Hydrology and
Earth System
Sciences

Open Access

Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions?

C. Teutschbein[1,3] **and J. Seibert**[1,2,3]

[1]Department of Physical Geography and Quaternary Geology, Stockholm University, 10691 Stockholm, Sweden
[2]Department of Geography, University of Zurich, 8057 Zurich, Switzerland
[3]Department of Earth Sciences, Uppsala University, 75236 Uppsala, Sweden

Correspondence to: C. Teutschbein (claudia.teutschbein@natgeo.su.se)

Abstract. In hydrological climate-change impact studies, regional climate models (RCMs) are commonly used to transfer large-scale global climate model (GCM) data to smaller scales and to provide more detailed regional information. Due to systematic and random model errors, however, RCM simulations often show considerable deviations from observations. This has led to the development of a number of correction approaches that rely on the assumption that RCM errors do not change over time. It is in principle not possible to test whether this underlying assumption of error stationarity is actually fulfilled for future climate conditions. In this study, however, we demonstrate that it is possible to evaluate how well correction methods perform for conditions different from those used for calibration with the relatively simple differential split-sample test.

For five Swedish catchments, precipitation and temperature simulations from 15 different RCMs driven by ERA40 (the 40 yr reanalysis product of the European Centre for Medium-Range Weather Forecasts (ECMWF)) were corrected with different commonly used bias correction methods. We then performed differential split-sample tests by dividing the data series into cold and warm respective dry and wet years. This enabled us to cross-evaluate the performance of different correction procedures under systematically varying climate conditions. The differential split-sample test identified major differences in the ability of the applied correction methods to reduce model errors and to cope with non-stationary biases. More advanced correction methods performed better, whereas large deviations remained for climate model simulations corrected with simpler approaches. Therefore, we question the use of simple correction methods such as the widely used delta-change approach and linear transformation for RCM-based climate-change impact studies. Instead, we recommend using higher-skill correction methods such as distribution mapping.

1 Introduction

In hydrological climate-change impact studies, large-scale climate variables for current and future conditions are generally provided by global climate models (GCMs). To resolve processes and features relevant to hydrology at the catchment scale, regional climate models (RCMs) are commonly used to transfer coarse-resolution GCM data to a higher resolution. Although this provides more detailed regional information (Fowler et al., 2007; Grotch and MacCracken, 1991; IPCC, 2007; Maraun et al., 2010; Salathé Jr., 2003) for hydrological simulations, there is still a mismatch of scales especially for meso- and small-scale watersheds that are often captured by only one RCM grid cell. In addition, impact modelers are also facing a risk of improper RCM simulations (Christensen et al., 2008; Teutschbein and Seibert, 2010; Varis et al., 2004) due to systematic (i.e., biases) and random model errors. Mismatching scales in combination with such errors have led to many recently developed correction approaches (Chen et al., 2013; Johnson and Sharma, 2011; Maraun et al., 2010; Teutschbein and Seibert, 2012; Themeßl et al., 2011) that help impact modelers to cope with the various problems linked to biased RCM output.

These correction approaches can be classified according to their degree of complexity and include simple-to-apply

methods such as linear transformations but also more advanced methods such as distribution mapping. The correction procedures usually identify possible differences between observed and simulated climate variables, which provide the basis for correcting both control and scenario RCM runs with a transformation algorithm. Although the correction of RCM climate variables can considerably improve hydrological simulations under current climate conditions (Chen et al., 2013; Teutschbein and Seibert, 2012), there is a major drawback: most methods follow the assumption of stationarity of model errors, which means that the correction algorithm and its parameterization for current climate conditions are assumed to also be valid for a time series of changed future climate conditions. Whether or not this condition is actually fulfilled for our future climate cannot be evaluated directly. This motivated us to address this issue and to test how well different correction methods perform for conditions different from those used for calibration. We applied the idea of a differential split-sample test, originally proposed by Klemeš (1986) for hydrological models, to analyze the performance of different correction methods for use with simulations under changed conditions. The testing presented here was done for different commonly used and rather simple correction procedures (Johnson and Sharma, 2011; Maraun et al., 2010; Teutschbein and Seibert, 2012) based on 15 RCM-simulated temperature and precipitation series for five mesoscale catchments in Sweden.

We would like to emphasize that this paper was written by local-impact modelers for impact modelers. Thus, our intention was not to compare all available methods to deal with biased RCM simulations and bias non-stationarities (this study is by no means exhaustive). We simply present one possible approach to analyze correction methods that are frequently used by impact modelers – especially on smaller scales. In addition, we outline the most common terminology related to climate models and biases as some of these terms are not used consistently by impact modelers.

2 Terminology

The terms *climate model bias* and *bias correction* are frequently used in climate change and impact research. However, these terms are not always used consistently in the literature and in many studies it is not clear whether they are actually dealing with model biases or rather model errors, model shortcomings or other uncertainties. For clarification we, therefore, briefly summarize the most commonly used terminology.

2.1 Distinction between model shortcomings, model errors and model biases

Allen et al. (2006) suggested differentiating between the terms *model shortcomings*, *model errors* and *model biases*.

Model shortcomings are based on the fact that some models do not represent some parts of the climate system or are not able to resolve certain processes. Model shortcomings might also originate from numerical issues causing problems such as the violation of mass conservation observed in several climate models (Liepert and Previdi, 2012). These deficiencies can generally be resolved by improving the model, for instance, through the introduction of new physical descriptions or through increased spatial and temporal resolution (Allen et al., 2006). Model shortcomings can lead to *model errors*.

Model errors can be caused by initial and boundary conditions, parameterizations, physical and numerical formulations, lacking knowledge of external factors or general model shortcomings (Deser et al., 2012; Eden et al., 2012; Jung, 2005; Ménard, 2010; Palmer et al., 2005). Model errors can appear as *unsystematic (random)* and *systematic errors* (Ménard, 2010). Mathematically, the time-dependent model error (e_t) is the difference between the model simulation (s_t) and the best estimate of the truth (o_t), i.e., observations (Eq. 1, modified from Jung, 2005).

$$e_t = s_t - o_t \tag{1}$$

Unsystematic (random) model errors cause random variations in model simulations. They have their origin in the internal variability of climate models, i.e., in hidden non-linearities and complex (random) dynamical processes (Allen et al., 2006; Deser et al., 2012; Majda and Gershgorin, 2010; Ménard, 2010). This internal variability is associated with a model's degree of freedom to develop its own dynamic feedback mechanisms (Christensen et al., 2001). For shorter (decadal) timescales, these random errors (internal variability) are the dominant sources of uncertainty in model simulations (Hawkins and Sutton, 2011).

Systematic model errors, also commonly termed *model biases*, produce predictably inaccurate (i.e., biased) model simulations. They are defined as systematic differences between model simulations (s_t) and observations (o_t), which is for a certain diagnostic d given by Eq. (2) (Jung, 2005), where $\hat{d}_{SE}$ stands for the estimated systematic error of the diagnostic (d) with the hat in $\hat{d}$ indicating that this is an estimate of the true value. The diagnostic d can be the mean value, but can also address other aspects of the model error.

$$\hat{d}_{SE} = \hat{d}(s_t) - \hat{d}(o_t) \tag{2}$$

Systematic model errors can originate either from inadequately constrained parameters or from model structures that are unable to describe the physical process of interest (Allen et al., 2006). These systematic model errors, or *model biases*,

are generally the most dominant source of uncertainty for longer (centennial) timescales (Hawkins and Sutton, 2011).

2.2 Climate model bias: definition and detection

Model bias is defined as a systematic distortion of statistical findings from the expected value. According to this definition, *climate model biases* describe *systematic climate model errors* (see definition above) only. It should, however, be noted that the term *bias* in the context of climate change impact studies is often misleadingly used to describe model errors in general (i.e., a combination of both systematic and random error).

