalized point scale statistics of the surrounding rainfall network (Gessner et al., 2015).

The methodology is tested for the central Sudanian Savanna of West Africa. This region is chosen due to the high importance of the WASCAL program as core research area (Yira et al., 2016; Danso et al., 2018; Bliefernicht et al., 2018; Salack et al., 2019; Berger et al., 2019) to provide bias-corrected climate scenarios for local impact studies. Another reason is the availability of long-term daily precipitation observations from rainfall gauges
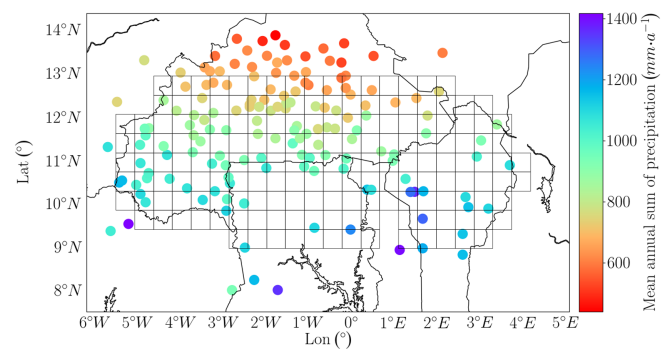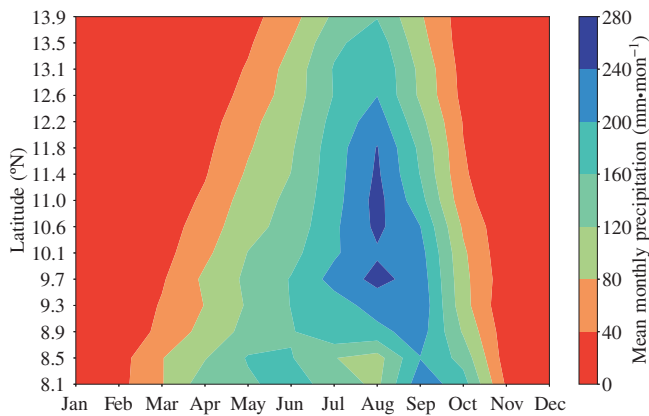


**FIGURE 1** Mean annual sum of precipitation of the West African observation data (1950–2005). The circles correspond to the rainfall network which was used for the bias correction. [Colour figure can be viewed at wileyonlinelibrary.com]

for this region (Bliefernicht et al., 2019). The analysis of the projected climate change signal is carried out for several precipitation characteristics (annual and monthly precipitation amount and the onset of the rainy season) for three projections (RCP2.6, RCP4.5, RCP8.5) and two time periods (2020–2049 and 2070–2099) using the full CORDEX-Africa ensemble based on 23 GCM-RCM model combinations. This study is therefore one of the first that uses the full CORDEX-Africa ensemble (48 RCM scenarios) for bias correction.

## 2 | STUDY REGION AND DATA SETS

The study region is shown in Figure 1. It covers regions of different countries in West Africa, primarily Burkina Faso in the North, as well as Ghana, Benin, and Togo in the Southern domain. The spatial grid stems from the CORDEX-Africa RCM ensemble with a spatial resolution of 0.44°. The centres of the grid cells are chosen in this study as ungauged sites to cover the entire region in a homogeneous way. However, the current methodology can be also applied for any irregular network. The station network is relatively dense for West Africa but it is still a magnitude lower compared with rainfall networks in Europe or North America. The rainfall sites were chosen from a novel precipitation database that has been collected and merged within the BMBF research program WASCAL from the global, regional and national databases (Bliefernicht et al., 2021). One hundred and seventy two stations are located in the proximity of the study region. Daily precipitation time series have been extracted for the period 1950–2005 for these sites. Subsets of this precipitation dataset were also used by various studies, for example, by Dieng et al. (2017) for RCM evaluation and Ascott et al. (2020) for groundwater reconstruction.

**FIGURE 2** Hovmöller diagram of mean monthly precipitation of the observation data used in this study (1950–2005). [Colour figure can be viewed at wileyonlinelibrary.com]

Figure 1 also shows that the observed mean annual sum of precipitation decreases from South (up to 1418 mm·a$^{-1}$) to North (as low as 439 mm·a$^{-1}$). Stations on the same degree of latitude have relatively similar annual sums which indicates that the statistics are anisotropic. The rainy season is also very distinct in the study region and is dominated by the West African Monsoon. From February, the southern locations already receive precipitation (Figure 2). Over the course of the rainy season, the monthly amounts increase until August. At this time, the maxima occur at the southern border of Burkina Faso. From September on, the precipitation amounts decrease quickly.

The CORDEX-Africa ensemble consists of 23 different GCM-RCM combinations (Table 1). The majority of the models from the CORDEX-Africa ensemble overestimate the annual sums of precipitation in the presented study region for the historical period (1950–2005). Annual sums of more than 2500 mm have been simulated by individual ensemble members which is a positive bias of more than 100%. Such an overestimation poses tremendous problems for subsequent impact models like crop models because it may be assumed that much more water is available than in reality. Some regions, especially Northern Ghana, have only a few measurement stations which means that the climatological statistics are unknown for many locations. This data scarcity of the measurement stations and the anisotropy of the statistics motivated a geostatistical approach to estimate the bias transfer function for ungauged locations.

## 3 | METHODS

In this study, double quantile mapping (Bárdossy and Pegram, 2011) has been chosen to bias correct the daily

precipitation time series of the CORDEX-Africa ensemble for historical and future time periods. Gudmundsson et al. (2012) found that empirical quantile mapping resulted in the best bias corrected simulations, but this approach requires complete observation time series for every location in the simulation period, which are often not available. This is also the case for the study region, and therefore the unknown point scale distribution functions for each grid cell in the study region were estimated by interpolating the parameters of the observed distribution functions to all grid cell centres in the region of interest using ordinary kriging. Separate statistics were used for each month of the rainy season. The dry season from November to February was grouped into a single season to obtain enough values for a robust estimation of the local statistics. The estimated distribution was utilized to generate so called "simulated observations." A double quantile mapping with dry day correction was then carried out for each RCM cell with the estimated CDFs. The full process of the bias correction method is illustrated in Figure 3.

### 3.1 | Dry day correction

For precipitation a correction of the frequency of wet values is necessary in most cases because RCM simulations typically exhibit more time steps with precipitation than is observed. One cause for an excessive precipitation probability is the drizzle effect. RCMs often simulate too many low intensity (high frequency) precipitation events when compared with observations (e.g., Sun et al., 2006). The probability that a grid cell is wet is also scale-dependent as larger cells are more likely to exhibit precipitation (Argüeso et al., 2013). In practice, it is often the case that there is a mismatch between the spatial resolution of the RCM and the observations. If we assume that $x_{obs}$ is observed precipitation measured by a gauge, this information corresponds to a single point in space and therefore a difference between observed precipitation probability $p_{w,obs}$ and simulated precipitation probability $p_{w,sim}$ is to be expected. Furthermore, rain gauges may miss very light precipitation amounts that are below the detection limit. Nevertheless, gauge data is commonly used as reference for bias correction—either because there is no other data available or because the impact models are usually calibrated with gauge observations.

An overestimated precipitation probability can be corrected by setting all values below a chosen threshold $\vartheta$ (e.g., 1.0 mm·d$^{-1}$) to zero. This threshold should be calculated individually for each cell so that the frequency of values above the threshold is equal to the observed precipitation probability. After calculating $p_{w,obs}$ only the $n_{sim} \cdot p_{w,obs}$ largest values of the RCM simulations will be considered as actual precipitation. $n_{sim}$ is the number of

**TABLE 1** CORDEX-Africa ensemble members which were used in this study

| Institute/research initiative | Driving model | RCM |
| --- | --- | --- |
| CCCma (Canadian Centre for Climate Modelling and Analysis) | CCCma-CanESM2 | CanRCM4 v4 |
| CLMcom (Climate Limited-area Modelling Community) | CNRM-CERFACS-CNRM-CM5 | CCLM4-8-17 v1 |
| CLMcom | ICHEC-EC-EARTH | CCLM4-8-17 v1 |
| CLMcom | MOHC-HadGEM2-ES | CCLM4-8-17 v1 |
| CLMcom | MPI-ESM | CCLM4-8-17 v1 |
| DMI (Danish Meteorological Institute) | ICHEC-EC-EARTH | HIRHAM5 v2 |
| DMI | NCC-NorESM1-M | HIRHAM5 v1 |
| KNMI (Koninklijk Nederlands Meteorologisch Instituut) | ICHEC-EC-EARTH | RACMO22T v1 |
| KNMI | MOHC-HadGEM2-ES | RACMO22T v1 |
| MPI-CSC (Max Planck Institute for Meteorology—Climate Service Center) | ICHEC-EC-EARTH | REMO2009 v1 |
| MPI-CSC | MPI-ESM | REMO2009 v1 |
| SMHI (Swedish Meteorological and Hydrological Institute) | CCCma-CanESM2 | RCA4 v1 |
| SMHI | CNRM-CERFACS-CNRM-CM5 | RCA4 v1 |
| SMHI | CSIRO-Mk3.6.0 | RCA4 v1 |
| SMHI | ICHEC-EC-EARTH | RCA4 v1 |
| SMHI | NOAA-GFDL-GFDL-ESM2M | RCA4 v1 |
| SMHI | MOHC-HadGEM2-ES | RCA4 v1 |
| SMHI | IPSL-CM5A-MR | RCA4 v1 |
| SMHI | MIROC-MIROC5 | RCA4 v1 |
| SMHI | MPI-ESM | RCA4 v1 |
| SMHI | NCC-NorESM1-M | RCA4 v1 |
| UQAM (Université du Québec à Montréal) | CCCma-CanESM2 | CRCM5 v1 |
| UQAM | MPI-ESM | CRCM5 v1 |

days of the RCM time series. The threshold $\vartheta$ is thus the value that satisfies:

$$n_{\text{sim}} \cdot p_{w,\text{obs}} = \#\{x_{\text{sim}} | x_{\text{sim}} \geq \vartheta\} \qquad (1)$$

with $x_{\text{sim}}$ the simulated precipitation value of the RCM. After the threshold has been found, all values below it are set to zero and the remaining values are shifted toward zero to allow for the fitting of a parametric distribution function. This approach has been used among others in Volosciuk *et al.* (2017) and Lafon *et al.* (2012). A correction for the rare converse case, that the wet day probability is higher in the observations than in the simulations, was developed by Themeßl *et al.* (2010) who

infilled very low intensities until the observed wet day probability was matched.

## 3.2 | Quantile mapping and double quantile mapping

Quantile mapping inverts the CDF of the observed variable $F_{\text{obs}}$ with the CDF value of the RCM simulations $F_{\text{sim}}(x_{\text{sim}})$ to generate a corrected value $x_{\text{BC}}$. Thus, the general characteristics of the RCM time series, such as when the highest values occur, remain the same but each value is mapped to its corresponding observed quantile.