Biases in climate model simulations are commonly detected by validation (i.e., comparison) with observations (Eq. 2), where the observations are considered to be "true" and unbiased (Jung, 2005; Ménard, 2010). Jung (2005) highlights the mean (μ) as one of the simplest and most widely used diagnostics to detect climate model biases, so that Eq. (2) can be modified as follows:

$$\hat{\mu}_{SE} = \hat{\mu}(s_t) - \hat{\mu}(o_t), \tag{3}$$

where $\hat{\mu}_{SE}$ is the estimated *mean* systematic error over the time period. One should be aware that $\hat{\mu}_{SE}$ can be zero (i.e., detecting no systematic error) due to error cancellation although simulations (s_t) and observations (o_t) might be characterized by different variability or distributions. This emphasizes the need for considering systematic errors for other diagnostics, i.e., replacing $\hat{\mu}$ in Eq. (3) by other statistics, amongst others the standard deviation (σ), 10th/90th percentiles (X_{10}/X_{90}) or probabilities (P). The detection and estimation of climate model biases by comparing model simulations to observations is, however, not solely limited to Eqs. (2) and (3). For example, Hanna (1993) and Chang and Hanna (2004) recommended to use the *fractional bias (FB)*, the *geometric mean bias (MG)*, the *normalized mean square error* (*NMSE*) and the *geometric variance (VG)*. These additional performance measures are mentioned here to show further possibilities of analyzing climate model biases. Depending on the focus of a climate change impact study, other measures can be defined as well (Chang and Hanna, 2004). As each of these measures has advantages and disadvantages (for more information see Chang and Hanna, 2004), any bias analysis should always be based on multiple diagnostics.

2.3 Bias correction methods

RCM simulations are typically affected by systematic and random model errors. Misestimated climate variables in general, incorrect seasonal variations of precipitation (Christensen et al., 2008; Terink et al., 2009; Teutschbein and Seibert, 2010) and the simulation of too-many drizzle (i.e., low-intensity rain) days (Ines and Hansen, 2006) are just a few examples of common systematic errors (biases). In other words, climate variables simulated by individual RCMs often do not
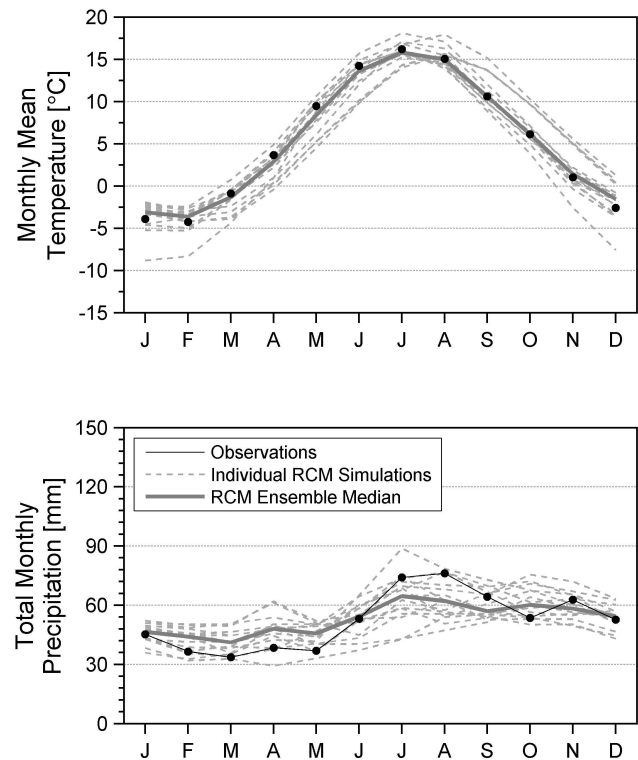

Fig. 1. Monthly mean temperature (top) and total monthly precipitation (bottom) for the period 1961–2000 as simulated by 15 individual ERA40-driven RCMs (gray dashed lines) for the Vattholmaån catchment in southeastern Sweden. Observations (black circles) and the RCM-ensemble medians (gray continuous line) are displayed as well.

agree with observed time series (Fig. 1). This poses a problem for using these simulations as input data for hydrological impact studies. One possible solution is to use an ensemble of RCM simulations (Déqué et al., 2007; Giorgi, 2006) as ensembles have two advantages: (1) the spread of individual ensemble members covers a more realistic range of uncertainty and (2) the ensemble median may fit observations better (Jacob et al., 2007), which is especially true for temperature simulations (Fig. 1, top). However, for precipitation simulations even the ensemble median often deviates considerably from observations and is not able to capture the variability in the observations (Fig. 1, bottom). This shows that it is not enough to only employ an RCM ensemble, but that additional correction procedures are needed.

Several bias correction methods have already been applied in weather forecasting under the name *model output statistics* (MOS) about four decades ago (Glahn and Lowry, 1972; Klein and Glahn, 1974). In the context of correcting RCM output, however, it is today a controversial subject (Ehret et al., 2012; Muerth et al., 2013): despite their advantageous ability to reduce errors in climate model output, most correction methods are criticized to diminish the advantages of climate models (Ehret et al., 2012) and to not have much added

value in a complex modeling chain when considering other sources of uncertainty (Muerth et al., 2013).

Typical correction approaches aim at correcting the systematic error (bias) in RCM-simulated climate variables by employing a transformation algorithm and are therefore called *bias correction methods*. The concept is based on the identification of possible biases between observed and simulated climate variables, which is the starting point for correcting both control and scenario RCM runs. It should be noted that there is a risk of not only correcting systematic errors (biases) but also unintentionally modifying simulations due to unsystematic (random) model errors (Maraun et al., 2010).

2.4　Stationarity assumption of model bias

A common assumption of most bias correction methods is stationarity, or time invariance, of the model errors. This implies that the empirical relationships in the correction algorithm and its parameterization for current climate conditions do not change over time and are also valid for future conditions. This assumption is, however, likely not met under changing climate conditions (Ehret et al., 2012; Maraun, 2012; Maraun et al., 2010; Vannitsem, 2008). In fact, Maraun (2012) was able to identify and distinguish between different types of bias changes, which are briefly described in Table 1. This highlights that there are potential issues when correction methods are applied to adjust RCM simulations.

3　Methods

3.1　Study catchments

The analysis in this study was performed for five mesoscale catchments (Fig. 2) with areas ranging from 147 to 293 km^2, as this scale is relevant for local climate change impacts (e.g., local heavy rainfall events, flooding, permafrost melt or droughts). These catchments fall all below the standard RCM grid cell size of approximately 25 km × 25 km and are, therefore, potentially affected by the scaling issue. The chosen catchments represent different typical Swedish climatic conditions and land-use types (Table 2). Continuous temperature and precipitation measurements for all five catchments were available for the period 1961–2000.

3.2　Data

Daily temperature and precipitation measurements for the period 1961–2000 were taken from the spatially interpolated 4 km × 4 km national grid PTHBV (Johansson, 2002) provided by the Swedish Meteorological and Hydrological Institute (SMHI). Climate simulations were obtained from the ENSEMBLES project (Van der Linden and Mitchell, 2009): we used daily precipitation and temperature series for the period 1961–2000 simulated by 15 RCMs (Table 3), which were all driven by ERA40 data (the 40 yr reanalysis
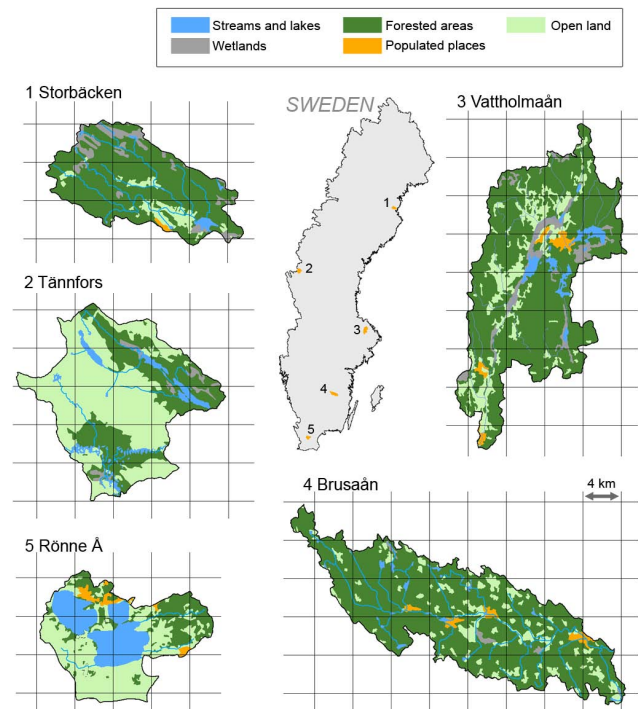

Fig. 2. Map showing locations of the Swedish study sites and the spatially interpolated 4 km × 4 km national grid of observed precipitation and temperature. Catchments: (1) Tännfors, (2) Storbäcken, (3) Vattholmaån, (4) Brusaån and (5) Rönne Å.

product of the European Centre for Medium-Range Weather Forecasts (ECMWF)). The chosen RCMs have a resolution of 25 km and, thus, the area of a single grid cell clearly exceeds the size of the study catchments. We chose to average precipitation and temperature values from the RCM grid cell with center coordinates closest to the center of the catchment and its eight neighboring grid cells.