**FIGURE 3** Flowchart of the geostatistical bias correction method
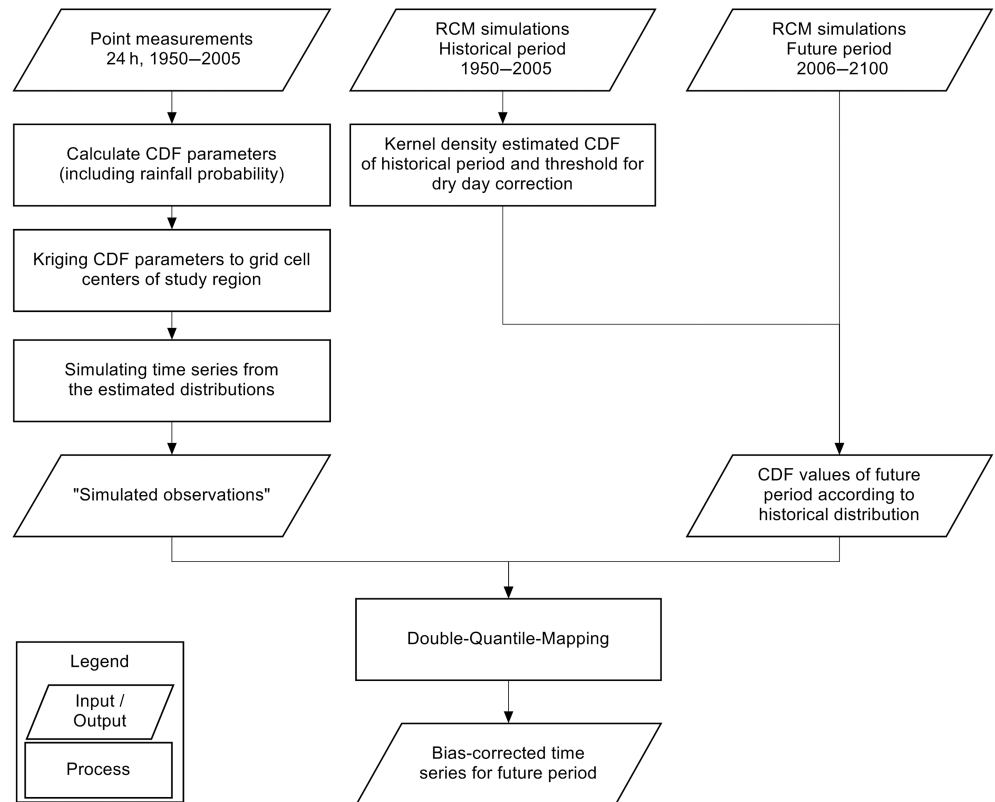


$$x_{BC} = F_{obs}^{-1}\{F_{sim}(x_{sim})\} \qquad (2)$$

$F_{obs}$ and $F_{sim}$ can either be empirical or parametric function with parameter set $\{\Theta_1, ..., \Theta_n\}$. One problem of choosing an empirical CDF is that the observed maximum cannot be exceeded. Also, the time series of observations should be as long as the RCM time series which can be circumvented by interpolating between the two values with a given rank or by sampling from the observation set until it is as large as the simulation set (Piani and Haerter, 2012). Parametric CDFs are capable of generating values larger than the observed maximum and the discrete nature of the measurements (e.g., a resolution of 0.1 mm of the measurement device) is less apparent in the bias corrected time series. Finding a function $F_{sim}$ that fits the skewed precipitation intensities simulated by RCMs can be challenging, as discussed in Gudmundsson et al. (2012).

For climate change studies, the traditional quantile mapping cannot be used directly. Fitting a distribution function $F_{sim}$ to the future period and inverting the observed distribution $F_{obs}$ with the CDF values in the future, would result in a bias corrected time series with a distribution function that is identical to the one of the observations. The only difference would be how the large and small values tend to cluster in space and time in the different time periods. The double quantile mapping method utilizes the historical CDF to calculate the
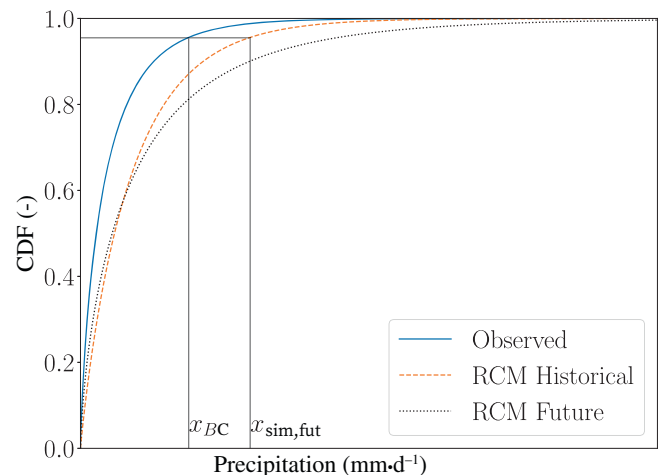


**FIGURE 4** Transformation of nonzero future precipitation amounts $x_{sim,fut}$ simulated by a regional climate model (RCM) using double quantile mapping to a bias-corrected value $x_{BC}$. The value of the RCM's historical cumulative distribution function (CDF) is used to invert the observed CDF [Colour figure can be viewed at wileyonlinelibrary.com]

CDF values of the future period. A parametric distribution function $F_{sim,hist}$ is fitted to the historical RCM time series and the CDF values of the future period are calculated with this CDF. This way, a change of the intensity distribution leads to a bias corrected time series whose distribution is no longer identical to the observed one.

$$x_{BC} = F_{obs}^{-1}(F_{sim,hist}(x_{sim,fut})) \qquad (3)$$

The double quantile mapping is illustrated in Figure 4 with artificial data. Since the CDFs of the historical and future RCM differ, the precipitation amount $x_{sim,fut}$ which is the 90%-Quantile in the future period attains a larger CDF value $F_{sim,hist}(x_{sim,fut}) = 0.95$.

## 3.3 | Estimating distribution parameters with kriging

Utilizing the closest measurement station for the centre of each grid cell to estimate the distribution parameters would lead to high uncertainties for sites that have no nearby stations. This is especially problematic in regions with a highly variable local climatology as in mountainous regions or in this case West Africa where the climatology is anisotropic and depends strongly on the degree of latitude (Figures 1 and 2). Although there are several spatial interpolation methods available for environmental variables (Li and Heap, 2013) and specifically for precipitation (Ly et al., 2013), most of these methods perform a spatial smoothing of the variable of interest leading to an underestimation of the observed variability at ungauged sites, especially for variables with a strongly skewed distribution like daily precipitation. This is the main reason why instead of a direct interpolation of daily precipitation, an interpolation of the parameters of distribution functions and a simultaneous simulation of the daily precipitation amount is performed. For instance, Mamalakis et al. (2017) interpolated the parameters of a Generalized Pareto distribution to ungauged locations to perform a bias correction.

The Kriging method is shortly explained for a single CDF parameter denoted by $\Theta$. If several parameters are required to estimate the distribution function at unmeasured locations, the following procedure has to be done for each parameter individually. Kriging estimates the unknown CDF parameter $\Theta^*$ as a linear combination of $n$ observed parameters $\Theta_i$ which each have a Kriging weight $\lambda_{K,i}$ which must fulfil the condition $\sum_{i=1}^{n} \lambda_{K,i} = 1$.

$$\Theta^* = \sum_{i=1}^{n} \lambda_{K,i} \Theta_i \qquad (4)$$

The Kriging weights $\lambda_K$ are obtained by solving the Kriging equation system that minimizes the estimation uncertainty of the unknown value. The estimation uncertainty is calculated with parametric variogram models $\widehat{\gamma}(h)$ that estimate the semi-variance of values at a given spatial separation distance $h$. Anisotropic variograms can be calculated as a linear combination of the directional variograms $\widehat{\gamma}_x$ and $\widehat{\gamma}_y$, so if the variogram value increases more strongly in one direction, the expected overall semi-variance $\widehat{\gamma}$ between the unknown point and the observation location will be higher if the two points are mainly separated along this axis.

$$\widehat{\gamma}(h) = \widehat{\gamma}_x(h_x) + \widehat{\gamma}_y(h_y) \qquad (5)$$

For instance, in West Africa precipitation pairs on the same degree of latitude have a lower semi-variance than pairs on the same degree of longitude because $\widehat{\gamma}_x(h) < \widehat{\gamma}_y(h)$. As Kriging minimizes the estimation uncertainty, a higher Kriging weight is given to neighbouring gauges that are on the same degree of latitude.

## 3.4 | Generating simulated observations

With the interpolated precipitation probability $p_w$ and distribution parameters $\Theta_1, ..., \Theta_n$, a surrogate for the unknown distribution $F_{obs}$ of each grid cell centre is provided. Afterward, stochastic simulations can be performed to obtain a time series of "simulated observations." A set of uniform random numbers $u_{sim} \sim U(0,1)$ is drawn and inverted via the parametric CDF to obtain realizations $x$. In this study, daily precipitation time series were simulated for the period 1950–2005 for each grid cell centre.

$$x = F^{-1}(u_{sim}) \qquad (6)$$

In case of precipitation it is necessary to also simulate zero precipitation amounts with a dry probability of $p_d = 1 - p_w$ which is not possible with a single parametric distribution. The overall distribution is constructed as a mixed discrete-continuous (or truncated) distribution.

$$u_{sim}(x) = \begin{cases} p_d + p_w F(x) & \text{if } x > 0 \\ \leq p_d & \text{else} \end{cases} \qquad (7)$$

The zero amounts obtain a censored CDF value $u_{sim} \leq p_d$. This means that $u_{sim}$ is unknown and can take on any value between 0 and $p_d$. To invert this truncated distribution, the random numbers are compared with the dry probability $p_d$. If a random number $u_{sim}$ is below $p_d$, the simulated value will become $x_{sim} = 0$. In the other case, the CDF value of the nonzero precipitation amounts is calculated as $u_w = \frac{u_{sim} - p_d}{1 - p_d}$ and the CDF $F$ is inverted with this rescaled value (Equation (8)).
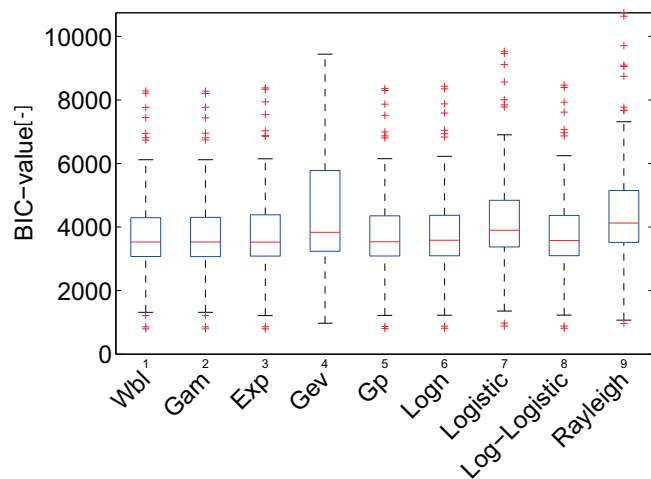
**FIGURE 5** Boxplot of Bayesian information criterion (BIC) values of nine parametric distribution functions (Weibull—Wbl, Gamma—Gam, exponential—Exp, generalized extreme value— GEV, generalized Pareto—Gp, log-normal—Logn, logistic, log— logistic and Rayleigh) fitted to daily precipitation intensities in August (1950–2005). [Colour figure can be viewed at wileyonlinelibrary.com]

$$x_{\text{sim}} = \begin{cases} F^{-1}\left(\dfrac{u_{\text{sim}} - p_d}{1 - p_d}\right) & \text{if } u > p_d \\ 0 & \text{if } u \leq p_d \end{cases} \quad (8)$$

# 4 | RESULTS

In this section a description of the results are given. At first, the outcomes of the calibration of the bias correction method are shown (Section 4.1). Afterward, the results of the bias correction are shown for the historical period (Section 4.2) and for the future periods (Section 4.3).