3.3　Bias correction methods

In addition to the original (i.e., uncorrected) RCM output data, we applied and analyzed the following six bias correction methods (Table 4) to adjust RCM simulations: (1) linear transformation, (2) local intensity scaling (LOCI), (3) power transformation, (4) variance scaling, (5) distribution mapping and (6) the delta-change approach. Furthermore, a precipitation threshold was used in combination with other bias correction procedures (namely LOCI, power transformation and distribution mapping), but not considered an appropriate "stand-alone" method. More detailed descriptions of these methods can be found in Teutschbein and Seibert (2012), Gudmundsson et al. (2012), Johnson and Sharma (2011) and the original method publications provided in Table 4. All bias correction methods were applied to daily values on a monthly basis as described by Teutschbein and Seibert (2012).

Table 1. Classification and definitions of different bias changes based on Maraun (2012).

Class	Type	Abbreviation	Definition
1. Real Bias Changes	(a) Sensitivity Related Bias Changes	SBC	Systematic model errors lead to biases in the simulated local response to external forcings (so-called climate sensitivity). As the climate system might respond different to external forcings in the future, associated bias changes are defined as sensitivity related bias changes (SBC).
2. Apparent Bias Changes	(a) Variability Related Apparent Bias Changes	VABC	As biases are often detected based on relatively short time periods, sampling uncertainty mainly caused by internal variability can be a serious problem. Thus, associated bias changes are defined as variability related apparent bias changes (VABC).
	(b) Mixture Related Apparent Bias Changes	MABC	The overall bias might be a mixture of several underlying biases depending on different weather types. As the relative occurrence of these weather types might change in the future, associated bias changes are defined as mixture related apparent bias changes (MABC).

Table 2. Main characteristics of the five Swedish study sites including area, total annual precipitation, mean annual temperature, climate zones and land-use properties.

Catchments	Area	Precipitation	Temperature	Climate zone	Land-use [%]			
	$[km^2]$	$[mm\,yr^{-1}]$	$[°C]$	(Köppen–Geiger)	Forest	Open land	Lakes/wetlands	Residential
1. Tännfors	227	775	−0.5	Continental subarctic towards polar tundra	32	60	8	0
2. Storbäcken	150	617	2.1	Continental subarctic	79	9	2	0
3. Vattholmaån	293	633	5.2	Warm summer continental	81	7	10	2
4. Brusaån	240	632	5.7	Warm summer continental towards maritime temperate	83	12	3	2
5. Rönne Å	147	786	7.3	Maritime temperate	23	46	27	4

The above mentioned methods were chosen based on their frequent application in climate change impact studies. Although some of these methods might seem outdated from a climate modeler's perspective, they are all still commonly used by impact modelers especially at smaller scales, partly because they are relatively simple to apply.

3.4 Testing the stationarity assumption

To test how well bias correction methods work for conditions different from those that they were calibrated to, we employed an approach that is based on one of the operational testing methods presented by Klemeš (1986). Klemeš (1986) presented two methods of interest for systematic testing of hydrological model transposability: split-sample testing (SST) for stationary conditions and *differential* split-sample testing (DSST) for non-stationary conditions. SST implies the splitting of an available data record into two (preferably equally sized) segments in order to use one as calibration and one as validation period. DSST however should, according to Klemeš (1986), be used under changing con-

ditions. The first step of this test includes the identification of two periods with the climate variable of interest having different values, for instance a warm versus a cold or a wet versus a dry period. The model is then calibrated on the period with one condition and validated on the period with the other condition, which allows analyzing the model's ability to perform under shifting conditions. SST can automatically transform into DSST, if the two segments by nature show substantial differences in their conditions (Klemeš, 1986).

To test the ability of different correction procedures to reliably work for changed climate conditions, we applied DSST proposed by Klemeš (1986) that was originally intended for hydrological models. Both SST and DSST are seldom used to evaluate bias correction methods. We are aware of only a few other studies using such a test: Bennett et al. (2010) and Terink et al. (2010) evaluated bias correction methods using SST with two different time periods for which observations were available. A major limitation of this approach is that the periods should be long enough to represent natural climate variability satisfactorily (Bennett et al., 2010). Furthermore, unless the two periods are different in their conditions, the

Table 3. ERA40-driven RCM experiments (1961–2000) with a resolution of 25 km from the ENSEMBLES EU project used in this study.

No.	Institute	Model
1	C4I	RCA3
2	CHMI	Aladin
3	DMI	HIRHAM
4	EC	GEMLAM
5	ETHZ	CLM
6	HC	HadRM3Q0
7	HC	HadRM3Q3
8	HC	HadRM3Q16
9	ICTP	RegCM
10	KNMI	RACMO
11	METNO	HIRHAM
12	MPI	REMO
13	OUR	CRCM
14	SMHI	RCA
15	UCLM	PROMES

methods are not evaluated for use under changed conditions. This issue motivated us to rather use DSST that is better able to evaluate performance under changing climate conditions (Li et al., 2012; Seiller et al., 2012; Tramblay et al., 2013).

The available 40 yr period 1961–2000 was separated into two 20 yr subsets with different climate conditions, one representing current climate and the other one future climate. Our available 40 yr period was not long enough to show a considerable trend in precipitation or temperature data (Fig. 3a), so we instead constructed the two subsets for each catchment as follows.

1. All years were sorted ascending according to their annual amount of observed precipitation (Fig. 3b).

2. The first 20 yr of the sorted data (i.e., the driest years) were included in the first subset and the last 20 yr (i.e., wettest years) in the second subset for the precipitation-bias correction assessment.

3. Each RCM-simulated precipitation time series was rearranged to match the annual order of sorted observed precipitation data and thereafter split into two subsets as above.

The same procedure was used for constructing two subsets for the evaluation of temperature-bias correction methods. Ranking all years according to their observed annual mean temperature resulted in two series, where the first consisted of the 20 coldest years and the second of the 20 warmest years (Fig. 3c). Again, each RCM-simulated temperature time series was rearranged in the same annual order as the sorted observed temperature data and thereafter split into two subsets. This procedure resulted in series where the years

were not consecutive and the two subsets consisted of different years for the evaluation of precipitation (Fig. 3b) and temperature-bias correction methods (Fig. 3c). Note that the procedure ensured that the two subsets included the same years for each RCM simulation and the observations.

To fully apply DSST, we performed a twofold cross-validation (Fig. 3b and c). First, all correction methods were calibrated based on the first subset of years and then evaluated for the second subset of years (case 1). In addition, the two periods were switched and the correction methods were calibrated based on the second subset and validated using the first subset (case 2). This way, DSST allowed the evaluation of bias correction methods under challenging conditions, namely considerably varying climate conditions for calibration and validation (Coron et al., 2012).

3.5 Evaluation of bias correction methods

Different diagnostics (Table 5) were used to detect and estimate model errors in uncorrected RCM simulations according to Eq. (2) for both the calibration and validation period of temperature as well as precipitation. Then, the same calculations were done to analyze the performance of each bias correction method. This implies that we (1) studied whether model errors were still present for the calibration data, (2) estimated the amount of model errors present for the validation data and (3) assessed the model error growth, i.e., the absolute difference between model errors in validation and calibration data. The model error growth measure allowed studying the transferability of a bias correction method to different climatic conditions.

As the above diagnostics were applied to the entire data series of a subset, they give no information about seasonal differences. Thus, we additionally included an analysis of the four seasons: winter (DJF), spring (MAM), summer (JJA) and autumn (SON).

4 Results

4.1 DSST-induced climate change signal

The two designed subsets used in the conducted DSST featured different climate conditions and were clearly non-stationary. In this study, the differences between the two subsets were within a range of 6–30 % for precipitation (Fig. 4, left) and 0.9–1.7 °C for temperature (Fig. 4, right). These values are in the same order of magnitude as the climate change signals for Sweden that are projected by the ENSEMBLES project (Van der Linden and Mitchell, 2009) and the ClimateCost project (Christensen et al., 2011) to occur until the 2050s (2041–2070) under emission scenario A1B or until the 2080s (2070–2099) under mitigation scenario E1 compared to the baseline 1961–1990 (Christensen et al., 2011; Van der Linden and Mitchell, 2009).

Table 4. Overview of methods used to correct RCM-simulated precipitation and/or temperature data, for more information on the methods see Teutschbein and Seibert (2012) and the specified references.

| Method	Variable	Short Description	Advantages (+) and Disadvantages (-)	References
Raw RCM Output Data	Precipitation Temperature	• RCM-simulated time series are used directly without any bias correction	+ simplest way to use RCM data − systematic model errors are ignored − can cause substantial errors in impact studies	
Precipitation Threshold	Precipitation	• an RCM-specific threshold is calibrated such that the number of RCM-simulated days exceeding this threshold matches the number of observed days with precipitation • rarely used as a "stand-alone" method but often combined with other correction procedures	+ wet-day frequencies are corrected − mean, standard deviation (variance) and wet-day intensities are not adjusted	Schmidli et al. (2006)
Delta-Change Correction	Precipitation Temperature	• RCM-simulated future change signals (anomalies) are superimposed upon observational time series • usually done with a multiplicative correction for precipitation and an additive correction for temperature	+ observations are used as a basis, which makes it a robust method + corrects the mean − standard deviation (variance), wet-day frequencies and intensities are not corrected − potential future changes in climate dynamics and variability are not accounted for − all events change by the same amount	Gellens and Roulin (1998) Graham et al. (2007a, b) Johnson and Sharma (2011) Lettenmaier et al. (1999) Mpelasoka and Chiew (2009) Middelkoop et al. (2001) Moore et al. (2008) Rasmussen et al. (2012) Shabalova et al. (2003)
Linear Transformation	Precipitation Temperature	• adjusts RCM time series with correction values based on the relationship between long-term monthly mean observed and RCM control run values • precipitation is typically corrected with a factor and temperature with an additive term	+ corrects the mean + variability of corrected data is more consistent with original RCM data − standard deviation (variance), wet-day frequencies and intensities are not corrected − all events are adjusted with the same correction factor	Lenderink et al. (2007)
Local Intensity Scaling (LOCI)	Precipitation	• combines a precipitation threshold with linear scaling (both described above)	+ corrects mean, wet-day frequencies and intensities + variability of corrected data is more consistent with original RCM data − standard deviation (variance) is not corrected − all events are adjusted with the same correction factor	Schmidli et al. (2006)
Power Transformation	Precipitation	• a precipitation threshold can be introduced a priori to avoid too many drizzle days (i.e., very low but non-zero precipitation) • is a non-linear correction in an exponential form ($a \times P^b$) that combines the correction of the coefficient of variation (CV) with a linear scaling	+ corrects mean and standard deviation (variance) + events are adjusted non-linearly + variability of corrected data is more consistent with original RCM data ± adjusts wet-day frequencies and intensities only to some extend	Leander and Buishand (2007) Leander et al. (2008)
Variance Scaling	Temperature	• combines standard linear scaling with a scaling based on standard deviations	+ corrects mean and standard deviation (variance) + variability of corrected data is more consistent with original RCM data − all events are adjusted with the same addends and correction factor	Chen et al. (2011)
Distribution Mapping	Precipitation Temperature	• matches the distribution functions of observations and RCM-simulated climate values • a precipitation threshold can be introduced to avoid substantial distortion of the distribution caused by too many drizzle days (i.e., very low but non-zero precipitation) • also known as "quantile-quantile mapping", "probability mapping", "statistical downscaling" or "histogram equalization"	+ corrects mean, standard deviation (variance), wet-day frequencies and intensities + events are adjusted non-linearly + variability of corrected data is more consistent with original RCM data	Block et al. (2009) Boe et al. (2007) Déqué et al. (2007) Ines and Hansen (2006) Johnson and Sharma (2011) Piani et al. (2010) Rojas et al. (2011) Sennikovs and Bethers (2009) Sun et al. (2011)
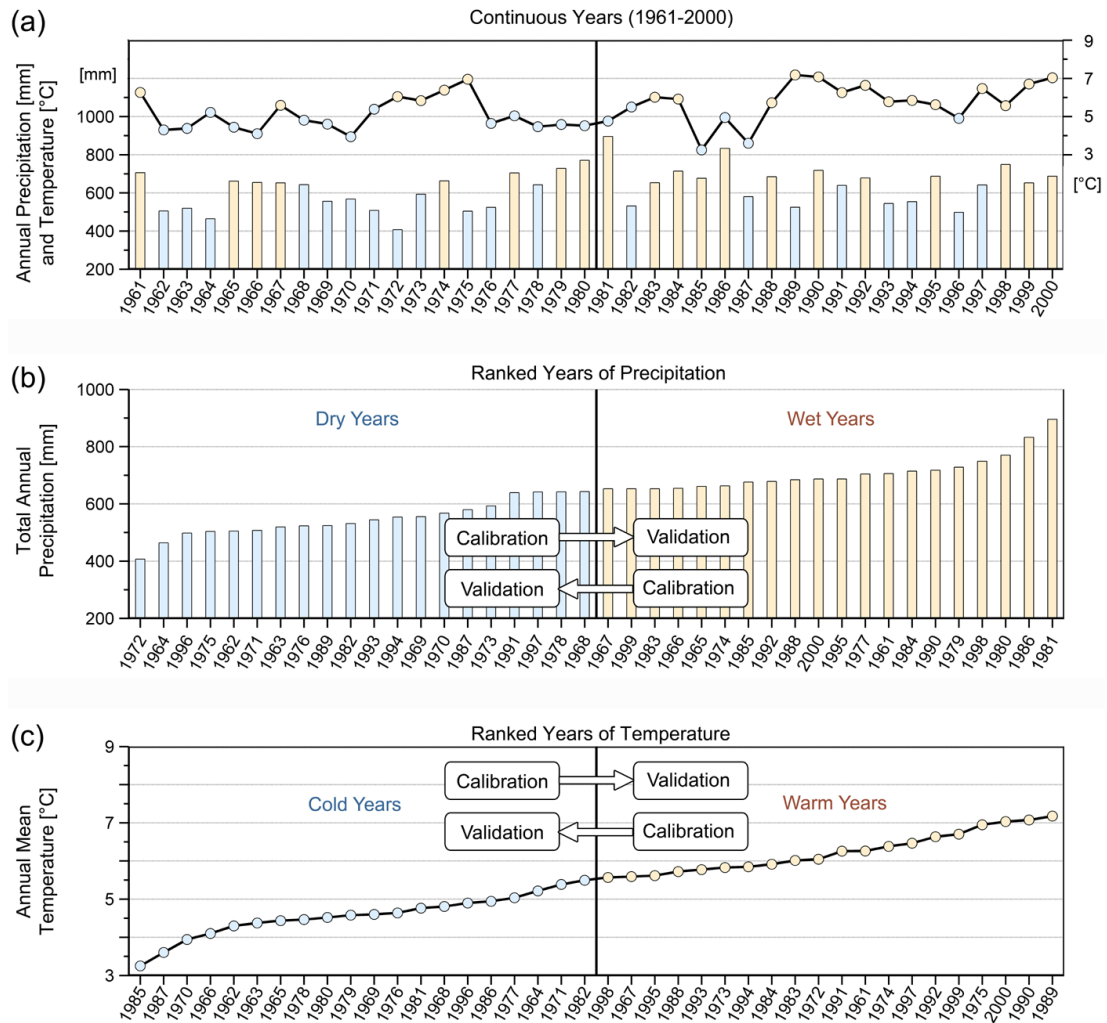
Fig. 3. Exemplary procedure of the differential split-sample test (DSST). First, the annual values **(a)** were sorted ascending **(b, c)**. For the twofold cross-validation, first the lower-value years were used for calibration and the higher-value years for validation. In a second step, calibration and validation periods were switched. This procedure was performed independently for precipitation and temperature.