## 4.1 | Calibration of the bias correction method

The bias correction model depicted in Figure 3 requires a surrogate of the CDFs $F_{\text{obs}}$ for the ungauged locations and CDFs $F_{\text{sim,hist}}$ fitted to the dry day corrected RCM precipitation time series in the historical period 1950–2005. The historical period stems from the CORDEX-Africa ensemble and was chosen for the observed data. The estimation of the CDF $F_{\text{obs}}$ of daily precipitation is based on kriging the parameters of the observed CDF and the precipitation probability $p_w$ to the ungauged locations. The CDF $F_{\text{sim,hist}}$ was estimated for each cell directly with a kernel density estimation (KDE, Rosenblatt, 1956). The required

steps to calibrate the model are the selection of a parametric distribution function (Section 4.1.1), the calculation of variograms of the distribution parameters (Section 4.1.2) and the fitting of a distribution function to the RCM simulations and dry day correction (Section 4.1.3).

## 4.1.1 | Parametric distribution function of observed precipitation

According to the Bayesian information criterion (BIC; Schwarz, 1978) values of nine fitted distribution functions, there is no parametric distribution function that clearly outperforms all other functions as the differences are rather small. As an example, the BIC values for the month of August are shown in Figure 5. While the BIC values of some other distributions were slightly lower and therefore better, the exponential distribution was chosen to model $F_{\text{obs}}$ for the observed daily precipitation intensities. The exponential distribution is defined by a single parameter $\lambda_{\text{exp}}$ which is the reciprocal value of the mean wet day amount $\overline{x_w}$.

$$F(x) = 1 - e^{-\lambda_{\text{exp}} x} \quad (9)$$

Even though the exponential distribution did not result in the minimum BIC values, it was chosen for the following reasons:

1. Piani *et al.* (2010) argue that a robust transfer function with few parameters is favourable for climate change studies.
2. As the rainy season is very pronounced in West Africa, a subdivision of the year into nine seasons was made (one season for the dry season November to February and separate parameters for the other months). Calculating experimental anisotropic variograms requires splitting the sample into subsets which reduces the sample size for the calculation of the directional variograms. As the estimation of variograms and the fitting of the distribution parameters require a sound basis of observation data, a distribution function with a single parameter can be fitted more easily.
3. Quantile–quantile-plots (QQ-Plots) of the simulated daily intensities against the observed ones showed a good fit for most locations.
4. Fitting a parametric CDF with more than one parameter can result in a high variability of the parameters between neighbouring locations which leads to variograms with nearly constant semi-variances. These cause nearly equal kriging weights and
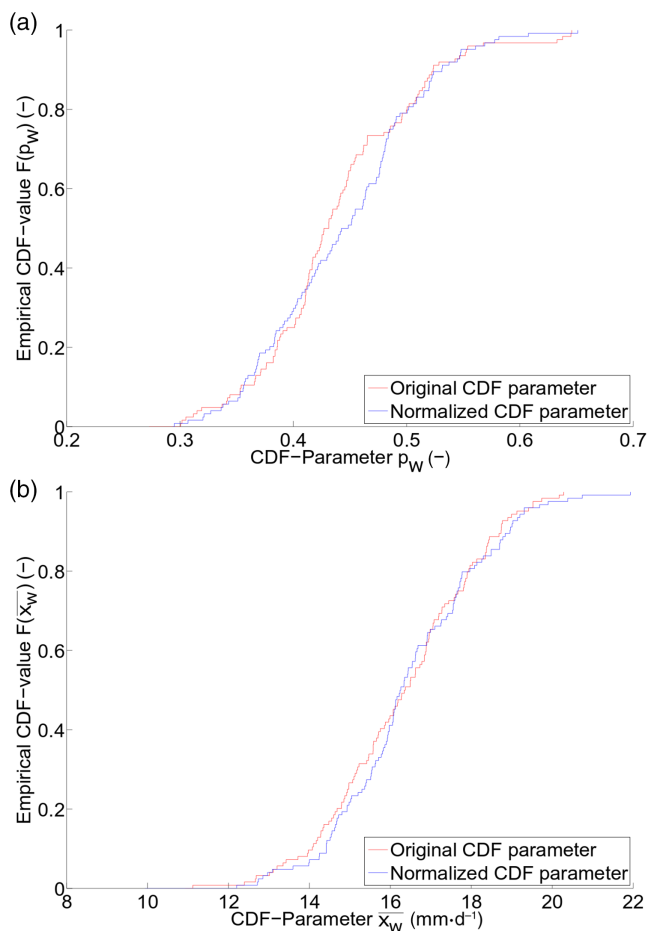
**FIGURE 6** Empirical and Gaussian distribution of daily precipitation probability $p_w$ (a) and mean wet day amount $\overline{x_w}$ (b) in August (1950–2005). [Colour figure can be viewed at wileyonlinelibrary.com]

therefore overly smooth interpolated maps and at times noninvertible kriging matrices.

### 4.1.2 | Variograms of distribution parameters

It has been discussed that the estimation of the variogram and kriging with non-normally distributed random variables can be problematic (e.g., Cressie and Hawkins, 1980) and to this end a transformation of non-normal variables like precipitation is often performed (e.g., Erdin et al., 2012). An investigation of the distribution parameters $p_w$ and $\overline{x_w}$ showed that they are approximately normally distributed for all seasons and no further transformation was performed. Figure 6 exemplifies this investigation for the month of August. The theoretical Gaussian CDFs were obtained by simulating from a Gaussian distribution with the standard deviation and mean of the observed CDF parameters within the study region.
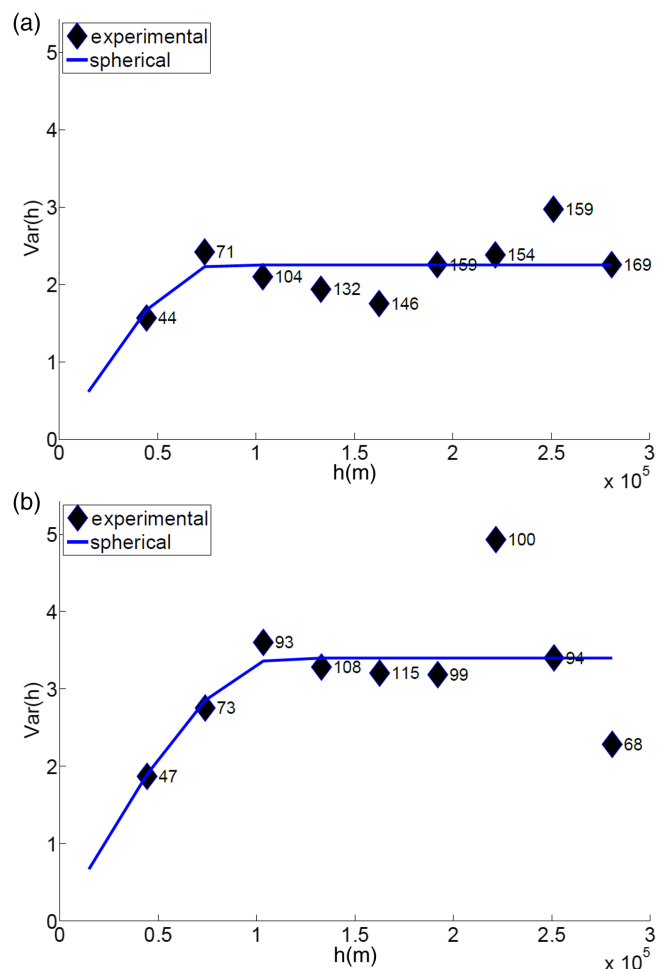


**FIGURE 7** Experimental and fitted variograms var(h) of the mean wet day amount in east–west (a) and north–south direction (b) in August (1950–2005). The numbers at the experimental variogram markers are the numbers of gauge pairs corresponding to the respective distances. [Colour figure can be viewed at wileyonlinelibrary.com]

Experimental anisotropic variograms of $\overline{x_w}$ and $p_w$ were calculated for 10 separation distances ranging from 0 to 300 km and four directions $(0°, 45°, 90°, 135°)$ to find the direction of maximum anisotropy (Krūminiene, 2006). It was found that lower semi-variances of $\overline{x_w}$ are to be expected when going from east to west (Figure 7a) as when going from north to south (Figure 7b) for both parameters. This anisotropy relates to the rainfall band that moves across West Africa from south to north during the rainy season.

### 4.1.3 | Fitting a distribution function to the RCM simulations

RCM precipitation intensities are typically highly skewed and the probability of precipitation is generally
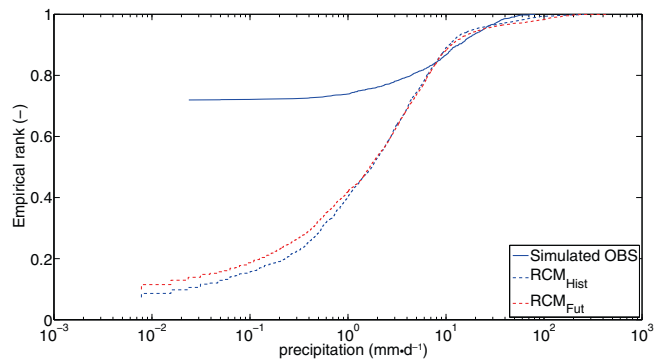
**FIGURE 8** Empirical CDFs of historical (1950–2005) and future (2006–2100, RCP8.5) precipitation of the KNMI-RACMO22-HadGEM2 model and of simulated observations in June for one cell. [Colour figure can be viewed at wileyonlinelibrary.com]

overestimated when compared with observation data. Figure 8 shows the empirical distribution of one model (KNMI-RACMO22-HadGEM2) in June for one cell. The historical and future (RCP 8.5) precipitation probability $p_w$ amounts to approximately 90%, whereas the interpolated $p_w$ is only about 28%. While the future $p_w$ is a bit lower than in the historical period, the values above $40 \ mm \cdot d^{-1}$ are higher.

In a first step, the RCM precipitation was dry day corrected. The dry day correction was calibrated by calculating the threshold $\vartheta$ for each season, grid cell and CORDEX-Africa model. The assumption was that the threshold $\vartheta$ remains constant for the future period. The remaining values were shifted toward 0 and a distribution was fitted to the positive amounts. Finding a parametric distribution $F_{sim,hist}$ that fits such highly skewed data is very problematic. Thus, a KDE-CDF was utilized. This type of CDF was chosen due to the very high extreme values that can occur in a large ensemble of RCM simulations. The data set is large enough to allow for a robust fit of the KDE-CDF on a monthly basis with the dry season November–February pooled into one season.