Table 5. Overview of diagnostics used to detect and estimate model errors in raw and corrected RCM-simulated precipitation and/or temperature data.

RCM-Simulated Climate Variable	Diagnostic	Abbreviation
Precipitation	1. Mean	(μ)
	2. 90th percentile	(X_{90})
	3. Standard deviation	(σ)
	4. Probability of wet days	(P_{wet})
	5. Intensity of wet days	(i_{wet})
	6. Maximum of consecutive 5 day precipitation	$(\mathrm{Precip}_{5\mathrm{max}})$
Temperature	1. Mean	(μ)
	2. 10th percentile	(X_{10})
	3. 90th percentile	(X_{90})
	4. Standard deviation	(σ)

Table 6. Mean absolute error (MAE) of seasonally averaged precipitation [mm d^{-1}] of differently corrected ERA40-driven RCM simulations compared to observed seasonally averaged precipitation for case-1 evaluation. Larger MAE values are shown in bold, medium MAE values appear in regular font and smaller MAE values are underlined.

| Season	Catchment	Bias Correction Method					
		Raw	Linear	LOCI	Power	Distribution	Delta
Winter (DJF)	1. Tännfors	**0.428**	0.190	0.185	**0.212**	**0.213**	**0.232**
	2. Storbäcken	0.174	_0.095_	_0.095_	_0.111_	_0.105_	_0.105_
	3. Vattholmaån	_0.085_	0.131	0.106	0.080	_0.070_	_0.140_
	4. Brusaån	**0.257**	_0.109_	_0.095_	_0.101_	_0.100_	_0.130_
	5. Rönne Å	**0.236**	**0.358**	**0.323**	**0.268**	**0.237**	**0.337**
Spring (MAM)	1. Tännfors	**0.244**	_0.081_	_0.078_	_0.143_	0.153	_0.081_
	2. Storbäcken	**0.290**	**0.348**	**0.367**	**0.436**	**0.455**	**0.353**
	3. Vattholmaån	**0.304**	_0.093_	_0.083_	_0.101_	_0.111_	_0.093_
	4. Brusaån	**0.462**	0.191	**0.203**	**0.234**	**0.250**	**0.203**
	5. Rönne Å	**0.438**	**0.236**	**0.250**	**0.289**	**0.311**	**0.267**
Summer (JJA)	1. Tännfors	**0.346**	0.181	0.174	0.150	_0.147_	**0.211**
	2. Storbäcken	**0.427**	**0.447**	**0.438**	**0.308**	**0.310**	**0.439**
	3. Vattholmaån	**0.285**	0.157	0.150	0.154	_0.144_	0.159
	4. Brusaån	**0.238**	_0.067_	_0.056_	_0.047_	_0.041_	_0.105_
	5. Rönne Å	**0.245**	0.168	0.168	0.167	0.173	**0.244**
Autumn (SON)	1. Tännfors	_0.129_	**0.248**	**0.243**	0.196	0.191	**0.260**
	2. Storbäcken	_0.140_	_0.102_	_0.093_	_0.079_	_0.074_	_0.143_
	3. Vattholmaån	0.174	**0.215**	**0.211**	**0.217**	**0.214**	**0.234**
	4. Brusaån	_0.111_	0.191	0.181	0.182	0.177	0.162
	5. Rönne Å	_0.098_	_0.144_	_0.131_	_0.120_	_0.107_	_0.145_
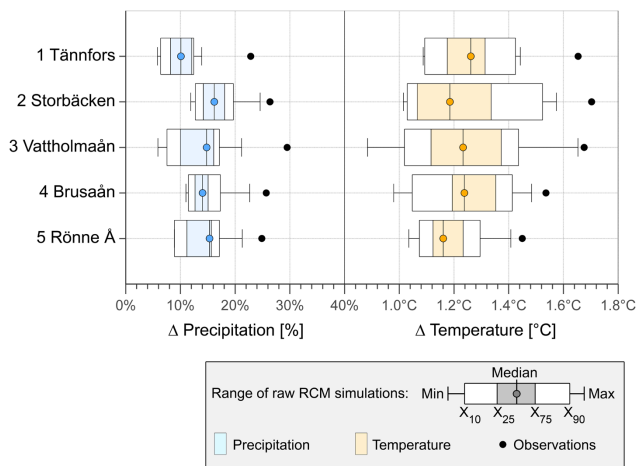
Fig. 4. Differences between mean values of the two constructed subsets for precipitation (left) and temperature (right) shown for raw RCM simulations (colored boxes) and observations (black circles).

In direct comparison to observations, the RCMs tended to underestimate the mean climate change signals for both precipitation (Fig. 4, left) and temperature (Fig. 4, right), which was most likely directly related to an underestimation of interannual variability by the RCMs.

4.2 RCM precipitation: model errors

The calculated precipitation model errors were displayed in gridded plots as a function of bias correction method (x axis) and catchment location (y axis) separately for each statistical diagnostic and separately for calibration and validation period. Considering the case-1 evaluation procedure from dry to wet years (Fig. 5), all precipitation-bias correction methods resulted in good estimates of the mean (μ) showing only small model errors during the designed calibration period (Fig. 5, upper left panel). Analyzing other statistical diagnostics, however, showed considerable differences between the methods already during the calibration period (Fig. 5, left panel column). Raw RCM simulations generally had large model errors. Linear transformation was not able to considerably improve other statistical properties than μ (Fig. 5, left panel column). For standard deviation (σ), 90th percentiles (X_{90}) and maximum 5 day precipitation (Precip$_{5\text{max}}$), power transformation and distribution mapping seemed to work best. The same could be observed for the probability of wet days (P_{wet}) and the intensity of wet days (i_{wet}), which were in addition also most correct after applying LOCI. The delta-change approach always performed perfectly during calibration by its definition.

The overall model error pattern was fairly similar for calibration and validation period. The major difference was
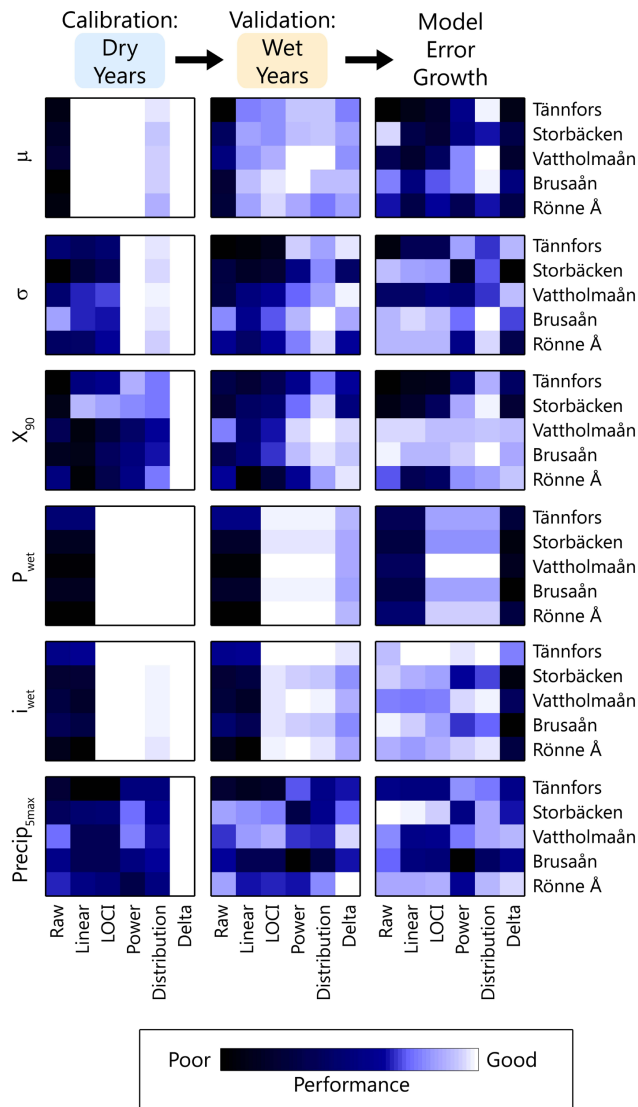
Fig. 5. Normalized precipitation model errors (different shades of blue) of 6 different statistical diagnostics (panel rows) for calibration period (left panel column) and validation period (central panel column). The model error growth, i.e., the absolute difference between model errors in validation and calibration period, is shown as well (right panel column). Each panel displays model error information depending on the 6 applied bias correction methods (x axis) and the 5 catchment locations (y axis). The figure is based on case-1 evaluation with calibration on dry years and validation on wet years.

Fig. 6. Normalized precipitation model errors (different shades of blue) of 6 different statistical diagnostics (panel rows) for calibration period (left panel column) and validation period (central panel column). The model error growth, i.e., the absolute difference between model errors in validation and calibration period, is shown as well (right panel column). Each panel displays model error information depending on the 6 applied bias correction methods (x axis) and the 5 catchment locations (y axis). The figure is based on case-2 evaluation with calibration on wet years and validation on dry years.

that the model error during validation period increased considerably (Fig. 5, central panel column, shown as darker blue shading). This fact was also supported by the calculated model error growth (Fig. 5, right panel column). Linear transformation tended to have a slightly larger model error growth, whereas distribution mapping had the least. An interesting fact is that the delta-change approach, despite the illusory perfect fit during calibration, was outperformed by other methods during validation: Delta-change corrected precipi-
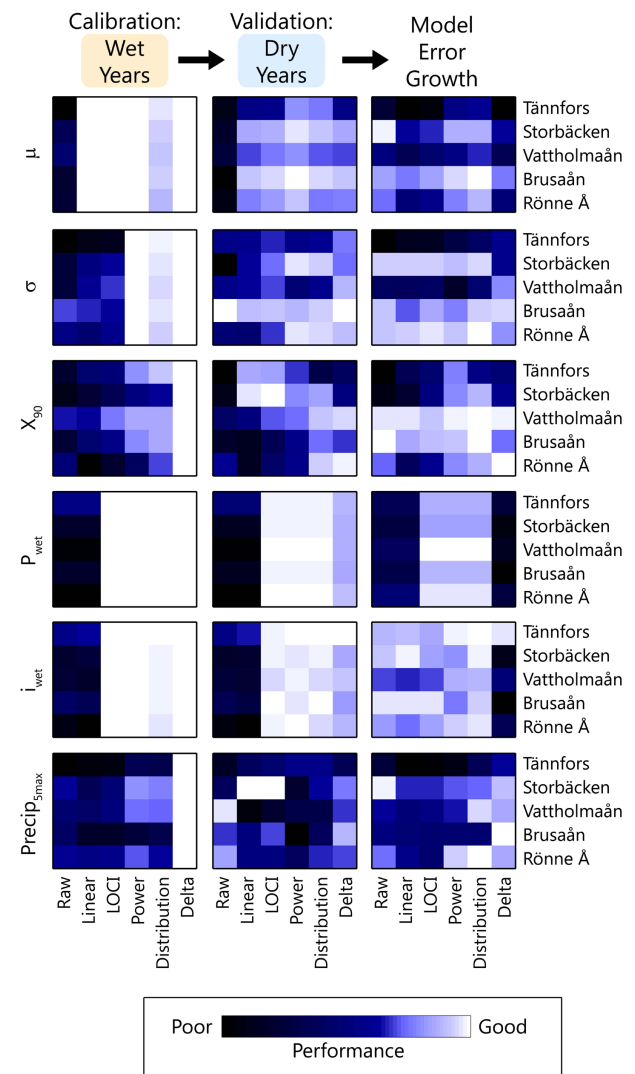
tation showed large deviations in μ as well as P_{wet} and i_{wet}. Consequently, the delta-change method showed the strongest model error growth. The case-2 evaluation procedure from wet to dry years (Fig. 6) mostly confirmed the results of the case-1 evaluation.
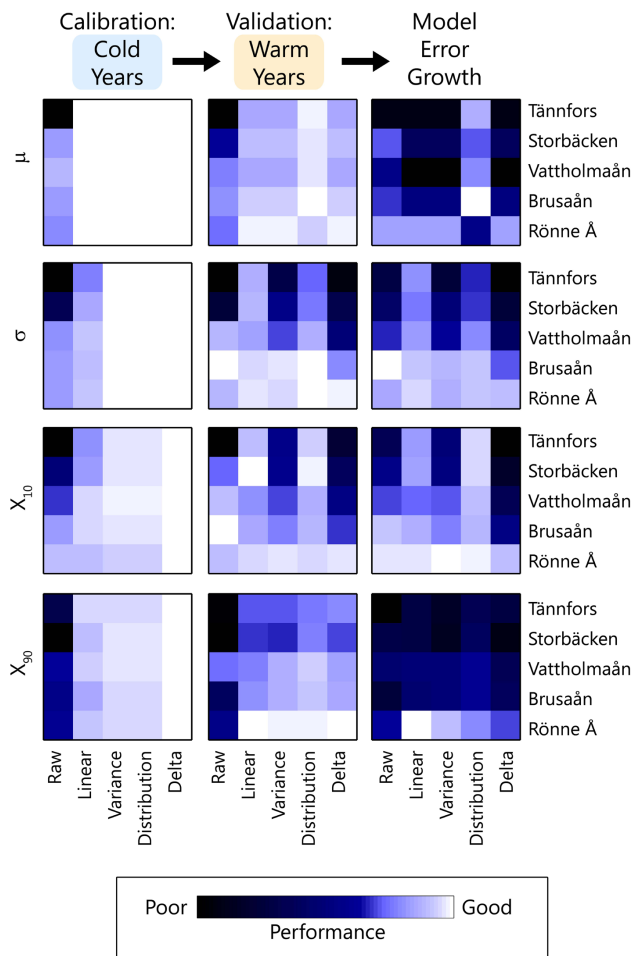
Fig. 7. Normalized temperature model errors (different shades of blue) of 4 different statistical diagnostics (panel rows) for calibration period (left panel column) and validation period (central panel column). The model error growth, i.e., the absolute difference between model errors in validation and calibration period, is shown as well (right panel column). Each panel displays model error information depending on the 5 applied bias correction methods (*x* axis) and the 5 catchment locations (*y* axis). The figure is based on case-1 evaluation with calibration on cold years and validation on warm years.

Fig. 8. Normalized temperature model errors (different shades of blue) of 4 different statistical diagnostics (panel rows) for calibration period (left panel column) and validation period (central panel column). The model error growth, i.e., the absolute difference between model errors in validation and calibration period, is shown as well (right panel column). Each panel displays model error information depending on the 5 applied bias correction methods (*x* axis) and the 5 catchment locations (*y* axis). The figure is based on case-2 evaluation with calibration on warm years and validation on cold years.