## 4.2 | Evaluation of the bias correction method

In this section the performance of the bias correction model is evaluated by comparing simulations from the estimated CDF $F_{obs}$ with observed data. Afterward, the bias-corrected RCM precipitation is compared with observations to evaluate the suitability of the KDE-CDF $F_{sim,hist,hist}$ to fit the historical time series of RCM precipitation. At first, a kriged map of the rainfall probability $p_w$ (Figure 9a) and the mean wet day amount $\overline{x_w}$ (Figure 9b)
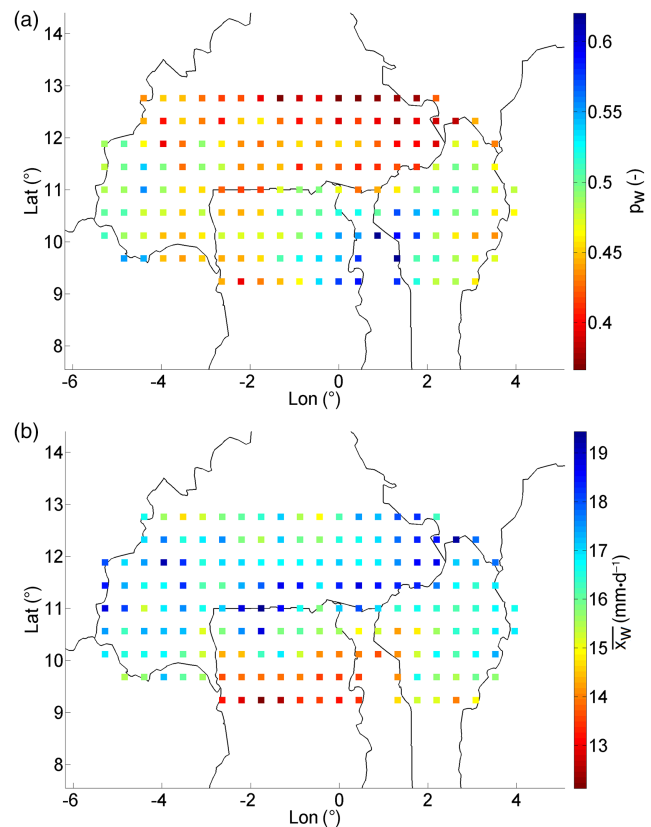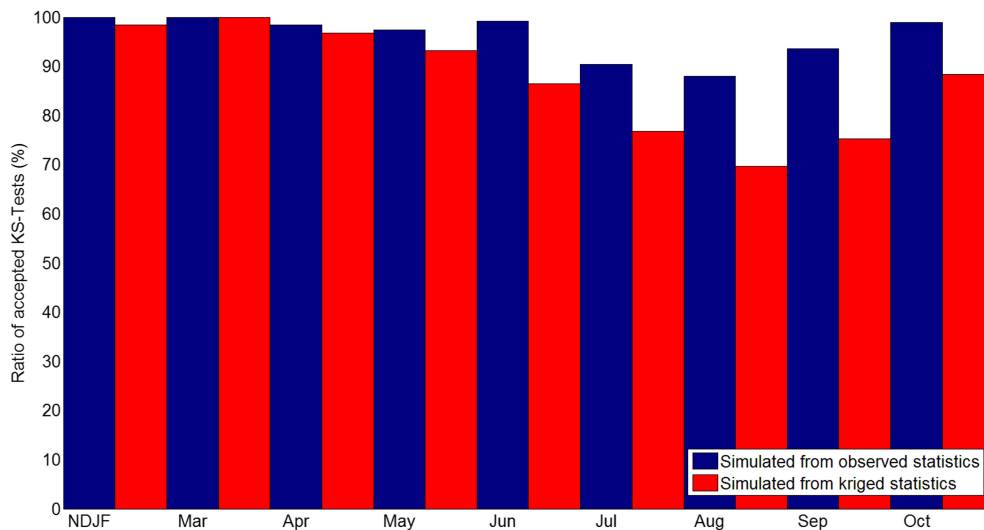


**FIGURE 9** Interpolated probability of daily precipitation $p_w$ (a) and mean wet day amount $\overline{x_w}$ (b) in August (1950–2005). [Colour figure can be viewed at wileyonlinelibrary.com]

is shown. With the fitted variograms $\widehat{\gamma}(h)$, the Kriging Equation System was built for each ungauged location. Measurement stations were considered as supporting points for the interpolation of $p_w$ if at least 500 valid daily values had been measured in the given month. For $\overline{x_w}$ it was required that at least 100 wet values had been measured at each location.

To evaluate the performance of the estimated CDF, a statistical test against a CDF fitted to observation was carried out. For each location with valid CDF parameters, the measured values of the corresponding month serve as a reference set $x_{ref}$. With the observed CDF parameters, a set of values $x_{sim,obs}$ of the same length as $x_{ref}$ was simulated with the truncated exponential distribution. Likewise, a set $x_{sim,krig}$ was simulated from the estimated CDF parameters that were interpolated from the neighbouring stations. Both simulated sets were then tested with a Kolmogorov–Smirnov test (KS test) at a significance level of $\alpha = 5\%$ against the reference set. Figure 10 presents the ratio of accepted tests for each season.

It can be seen, that the months of the rainy season do not always follow the exponential distribution and that the dry season is better represented by the exponential distribution. As Kriging tends to produce smoothed

estimates, less tests were accepted for the sets simulated from the kriged parameters. The kriged parameters' ratio of accepted KS tests lies in the range of 69.6 to 100%. Since the exponential distribution fitted to observed precipitation was not always capable of passing the KS test either, the ratios of accepted KS tests of kriged parameters was divided by the ratio of accepted KS tests with observed parameters to separate the Kriging performance from the suitability of the exponential distribution. In cases where the exponential distribution with observed parameters passed the KS test, between 79.1 and 100% of the estimated distributions also passed the test.

QQ-plots of the simulated and observed daily precipitation indicate that the data sets simulated from the exponential distribution is generally similar to the observed distribution of the nearest station. Figure 11a shows the simulated observations for Ouagadougou, Burkina Faso. The QQ-plots of some cells indicate an underestimation of the extreme values. Figure 11b is an example of one of the worst fits for a cell in the north of Ghana. This region exhibits high extreme values which the exponential distribution cannot reproduce accurately.

The CDF $F_{sim,hist}$ was fitted to the daily positive precipitation amounts $x_{RCM,hist}$ of the historical period 1950–2005 with a KDE-CDF. When the fit of a CDF is perfect, the CDF values are uniformly distributed. A regular quantile mapping was performed for the historical period and the mean monthly sums of the uncorrected and bias corrected models of the historical period were calculated and compared with the closest observed monthly sums (Figure 12). The spread of the uncorrected simulations is very large which further illustrates the need for a bias correction. As the spread of the monthly precipitation sums after the bias correction is very small, it can be concluded that the KDE-CDF fits the RCM precipitation quite well.

## 4.3 | Projected climatology of the bias corrected RCMs

A total of 48 time series simulated by different GCM-RCM model combinations were bias corrected. Five models were run for the RCP 2.6 scenario, 22 for RCP 4.5, and 21 for RCP 8.5. For an analysis of the projected climatology, the future period was split into the near future (2020–2049) and the far future (2070–2099). The historical period was chosen as 1970–1999 so that all periods are 30 years long.

The mean annual sum of precipitation is projected to change for all RCP scenarios but the magnitude and sign of change depend on the given RCP scenario and the geographical location. Figure 13a is a violin plot of the difference of the annual precipitation in the near future (2020–2049). The differences were averaged over all grid cells and the spread of the violins relates to the different models in the ensemble. To illustrate the spread of the change signal, the interquartile range (IQR) was calculated. This measure is the difference between the 25% and the 75% quantile of a dataset.
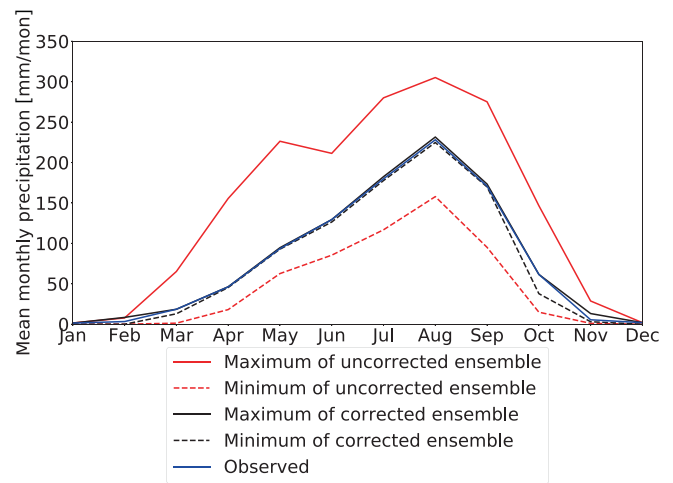
The median change amounts to $-9.4 \text{ mm} \cdot \text{a}^{-1}$ for the RCP 2.6 scenario. For the scenarios RCP 4.5 and RCP 8.5, the median change is positive and stronger ($36.6 \text{ mm} \cdot \text{a}^{-1}$ for RCP 4.5 and $78.6 \text{ mm} \cdot \text{a}^{-1}$ for RCP 8.5). For the far future (2070–2099), similar differences were calculated (Figure 13b). In the RCP 2.6 scenario, a median decrease of $-11.9 \text{ mm} \cdot \text{a}^{-1}$ is expected. For the other scenarios, the difference is again positive ($50.7 \text{ mm} \cdot \text{a}^{-1}$ for RCP 4.5 and $165.1 \text{ mm} \cdot \text{a}^{-1}$ for RCP 8.5). In contrast to the near future, the spread is larger and some models project very large differences of more than $400 \text{ mm} \cdot \text{a}^{-1}$.

The bias correction resulted in larger spreads of the average annual sums which is related to a change of the

**(a)**

**(b)**
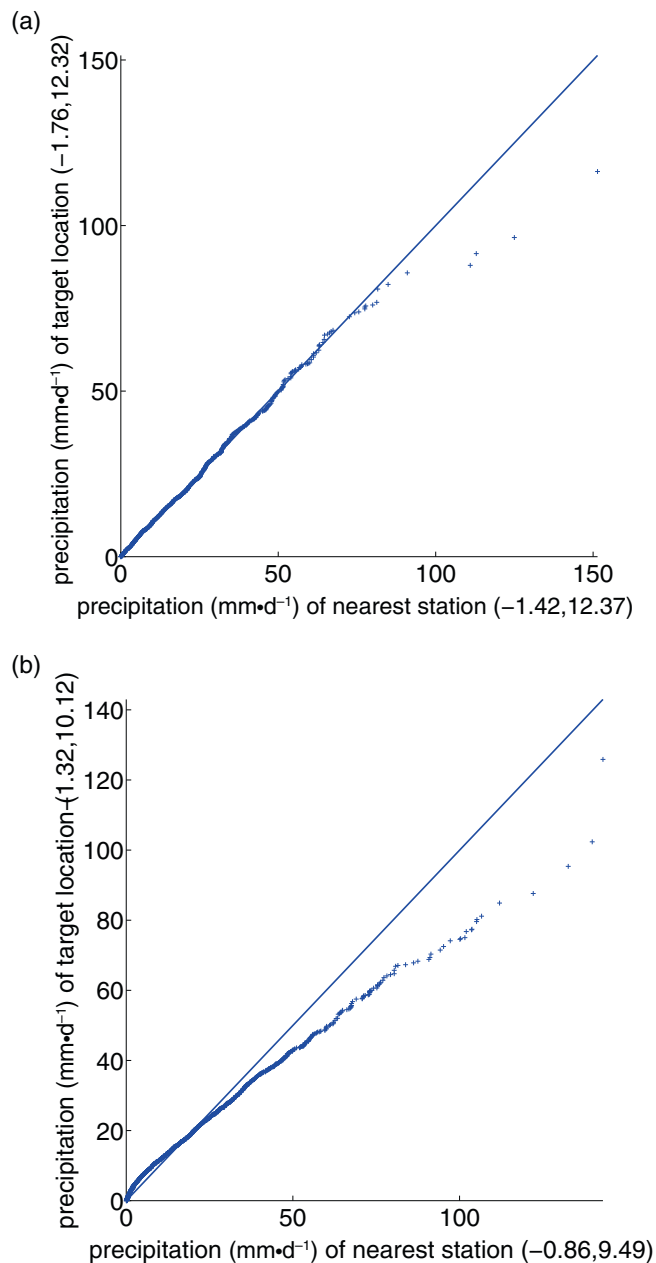
FIGURE 11    QQ-plot of simulated and observed daily precipitation for Ouagadougou, Burkina Faso (a) and for a cell in the north of Ghana (b) (1950–2005). [Colour figure can be viewed at wileyonlinelibrary.com]