4.3 RCM precipitation: seasonal analysis

The analysis of seasonally averaged raw and corrected RCM-simulated precipitation for the validation period revealed only a weak pattern in terms of the influence of different correction methods, seasons and catchments on model errors (Table 6). The mean absolute error (MAE) was generally large for raw RCM-simulated precipitation, except for autumn. During autumn, which is characterized by medium to high monthly precipitation, raw RCM simulations were relatively close to observations and the correction methods were not able to provide further enhancement (except for catchment 2, Storbäcken). During all other seasons the cor-
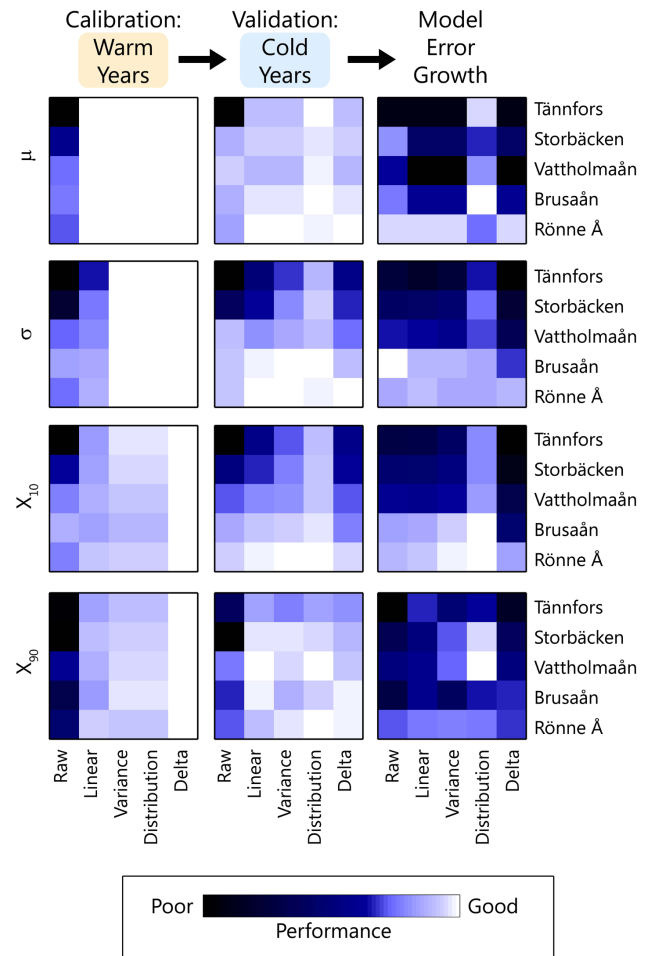
rection methods were generally able to improve raw RCM simulations (except for catchment 5, Rönne Å, in winter and catchment 2, Storbäcken, in spring). Power transformation and distribution mapping performed better than other methods in winter, summer and autumn, which are seasons that are characterized by somewhat higher monthly precipitation. On the other hand, linear scaling and LOCI performed generally better in spring, a season with lower monthly precipitation. Furthermore, bias correction methods worked better for catchments in south central Sweden (i.e., catchments 3, Vattholmaån, and 4, Brusaån), which are generally drier than the other three catchments.

4.4 RCM temperature: model errors

The same type of gridded plots were created to demonstrate the calculated temperature model errors: in terms of the case-1 evaluation procedure from cold to warm years (Fig. 7), all temperature-bias correction methods resulted in very good μ estimates during the designed calibration period (Fig. 7, upper left panel). Substantial differences between the correction methods became apparent with help of other statistical diagnostics (Fig. 7, left panel column): linear transformation was the only method not able to sufficiently correct σ. Variance-scaled and distribution-mapped RCM temperature had both the most correct X_{10} and X_{90} during the control period. Again, the delta-change approach is perfect (i.e., model-error-free) by definition.

During validation with warmer years (Fig. 7, central panel row), the bias correction methods performed somewhat differently. Especially variance scaling showed larger model errors in σ and X_{10} compared to the other methods. Distribution mapping, however, had relatively low μ, σ, X_{10} and X_{90} model errors. The model error growth (Fig. 7, right panel column) identified variance scaling and the delta-change method as the two approaches with the largest model error increase. The delta-change approach again had one of the largest model errors during validation and, thus, the strongest model error growth. Overall, a north–south gradient became apparent. Raw and adjusted RCM temperatures were characterized by a larger model error and a stronger model error growth for the northern catchments with cold climate conditions compared to the catchments with a warmer climate.

These findings were confirmed by the case-2 evaluation procedure from warm to cold years (Fig. 8). The results were essentially the same as for the case-1 evaluation; only that linear transformation performed worse and showed larger model errors during validation (Fig. 8, central panel column). Moreover, the north–south gradient was even more pronounced in all panels for validation and model error growth (Fig. 8, central and right panel column).

4.5 RCM temperature: seasonal analysis

The evaluation of seasonally averaged raw and corrected RCM-simulated temperature showed clear differences between correction methods, seasons and catchments for the validation period (Table 7). The mean absolute error (MAE) was generally large for raw RCM-simulated temperature. A north–south gradient was visible with northern catchments showing larger model errors in raw and corrected temperature. Furthermore, there was a clear seasonal difference: winter temperatures (cold season) were much more flawed than temperatures in all other (warmer) seasons. Distribution mapping consistently showed lowest MAE values.
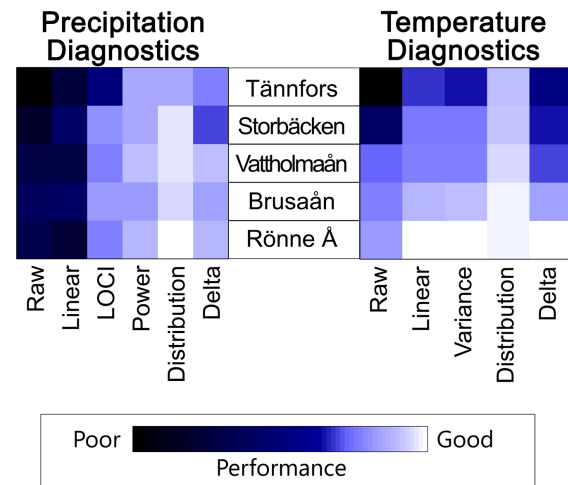

Fig. 9. Overall performance of precipitation (left) and temperature (right) biases correction methods. Each panel displays model error information depending on the applied bias correction methods (x axis) and the 5 catchment locations (y axis). Each plot combines normalized model errors of different statistical diagnostics and was created by superimposing the different panels from Figs. 5 and 6 for precipitation (left) as well as the panels from Figs. 7 and 8 for temperature (right). Light color fields represent small model errors and little model error growth; darker colors characterize larger model errors and more model error growth.

4.6 Overall performance of bias correction methods

In order to obtain information on the overall performance of each bias correction method and its transferability to different climate conditions, we first normalized the model errors of raw and corrected RCM simulations to bring them to a common comparable scale (shown in Figs. 5–8). These normalized errors were then averaged over different diagnostics, over both the case-1 and case-2 evaluation and over both subsets (calibration and validation). Thus, we obtained a combination of the results in Figs. 5 and 6 for precipitation and in Figs. 7 and 8 for temperature.

The obtained signal was reasonably clear for both precipitation (Fig. 9, left) and temperature (Fig. 9, right): raw RCM simulations had the largest model error. In general, the more advanced the algorithm of a bias correction method was, the smaller was the model error present after correction (Fig. 9, from left to right on the x axis). This means that linear transformation had the largest and distribution mapping the smallest model errors. Furthermore, the simple delta-change approach resulted in relatively large model errors.

Table 7. MAE of seasonally averaged temperature [°C] of differently corrected ERA40-driven RCM simulations compared to observed seasonally averaged temperature for case-1 evaluation. Larger MAE values are shown in bold, medium MAE values appear in regular font and smaller MAE values are underlined.