FIGURE 12    Maximum and minimum mean monthly sum of precipitation of uncorrected and corrected historical CORDEX-Africa RCM ensemble simulations and monthly sum of nearest observations (1950–2005). [Colour figure can be viewed at wileyonlinelibrary.com]

distribution of positive amounts and of the dry days. As was shown in Figure 8, the distribution of the simulations typically have too many nonzero amounts which were removed with the dry day correction method. The remaining positive were shifted toward 0 and transforming them via $F_{obs}$ resulted in the larger spread of the annual sums. The differences between the bias corrected historical and future climatology are caused by a change in the distribution. Since the dry day correction thresholds $\vartheta$ can be very large, a change of its nonexceedance probability in the future period introduces a change of

the bias corrected distribution that cannot be inferred from the raw simulations. Therefore the large number of uncertain low precipitation amounts can have a strong influence on the simulations' statistics. The assumption is that the threshold $\vartheta$ remains constant for the future period. For the presented model, location and season, the observed precipitation probability is $p_w = 28.1\%$ with corresponding $\vartheta = 4.38 \text{ mm} \cdot \text{d}^{-1}$. Applying this $\vartheta$ to the future period changes $p_w$ to 29.2%. Thus, the precipitation probability is slightly higher in the bias corrected RCP 8.5 time series even though the wet day probability ($\geq 0 \text{ mm} \cdot \text{d}^{-1}$) in the uncorrected future period is lower than in the uncorrected historical period (Figure 8). This is caused by more values exceeding the threshold $\vartheta$ in the future period. It can be seen that the future CDF in Figure 8 intersects the historical CDF at around $1 \text{ mm} \cdot \text{d}^{-1}$. A change of the occurrence of these uncertain low values cannot always be utilized to estimate a trend. Therefore, $\vartheta$ was assumed to remain constant for the future period as in Pierce *et al.* (2015). Estimating a trend of the precipitation probability $p_w$ in the raw RCM simulations to adjust this threshold for future conditions is not always feasible as there might be no discernible trend because an RCM might have no zeros at all (e.g., the Hirham-EC-EARTH model). In such a case, $p_w$ is 100% for the historical and future period and a change in $p_w$ cannot be estimated. Calculating $\vartheta$ for the historical period and assuming its validity for future conditions did lead to the most stable results. Because of the strong bias of most GCM-RCM combinations, very high thresholds ($\vartheta > 15 \text{ mm} \cdot \text{d}^{-1}$) were necessary for certain months and locations. Similar findings were reported by Polade *et al.*

(a)



(b)
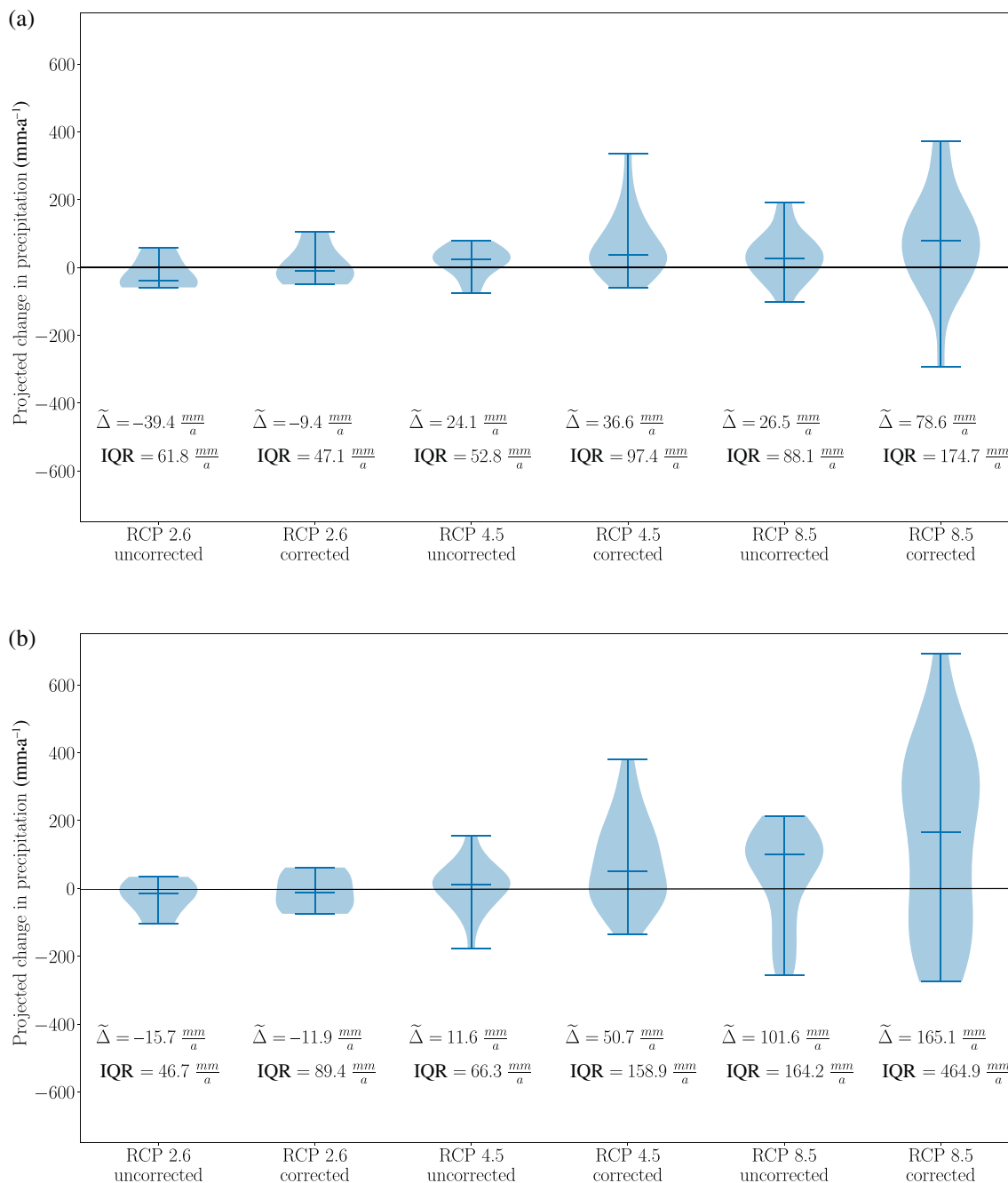


**FIGURE 13** Violin plot of change of future annual precipitation for the three uncorrected and corrected RCP scenarios in the near future 2020–2049 (a) and the far future 2070–2099 (b) in comparison to the historical period 1970–1999. $\widetilde{\Delta}$, median change; IQR, interquartile range. [Colour figure can be viewed at wileyonlinelibrary.com]

(2014) who separated the contribution of changes in the number of wet days ($>1\ mm \cdot d^{-1}$) from the changes in precipitation amounts on wet days to the annual sum of the CMIP5 (Coupled Model Intercomparison Project Phase 5) ensemble. Between $40^{\circ}$ S and $40^{\circ}$ N, changes in the wet day frequency contributed more than 50% to the change of annual precipitation.

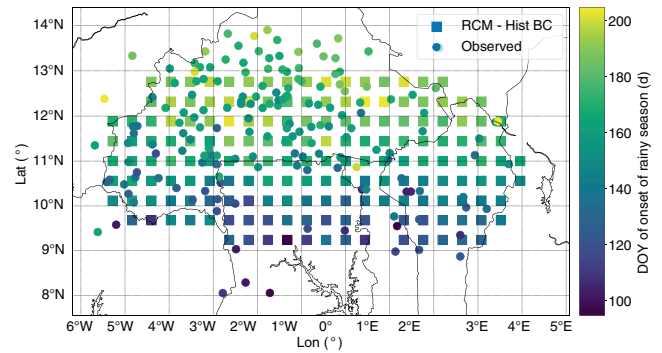The majority of the models project increasing mean monthly sums for the RCP 8.5 scenario when compared

with the historical period 1970–1999 (Figure 14). As the ensemble spread is very large, there exist models however which project lower mean monthly sums for the near and far future.

To assess how the climatology is projected to change for agriculturally relevant rainfall indices like the onset of the rainy season (ORS), a statistical approach proposed by Laux *et al.* (2008) was used for the calculation of the ORS dates. As an example for the historical period, a

**(a)**

**(b)**
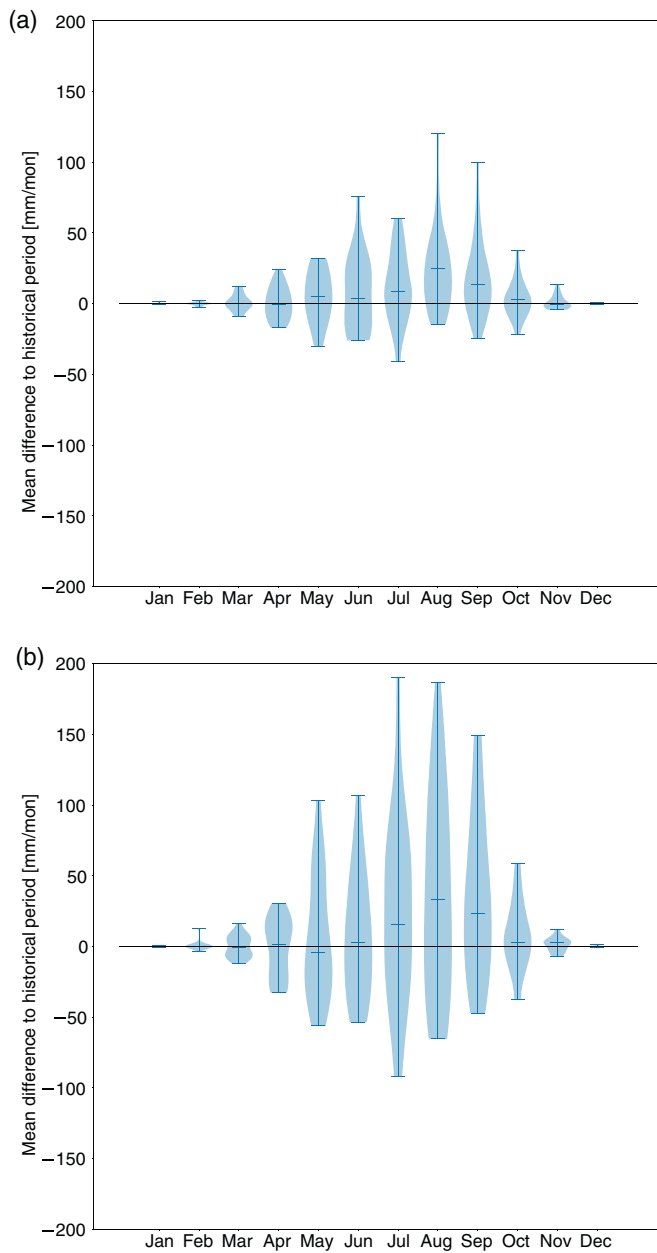
**FIGURE 14** Difference of mean monthly sums of precipitation in the bias corrected RCP 8.5 scenario in the near future 2020–2049 (a) and the far future 2070–2099 (b) compared with historical period (1970–1999). [Colour figure can be viewed at wileyonlinelibrary.com]



**FIGURE 15** Map of observed (circles) and bias corrected (squares) onset dates of the best model in the historical period 1970–1999. DOY, day of the year. [Colour figure can be viewed at wileyonlinelibrary.com]



**(a)**

**(b)**
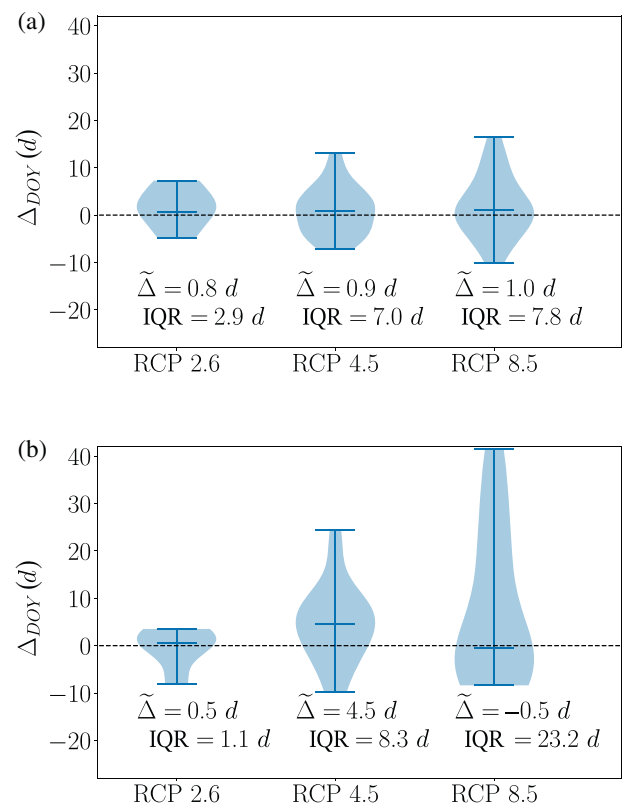
**FIGURE 16** Violin plots of the onset changes ($\Delta_{DOY}$) in terms of number of days of the bias corrected near future 2020–2049 (a) and far future 2070–2099 (b) in comparison to the historical period 1970–1999. $\widetilde{\Delta}$, median change; IQR, interquartile range. [Colour figure can be viewed at wileyonlinelibrary.com]

map of the ORS dates is given for the corrected RCM with the best performance (Figure 15). The general ORS pattern is well met for this region. In the southern parts, the rainy season starts earlier in the year (approximately in mid-April) as the monsoon rain belt moves from south to north. In the northern parts, the rainy season starts about 2 months later in mid-June. To investigate whether the rainy season is likely to start earlier or later in the future, the differences of the ORS dates between the historical and future periods were calculated for all models.

Depending on the RCP scenario and future period, the rainy season is projected to begin 0–5 days later in the future (Figure 16) even though the annual sum of precipitation is projected to increase (Figure 13). While the median differences for the onset dates of each RCP scenario are rather small, larger changes can be observed for individual models (between 10 days earlier and 41 days later). Note that the different spreads of the violins is

partially related to the number of available models in the different RCP scenarios (5 for RCP 2.6, 22 for RCP 4.5, and 21 for RCP 8.5).

# 5 | DISCUSSION

From the methodological point of view, the proposed bias correction method based on double quantile mapping has several advantages over common bias correction methods of climate projections, as it corrects the entire distribution of daily precipitation including the frequency of wet values and preserves the signal of climate models. Due to the geostatistical approach and the stochastic simulation, a bias corrected time series can be provided for an ungauged site based on regionalized point-scale statistics of the surrounding rainfall network, while the variability of the target variable is maintained. The estimation of distribution parameters can also be used for further models like stochastic weather generators (e.g., Wilks, 2009) or other statistical approaches used in climate sciences (e.g., Bárdossy and Pegram, 2016; Lorenz et al., 2018). The approach can also be easily transferred to other meteorological variables and climate regions if corresponding observations and GCM/RCM simulations are available. Moreover, instead of interpolation of point scale statistics, the approach can simply be extended using block kriging if bias corrected time series on regular grid for distributed impact models in hydrology or agriculture are needed.

We also showed that the bias correction eliminates the systematic differences between the RCM ensemble and the observational references for the historical period as shown in Figure 12 for the seasonal distribution of monthly precipitation, for instance. An analysis of the corrected climate change signal using the full CORDEX-Africa ensemble revealed that the central Sudanian Savanna of West Africa will most likely experience higher rainfall amounts in future and a slight delay of the onset. This is consistent to other investigations in this region using a limited RCM ensemble (Sylla et al., 2015; Badou et al., 2018; Heinzeller et al., 2018). Moreover, the sign of the ensemble signal is preserved and is therefore consistent to the uncorrected RCM simulations. However, more detailed investigations are needed to explore how certain daily precipitation characteristics (e.g., extremes, dry spells) are corrected and projected to change in this region and how the proposed method behaves in other climatological zones and seasons as shown in other studies (Oguntunde and Abiodu, 2013; Sylla et al., 2015; Laux et al., 2021).

We also illustrated that the ensemble spread is increased for the future period although the bias between RCM simulations and observational reference is eliminated. As explained in detail in Section 4.3, this is mainly related to the change of the distribution function of the nonzero precipitation amounts and the dry days and the challenges to account the dry day correction. It is therefore partly the result of the large statistical differences of several RCMs to the observations. Another problem is that tested distribution functions are not perfectly matching to the observed statistics. Thus, we cannot completely eliminate the biases for the ungauged sites and further uncertainties are introduced due to this problem. Thus, the elimination of less reliable RCMs simulations from the CORDEX-Africa ensemble and the further development of the current approach by using better fitted distribution functions would reduce the ensemble spread.

Moreover, the proposed approach should be only viewed as straightforward reference approach for a point-based bias correction of daily precipitation. There are more sophisticated approaches for a bias correction of regional climate projections (e.g., Bárdossy and Pegram, 2012), but they have not yet been applied on an ensemble basis or need to be advanced for point estimates. An important issue is the integration of large-scale atmospheric circulation patterns (CP; Huth et al., 2018) in the bias correction approach as shown in (Bárdossy and Pegram (2011)). In this case, the observed bias is determined for each CP and not for each month as done in this study. This approach would better capture the WAM dynamics on different spatio-temporal scales. Guèye et al. (2012) and Moron et al. (2018) showed that reliable atmospheric CPs can be also determined for this challenging region. In addition, different parametric CDFs can be selected as was done by Mamalakis et al. (2017) for the correction of daily precipitation. Moreover, a set of fitted distributions can be used to study the inherent uncertainty of the bias correction approach in comparison to the GCM/RCM spread. As pointed out by Laux et al. (2021) for the WAM region, the uncertainty of bias correction can be in the range of the GCM/RCM spread. However, the selection of other distribution functions or even an ensemble of distribution function to study this issue is not straightforward due to the reasons given in the Section 4.1.1. A further important task is to provide better estimates for the distribution function of daily precipitation at the ungauged site as in Mosthaf and Bárdossy (2017) since all fitted distribution functions have limitations (as illustrated in Figures 10 and 11 exemplary for the exponential distribution). This would reduce the uncertainty of this approach and would provide more robust estimates of future climate. Moreover, the choice of the interpolation methods for

estimating the distribution parameters can be investigated in more detail since several other interpolation methods are available for precipitation (Ly *et al.*, 2013). In addition, improved interpolation method for a direct interpolation of the variable of interest can be tested that can better capture the observed variability at ungauged sites. This task plays important role if the proposed approach is used for meteorological variables that are less skewed like daily temperature.

Another important aspect is the analysis how reliably the different CORDEX GCM/RCM-combinations can simulate the atmospheric processes for this region. This study indicated several limitations such as a strong over-estimation of the annual precipitation amount for certain RCMs (Section 2) and a large ensemble spread for the historical period (Figure 12). The reasons for the RCM biases can be manifold. According to Teutschbein and Seibert (2013) there are several error sources of RCM simulations: numerical effects, computationally limited spatiotemporal resolution, uncertainties of the initial and boundary conditions and using a single parameter to represent a certain property of a grid cell like, for example, the land use. Moreover, RCMs use parametrization schemes to solve small-scale processes and this can have a strong impact on precipitation simulations for the West African Monsoon as shown in Klein *et al.* (2015). Another reason for the bias is also the high observational uncertainty for this region as shown in Sylla *et al.* (2015), who used several standard global precipitation products as reference for comparison with RCM simulations. Moreover, the inherent bias of GCMs is another potential source of RCM bias. However, the exact bias reasons for the different RCMs are not investigated in this study. Thus, a more in-depth analysis of the GCM/RCM is needed to better relate the identified precipitation biases to the different sources of RCM biases and to filter less reliable RCM simulations for this region as, for example, done for the Amazon region for CMIP5 models by Baker *et al.* (2021). In addition, the CPs can be also used as tool for analysing how well large-scale atmospheric processes can be simulated by climate models in this region. These different steps would allow a more process-based bias correction as demanded, for example, by Maraun *et al.* (2017) and could ultimately provide more reliable precipitation scenarios for this region.

# 6 | SUMMARY AND CONCLUSIONS

A geostatistical bias correction technique that estimates the distribution functions of daily precipitation of ungauged locations was developed. The CDF parameters were interpolated from station locations using kriging and used to generate so-called "simulated observations" to perform a double quantile mapping of the precipitation time series of the CORDEX-Africa ensemble for a study region in West Africa. A comparison of the estimated distributions with actual observations showed that the main climatological characteristics are reproduced. The kriging procedure is flexible since it estimates the CDF parameters of ungauged locations as a function of distance to gauges.

An analysis of the projected climate change of the full CORDEX-Africa ensemble after the bias correction revealed that the central Sudanian Savanna of West Africa will most likely experience higher rainfall amounts in the period 2006–2100. The annual precipitation amounts will increase up to 79 mm·a$^{-1}$ for the near future 2020–2049 and 165.1 mm·a$^{-1}$ for the far future 2070–2099 under RCP 8.5. The sign of change is consistent with the uncorrected simulations but since the rainfall probability is generally too high in the models and the majority of the nonzero precipitation amounts is underestimated, the spread and uncertainty of the projected change increases after bias correction. However, the ORS date is not projected to change much.

The bias corrected precipitation ensemble of this study has been made available for future research applications and constitutes an important source of information for studying the impact of climate change in hydrology and agriculture in the Sudanian Savanna of West Africa. The full data set (raw and corrected) can be accessed via the WASCAL Scientific Research Data Catalogue using the key words "precipitation" and "CORDEX-Africa." This data portal also provides a gridded daily observed precipitation dataset (1970–2010) for the study region of this work and further meteorological datasets relevant for West Africa.

# AUTHOR CONTRIBUTIONS
**Harald Kunstmann:** Conceptualization; funding acquisition; supervision. **Manuel Lorenz:** Conceptualisation; methodological development; data analysis; writing and editing. **Jan Bliefernicht:** Conceptualisation; data provision; writing; funding acquisition; reviewing and editing.

## ORCID

*Manuel Lorenz* https://orcid.org/0000-0002-9914-5546
*Jan Bliefernicht* https://orcid.org/0000-0002-8591-6231
*Harald Kunstmann* https://orcid.org/0000-0001-9573-1743

## REFERENCES

Argüeso, D., Evans, J.P. and Fita, L. (2013) Precipitation bias correction of very high resolution regional climate models. *Hydrology and Earth System Sciences*, 17(11), 4379–4388. https://doi.org/10.5194/hess-17-4379-2013.

Ascott, M.J., Macdonald, D.M.J., Black, E., Verhoef, A., Nakohoun, P., Tirogo, J., Sandwidi, W., Bliefernicht, J., Sorensen, J. and Bossa, A.Y. (2020) In situ observations and lumped parameter model reconstructions reveal intra-annual to multidecadal variability in groundwater levels in sub-Saharan Africa. *Water Resources Research*, 56, e2020WR028056. https://doi.org/10.1029/2020WR028056.

Badou, D.F., Diekkrüger, B., Kapangaziwiri, E., Mbaye, M.L., Yira, Y., Lawin, E.A., Oyerinde, G.T. and Afouda, A. (2018) Modelling blue and green water availability under climate change in the Beninese Basin of the Niger River basin. *Hydrological Processes*, 32, 2526–2542. https://doi.org/10.1088/1748-9326/abfb2e.

Baker, J.C.A., Garcia-Carreras, L., Buermann, W., Castilho de Souza, D., Marsham, J.H., Kubota, P.Y., Gloor, M., Coelho, C. A.S. and Spracklen, D.V. (2021) Robust Amazon precipitation projections in climate models that capture realistic land–atmosphere interactions. *Environmental Research Letters*, 16, 7. https://doi.org/10.1088/1748-9326/abfb2e.

Bárdossy, A. and Pegram, G. (2011) Downscaling precipitation using regional climate models and circulation patterns toward hydrology. *Water Resources Research*, 47, 4. https://doi.org/10.1029/2010wr009689.

Bárdossy, A. and Pegram, G. (2012) Multiscale spatial recorrelation of RCM precipitation to produce unbiased climate change scenarios over large areas and small. *Water Resources Research*, 48, 9. https://doi.org/10.1029/2011wr011524.

Bárdossy, A. and Pegram, G. (2016) Space-time conditional disaggregation of precipitation at high resolution via simulation. *Water Resources Research*, 52, 2–937. https://doi.org/10.1002/2015wr018037.

Batté, L. and Deéqué, M. (2011) Seasonal predictions of precipitation over Africa using coupled ocean-atmosphere general circulation models: skill of the ENSEMBLES project multimodel ensemble forecasts. *Tellus A: Dynamic Meteorology and Oceanography*, 63(2), 283–299. https://doi.org/10.1111/j.1600-0870.2010.00493.x.

Berg, P., Wagner, S., Kunstmann, H. and Schädler, G. (2012) High resolution regional climate model simulations for Germany: part I—validation. *Climate Dynamics*, 40(1-2), 401–414. https://doi.org/10.1007/s00382-012-1508-8.

Berger, S., Bliefernicht, J., Linstädter, A., Canak, K., Guug, S., Heinzeller, D., Hingerl, L., Mauder, M., Neidl, F., Quansah, E., Salack, S., Steinbrecher, R. and Kunstmann, H. (2019) The impact of rain events on CO2 emissions from contrasting land use systems in semi-arid west African savannas. *Science of the Total Environment*, 647, 1478–1489. https://doi.org/10.1016/j.scitotenv.2018.07.39.

Bliefernicht, J., Salack, S., Guug, S., Hingerl, L., Heinzeller, D., Hingerl, L., Mauder, M., Steinbrecher, R., Steup, G., Bossa, A.Y., Waongo, M., Quansah, E., Balogun, A.A., Yira, Y., Arnault, J., Wagner, S., Klein, C., Gessner, U., Knauer, K., Straub, A., Schōnrock, R., Kunkel, R., Okogbue, E.C., Rogmann, A., Neidl, F., Jahn, C., Diekkrüger, Aduna, A., Barry, B. and Kunstmann, H. (2018) The WASCAL hydrometeorological observatory in the Sudan Savanna of Burkina Faso and Ghana. *Vadose Zone Journal*, 20, 180065. https://doi.org/10.2136/vzj2018.03.0065.

Bliefernicht, J., Salack, S., Waongo, M., Annor, T., Laux, P. and Kunstmann, H. (2021) Towards a historical precipitation database for West Africa: overview, quality control and harmonization. *International Journal of Climatology*, 2021, 1–23. https://doi.org/10.1002/joc.7467.

Bliefernicht, J., Waongo, M., Salack, S., Seidel, J., Laux, P. and Kunstmann, H. (2019) Quality and value of seasonal precipitation forecasts issued by the West African regional climate outlook forum. *Journal of Applied Meteorology and Climatology*, 58, 621–642. https://doi.org/10.1175/jamc-d-18-0066.1.

Bruni, G., Reinoso, R., van de Giesen, N.C., Clemens, F.H.L.R. and ten Veldhuis, J.A.E. (2015) On the sensitivity of urban hydrodynamic modelling to rainfall spatial and temporal resolution. *Hydrology and Earth System Sciences*, 19(2), 691–709. https://doi.org/10.5194/hess-19-691-2015.

Buizza, R., Bidlot, J.-R., Janousek, M., Keeley, S., Mogensen, K. and Richardson, D. (2017). New IFS cycle brings sea-ice coupling and higher ocean resolution. ECMWF Newsletter, 150 - Winter 2016/17, 14–17. https://doi.org/10.21957/xbov3ybily

Cannon, A.J. (2016) Multivariate bias correction of climate model output: matching marginal distributions and Intervariable dependence structure. *Journal of Climate*, 29(19), 7045–7064. https://doi.org/10.1175/jcli-d-15-0679.1.

Chen, J., Brissette, F.P., Chaumont, D. and Braun, M. (2013) Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water Resources Research*, 49(7), 4187–4205. https://doi.org/10.1002/wrcr.20331.

Cressie, N. and Hawkins, D.M. (1980) Robust estimation of the variogram: I. *Journal of the International Association for Mathematical Geology*, 12(2), 115–125. https://doi.org/10.1007/BF01035243.

Danso, I., Gaiser, T., Webber, H., Naab, J. and Ewert, F. (2018) Response of maize to different nitrogen application rates and

tillage practices under two slope positions in the face of current climate variability in the Sudan Savanna of West Africa. In: Saito, O., Kranjac-Berisavljevic, G., Takeuchi, K., A. and Gyasi, E. (Eds.) *Strategies for Building Resilience against Climate and Ecosystem Changes in Sub-Saharan Africa. Science for Sustainable Societies*. Singapore: Springer.

Dieng, D., Laux, P., Smiatek, G., Heinzeller, D., Bliefernicht, J., Sarr, A., Gaye, A.T. and Kunstmann, H. (2018) Performance analysis and projected changes of Agroclimatological indices across West Africa based on high-resolution regional climate model simulations. *Journal of Geophysical Research: Atmospheres*, 123, 7950–7973. https://doi.org/10.1029/2018jd028536.

Dieng, D., Smiatek, G., Bliefernicht, J., Heinzeller, D., Sarr, A., Gaye, A.T. and Kunstmann, H. (2017) Evaluation of the COSMO-CLM high-resolution climate simulations over West Africa. *Journal of Geophysical Research: Atmospheres*, 122, 1437–1455. https://doi.org/10.1002/2016jd025457.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K. and Liebert, J. (2012) HESS opinions: "Should we apply bias correction to global and regional climate model data?". *Hydrology and Earth System Sciences*, 16(9), 3391–3404. https://doi.org/10.5194/hess-16-3391-2012.

Erdin, R., Frei, C. and Künsch, H.R. (2012) Data transformation and uncertainty in geostatistical combination of radar and rain gauges. *Journal of Hydrometeorology*, 13(4), 1332–1346. https://doi.org/10.1175/jhm-d-11-096.1.

Feudale, L. and Tompkins, A.M. (2011) A simple bias correction technique for modeled monsoon precipitation applied to West Africa. *Geophysical Research Letters*, 38, L03803. https://doi.org/10.1029/2010GL045909.

Gbobaniyi, E., Sarr, A., Sylla, M.B., Diallo, I., Lennard, C., Dosio, A., Diedhiou, A., Kamga, A., Klutse, N.A.B., Hewitson, B., Nikulin, G. and Lamptey, B. (2014) Climatology, annual cycle and interannual variability of precipitation and temperature in CORDEX simulations over West Africa. *International Journal of Climatology*, 34, 2241–2257. https://doi.org/10.1002/joc.3834.

Gessner, U., Knauer, K., Kuenzer, C. and Dech, S. (2015) Land surface phenology in a West African Savanna: Impact of land use, land cover and fire. In: Kuenzer, C., Dech, S. and Wagner, W. (Eds.) *Remote Sensing Time Series. Remote Sensing and Digital Image Processing*, vol 22. Cham: Springer, pp. 202-223.

Gudmundsson, L., Bremnes, J.B., Haugen, J.E. and Engen-Skaugen, T. (2012) Technical note: downscaling RCM precipitation to the station scale using statistical transformations—a comparison of methods. *Hydrology and Earth System Sciences*, 16(9), 3383–3390. https://doi.org/10.5194/hess-16-3383-2012.

Guèye, A.K., Janicot, S., Sultan, B., Thiria, S., Niang, A., Sawadogo, S. and Diongue-Niang, A. (2012) Weather regimes over Senegal during the summer monsoon season using self-organizing maps and hierarchical ascendant classification. Part I: synoptic time scale. *Climate Dynamics*, 36, 1–18. https://doi.org/10.1007/s00382-010-0782-6.

Heinzeller, D., Dieng, D., Smiatek, G., Olusegun, C., Klein, C., Hamann, I., Salack, S., Bliefernicht, J. and Kunstmann, H. (2018) The WASCAL high-resolution regional climate simulation ensemble for West Africa: concept, dissemination and assessment. *Earth System Science Data*, 10(2), 815–835. https://doi.org/10.5194/essd-10-815-2018.

Huth, R., Beck, C., Philipp, A., Demuzere, M., Ustrnul, Z., Cahynová, M., Kyselý, J. and Tveito, O.E. (2018) Classifications of atmospheric circulation patterns. *Annals of the New York Academy of Sciences*, 1146(1), 105–152. https://doi.org/10.1196/annals.1446.019.

Klein, C., Heinzeller, D., Bliefernicht, J. and Kunstmann, H. (2015) Variability of west African monsoon patterns generated by a WRF multi-physics ensemble. *Climate Dynamics*, 45, 2733–2755. https://doi.org/10.1007/s00382-015-2505-5.

Klutse, N.A.B., Sylla, M.B., Diallo, I., Sarr, A., Dosio, A., Diedhiou, A., Kamga, A., Lamptey, B., Ali, A., Gbobaniyi, E.O., Owusu, K., Lennard, C., Hewitson, B., Nikulin, G., Panitz, H.J. and Büchner, M. (2016) Daily characteristics of west African summer monsoon precipitation in CORDEX simulations. *Theoretical and Applied Climatology*, 123, 369–386. https://doi.org/10.1007/s00704-014-1352-3.

Krūminiene, I. (2006) Analysis of anisotropic variogram models for prediction of the Curonian lagoon data. *Mathematical Modelling and Analysis*, 11(1), 73–86.

Lafon, T., Dadson, S., Buys, G. and Prudhomme, C. (2012) Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. *International Journal of Climatology*, 33, 1367–1381. https://doi.org/10.1002/joc.3518.

Laux, P., Kunstmann, H. and Bárdossy, A. (2008) Predicting the regional onset of the rainy season in West Africa. *International Journal of Climatology*, 28(3), 329–342. https://doi.org/10.1002/joc.1542.

Laux, P., Rötter, R.P., Webber, H., Dieng, D., Rahimi, J., Wei, J., Faye, B., Srivastava, A.K., Bliefernicht, J., Adeyeri, O., Arnault, J. and Kunstmann, H. (2021) To bias correct or not to bias correct? An agricultural impact modelers' perspective on regional climate model data. *Agricultural and Forest Meteorology*, 304, 304–305, 108406–305. https://doi.org/10.1016/j.agrformet.2021.108406.

Laux, P., Vogl, S., Qiu, W., Knoche, H.R. and Kunstmann, H. (2011) Copula-based statistical refinement of precipitation in RCM simulations over complex terrain. *Hydrology and Earth System Science*, 15(7), 2401–2419. https://doi.org/10.5194/hess-15-2401-2011.

Li, J. and Heap, A.D. (2013) Spatial interpolation methods applied in the environmental sciences: a review. *Environmental Modelling & Software*, 53, 173–189.

Lorenz, M., Bliefernicht, J., Haese, B. and Kunstmann, H. (2018) Copula-based downscaling of daily precipitation fields. *Hydrological Processes*, 32(23), 3479–3494. https://doi.org/10.1002/hyp.13271.

Ly, S., Charles, C. and Degré, A. (2013) Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale. A review. *Biotechnology, Agronomy and Society and Environment*, 17(2), 392–406.

Mamalakis, A., Langousis, A., Deidda, R. and Marrocu, M. (2017) A parametric approach for simultaneous bias correction and high-resolution downscaling of climate model rainfall. *Water Resources Research*, 53(3), 2149–2170. https://doi.org/10.1002/2016wr019578.

Mao, G., Vogl, S., Laux, P., Wagner, S. and Kunstmann, H. (2015) Stochastic bias correction of dynamically downscaled precipitation fields for Germany through copula-based integration of gridded observation data. *Hydrology and Earth System Sciences*, 19(4), 1787–1806. https://doi.org/10.5194/hess-19-1787-2015.

Maraun, D. (2016) Bias correcting climate change simulations: a critical review. *Current Climate Change Reports*, 2, 211–220. https://doi.org/10.1007/s40641-016-0050-x.

Maraun, D., Shepherd, T., Widmann, M., et al. (2017) Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, 7, 764–773. https://doi.org/10.1038/nclimate3418.

Mascaro, G., White, D.D., Westerhoff, P. and Bliss, N. (2015) Performance of the CORDEX-Africa regional climate simulations in representing the hydrological cycle of The Niger River basin. *Journal of Geophysical Research: Atmospheres*, 120(24), 12425–12444. https://doi.org/10.1002/2015jd023905.

Mbaye, M.L., Haensler, A., Hagemann, S., Gaye, A.T., Moseley, C. and Afouda, A. (2016) Impact of statistical bias correction on the projected climate change signals of the regional climate model REMO over the Senegal River basin. *International Journal of Climatology*, 36, 2035–2049. https://doi.org/10.1002/joc.4478.

Moron, V., Oueslati, B., Pohl, B. and Janicot, S. (2018) Daily weather types in February–June (1979–2016) and temperature variations in tropical North Africa. *Journal of Applied Meteorology and Climatology*, 57(5), 1171–1195. https://doi.org/10.1175/JAMC-D-17-0105.1.

Mosthaf, T. and Bárdossy, A. (2017) Regionalizing nonparametric models of precipitation amounts on different temporal scales. *Hydrology and Earth System Sciences*, 21(5), 2463–2481. https://doi.org/10.5194/hess-21-2463-2017.

Nikiema, P.M., Sylla, M.B., Ogunjobi, K., Kebe, I., Gibba, P. and Giorgi, F. (2017) Multi-model CMIP5 and CORDEX simulations of historical summer temperature and precipitation variabilities over West Africa. *International Journal of Climatology*, 37, 2438–2450. https://doi.org/10.1002/joc.4856.

Nikulin, G., Jones, C., Giorgi, F., Asrar, G., Büchner, M., Cerezo-Mota, R., Christensen, O.B., Déqué, M., Fernandez, J., Hänsler, A., van Meijgaard, E., Samuelsson, P., Sylla, M.B. and Sushama, L. (2012) Precipitation climatology in an ensemble of CORDEX-Africa regional climate simulations. *Journal of Climate*, 25(18), 6057–6078. https://doi.org/10.1175/jcli-d-11-00375.1.

Oettli, P., Sultan, B., Baron, C. and Vrac, M. (2011) Are regional climate models relevant for crop yield prediction in West Africa? *Environmental Research Letters*, 6, 014008. https://doi.org/10.1088/1748-9326/6/1/014008.

Oguntunde, P.G. and Abiodu, B.J. (2013) The impact of climate change on The Niger River basin hydroclimatology, West Africa. *Climate Dynamics*, 40(1-2), 81–94. https://doi.org/10.1007/s00382-012-1498-6.

Paeth, H., Hall, N.M., Gaertner, M.A., Alonso, M.D., Moumouni, S., Polcher, J., Ruti, P.M., Fink, A.H., Gosset, M., Lebel, T., Gaye, A.T., Rowell, D.P., Moufouma-Okia, W., Jacob, D., Rockel, B., Giorgi, F. and Rummukainen, M. (2011) Progress in regional downscaling of west African precipitation. *Atmospheric Science Letters*, 12, 75–82. https://doi.org/10.1002/asl.306.

Piani, C. and Haerter, J.O. (2012) Two dimensional bias correction of temperature and precipitation copulas in climate models. *Geophysical Research Letters*, 39, 20. https://doi.org/10.1029/2012gl053839.

Piani, C., Haerter, J.O. and Coppola, E. (2010) Statistical bias correction for daily precipitation in regional climate models over

Europe. *Theoretical and Applied Climatology*, 99(1), 187–192. https://doi.org/10.1007/s00704-009-0134-9.

Pierce, D.W., Cayan, D.R., Maurer, E.P., Abatzoglou, J.T. and Hegewisch, K.C. (2015) Improved bias correction techniques for hydrological simulations of climate change. *Journal of Hydrometeorology*, 16(6), 2421–2442. https://doi.org/10.1175/jhm-d-14-0236.1.

Polade, S.D., Pierce, D.W., Cayan, D.R., Gershunov, A. and Dettinger, M.D. (2014) The key role of dry days in changing regional climate and precipitation regimes. *Scientific Reports*, 4, 1. https://doi.org/10.1038/srep04364.

Rauch, M., Bliefernicht, J., Laux, P., Salack, S., Waongo, M. and Kunstmann, H. (2019) Seasonal forecasting of the onset of the rainy season in West Africa. *Atmosphere*, 10, 528. https://doi.org/10.3390/atmos10090528.

Rosenblatt, M. (1956) Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3), 832–837. https://doi.org/10.1214/aoms/1177728190.

Rummukainen, M. (2009) State-of-the-art with regional climate models. *Wiley Interdisciplinary Reviews: Climate Change*, 1(1), 82–96. https://doi.org/10.1002/wcc.8.

Salack, S., Bossa, A., Bliefernicht, J., Berger, S., Yira, Y., Sanoussi, K.A., Guug, S., Heinzeller, D., Avocanh, A.S., Hamadou, B., Meda, S., Diallo, B.A., Bado, I.B., Saley, I.A., Daku, E.K., Lawson, N.Z., Ganaba, A., Sanfo, S., Hien, K., Aduna, A., Steup, G., Diekkrüger, B., Waongo, M., Rogmann, A., Kunkel, R., Lamers, J.P.A., Sylla, M.B., Kunstmann, H., Barry, B., Sedogo, L.G., Jaminon, C., Vlek, P., Adegoke, J. and Moumini, S. (2019) Designing transnational Hydroclimatological observation networks and data sharing policies in West Africa. *Data Science Journal*, 18, 1–15. https://doi.org/10.5334/dsj-2019-033.

Schwarz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. https://doi.org/10.1214/aos/1176344136.

Siegmund, J., Bliefernicht, J., Laux, P. and Kunstmann, H. (2015) Toward a seasonal precipitation prediction system for West Africa: performance of CFSv2 and high-resolution dynamical downscaling. *Journal of Geophysical Research – Atmospheres*, 120, 7316–7339. https://doi.org/10.1002/2014JD022692.

Sultan, B., Guan, K., Kouressy, M., Biasutti, M., Piani, C., Hammer, G.L., McLean, G. and Lobell, D.B. (2014) Robust features of future climate change impacts on sorghum yields in West Africa. *Environmental Research Letters*, 9, 104006. https://doi.org/10.1088/1748-9326/9/10/104006.

Sun, Y., Solomon, S., Dai, A. and Portmann, R.W. (2006) How often does it rain? *Journal of Climate*, 19(6), 916–934. https://doi.org/10.1175/jcli3672.1.

Sylla, M.B., Diallo, I. and Pal, J. (2013) West African monsoon in state-of-the-science regional climate models. In: Tarhule, A. (Ed.) *Climate Variability, Regional and Thematic Patterns*. London: IntechOpen, pp. 3-36.

Sylla, M.B., Giorgi, F., Pal, J.S., Gibba, P., Kebe, I. and Nikiema, M. (2015) Projected changes in the annual cycle of high-intensity precipitation events over West Africa for the late twenty-first century. *Journal of Climate*, 28(16), 6475–6488. https://doi.org/10.1175/JCLI-D-14-00854.1.

Teutschbein, C. and Seibert, J. (2013) Is bias correction of regional climate model (RCM) simulations possible for non-stationary

conditions? *Hydrology and Earth System Sciences*, 17(12), 5061–5077. https://doi.org/10.5194/hess-17-5061-2013.

Themeßl, M.J., Gobiet, A. and Leuprecht, A. (2010) Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *International Journal of Climatology*, 31(10), 1530–1544. https://doi.org/10.1002/joc.2168.

Volosciuk, C., Maraun, D., Vrac, M. and Widmann, M. (2017) A combined statistical bias correction and stochastic downscaling method for precipitation. *Hydrology and Earth System Sciences*, 21(3), 1693–1719. https://doi.org/10.5194/hess-21-1693-2017.

Vrac, M. (2018) Multivariate bias adjustment of high-dimensional climate simulations: the rank resampling for distributions and dependences ($R^2D^2$) bias correction. *Hydrology and Earth System Sciences*, 22(6), 3175–3196. https://doi.org/10.5194/hess-22-3175-2018.

Wilks, D.S. (2009) A gridded multisite weather generator and synchronization to observed weather data. *Water Resources Research*, 45, 10. https://doi.org/10.1029/2009wr007902.

Yira, Y., Diekkrüger, B., Steup, G. and Bossa, A.Y. (2016) Modeling land use change impacts on water resources in a tropical west African catchment (Dano, Burkina Faso). *Journal of Hydrology*, 537, 187–199. https://doi.org/10.1016/j.jhydrol.2016.03.052.