Season	Catchment	Bias Correction Method					
		Raw	Linear	LOCI	Power	Distribution	Delta
Winter (DJF)	1 Tännfors	**3.659**	**1.544**	**3.659**	**3.659**	**0.610**	**1.704**
	2 Storbäcken	**0.983**	**1.326**	**0.983**	**0.983**	**0.476**	**1.438**
	3 Vattholmaån	0.326	**1.476**	0.326	0.326	**0.807**	**1.463**
	4 Brusaån	0.382	**0.888**	0.382	0.382	0.382	**0.968**
	5 Rönne Å	<u>0.280</u>	**0.486**	<u>0.280</u>	<u>0.280</u>	0.456	**0.554**
Spring (MAM)	1 Tännfors	**1.498**	<u>0.170</u>	**1.498**	**1.498**	0.349	0.357
	2 Storbäcken	**1.236**	<u>0.251</u>	**1.236**	**1.236**	0.459	<u>0.265</u>
	3 Vattholmaån	**0.754**	<u>0.242</u>	**0.754**	**0.754**	<u>0.227</u>	<u>0.261</u>
	4 Brusaån	**0.544**	<u>0.150</u>	**0.544**	**0.544**	<u>0.281</u>	<u>0.171</u>
	5 Rönne Å	**0.583**	<u>0.159</u>	**0.583**	**0.583**	<u>0.177</u>	<u>0.104</u>
Summer (JJA)	1 Tännfors	**0.858**	0.445	**0.858**	**0.858**	0.452	0.465
	2 Storbäcken	**1.469**	0.332	**1.469**	**1.469**	0.325	0.367
	3 Vattholmaån	0.419	<u>0.279</u>	0.419	0.419	<u>0.205</u>	<u>0.259</u>
	4 Brusaån	**0.486**	<u>0.213</u>	**0.486**	**0.486**	<u>0.276</u>	<u>0.140</u>
	5 Rönne Å	**0.665**	<u>0.165</u>	**0.665**	**0.665**	<u>0.207</u>	<u>0.238</u>
Autumn (SON)	1 Tännfors	**1.957**	<u>0.313</u>	**1.957**	**1.957**	<u>0.215</u>	**0.486**
	2 Storbäcken	<u>0.298</u>	<u>0.207</u>	<u>0.298</u>	<u>0.298</u>	0.293	0.366
	3 Vattholmaån	<u>0.187</u>	<u>0.127</u>	<u>0.187</u>	<u>0.187</u>	<u>0.106</u>	**0.569**
	4 Brusaån	<u>0.201</u>	<u>0.135</u>	<u>0.201</u>	<u>0.201</u>	0.153	0.437
	5 Rönne Å	<u>0.222</u>	<u>0.180</u>	<u>0.222</u>	<u>0.222</u>	<u>0.173</u>	<u>0.248</u>

5 Discussion

Based on all findings in this study, distribution mapping showed the best overall performance and transferability to potentially changed climate conditions, as it was able to correct statistical moments other than the mean and standard deviation. LOCI and power transformation (both for precipitation) as well as variance scaling (for temperature) performed moderately. It should be noted that variance scaling is not advisable as it is based on the invalid assumption that all local variability is related to larger-scale variability and, furthermore, tends to augment the mean square errors of uncorrected data (Von Storch, 1999). Linear transformation and the delta-change method were the least able to correct for overall model errors in the validation period.

In this study, we did not try to answer the "main question [. . .], whether and when the application of bias correction methods [. . .] is justified or not" (Ehret et al., 2012). Bias correction methods are often criticized to diminish the advantages of climate models, but even with today's much advanced climate models, bias correction is often unavoidable for climate-change impact studies as uncorrected RCM simulations are a source of large uncertainties and would consequently hamper subsequent impact simulations. However, one needs to be aware that there are several problematic aspects related to bias correction methods (Ehret et al., 2012):

– physical causes of model errors are not taken into account and, thus, a proper physical foundation is missing (Teutschbein and Seibert, 2012);

– spatiotemporal field consistency and relations between climate variables are modified (Ehret et al., 2012);

– conservation principles are not met (Ehret et al., 2012);

– feedback mechanisms are neglected (Ehret et al., 2012);

– the stationarity (time invariance) assumption is likely not met under changing climate conditions (Ehret et al., 2012; Maraun, 2012; Maraun et al., 2010; Vannitsem, 2008);

– variability ranges might be reduced without physical justification (Ehret et al., 2012);

– the climate-change signal might be altered (Dosio et al., 2012; Hagemann et al., 2011);

– the choice for a correction technique is an additional source of uncertainty (Chen et al., 2011; Teutschbein and Seibert, 2012; Teutschbein et al., 2011);

- the added value of bias correction methods is questionable in a complex modeling chain with other major sources of uncertainty (Muerth et al., 2013);

- impacts of bias correction methods and related uncertainties are not communicated to end-users (Ehret et al., 2012);

- effects of unsystematic (random) model errors could by mistake be blamed on systematic errors and, therefore, accidentally be modified by correction methods (Maraun et al., 2010).

For current climate conditions, Teutschbein and Seibert (2012) demonstrated that most correction approaches applied are able to improve raw RCM data to some extent, but that there are considerable differences in the quality of adjusted RCM temperature and precipitation. In this paper, we showed how DSST can be used to analyze the transferability of correction approaches to different climate conditions. Using DSST allows identifying clear differences in reproducing conditions similar to and conditions different from those that the correction approaches were calibrated to. These differences are an indicator for improper algorithm and parameter transfers.

By using the coldest/driest and warmest/wettest years for separation of the periods, we certainly pushed the correction methods. This was done on purpose, because we believe that reliable simulations of the more extreme years are essential for certain impact assessments, such as drought and flood modeling under future climate conditions. To test the transferability of correction approaches on conditions of a less extreme climate-change signal, it is also possible to use more moderate extrapolations by applying, for instance, the generalized split-sample test (GSST) as proposed by Coron et al. (2012).

We would like to emphasize that DSST is a rather simple approach to demonstrate algorithm transferability to different conditions. Another feasible approach to identify bias non-stationarities is to use a pseudo-reality, which is based on using one RCM simulation as reference and performing an inter-model cross-validation with other RCM simulations (Maraun, 2012; Räisänen and Räty, 2012). Such a pseudo-reality can identify potential issues with bias correction methods, but it is a rather complex exercise, expensive in terms of computing power and does not necessarily identify where correction approaches might be successful when compared to real observations (Maraun, 2012).

6 Conclusions

The choice between bias correction algorithms plays a large role in assessing hydrological climate change impacts. For current conditions, we could easily limit this choice to the one that performed best. For simulations of future climate

this is more difficult and the fundamental question is how transferable the different methods are. The differential split-sample test suggested here is a simple and yet powerful tool to evaluate this. It is possible to create two subsets of data with considerably different climate conditions and non-stationary model errors based on time series of observations and RCM simulations of current climate (no future simulations necessary). Thus, the transferability of different bias correction methods can be tested under non-stationary conditions.

The delta-change approach and the linear transformation are the two most common transfer methods and have been widely used (Gellens and Roulin, 1998; Graham et al., 2007a, b; Lettenmaier et al., 1999; Middelkoop et al., 2001; Moore et al., 2008; Shabalova et al., 2003), because they are straightforward and easy to implement due to their simplicity. Yet, our validation of these correction approaches with the differential split-sample test showed that these two methods result in large deviations and are the least reliable under changed conditions. These findings remain to be confirmed for other catchments and other geographic regions, but based on the findings in this study we question the use of the delta-change method or the linear transformation to correct RCM scenarios of future conditions for climate change impact studies. Instead, we would like to recommend distribution mapping as the best-performing correction method, because it was best able to cope with non-stationary conditions. However, regardless of the used method, our results demonstrate that the – in most climate impact studies unavoidable – use of bias correction approaches for conditions different from those being used for their parameterization, might result in significant uncertainties. In this study RCMs driven by ERA40 reanalysis data were evaluated and uncertainties can be expected to be even larger when using GCM driven RCMs.

Edited by: H. H. G. Savenije

References

Allen, M., Frame, D., Kettleborough, J., and Stainforth, D.: Model error in weather and climate forecasting, in: Predictability of Weather and Climate, edited by: Palmer, T. and Hagedorn, R., 391–427, Cambridge University Press, 2006.

Bennett, J. C., Ling, F. L. N., Graham, B., Grose, M. R., Corney, S. P., White, C. J., Holz, G. K., Post, D. A., Gaynor, S. M., and Bindoff, N. L.: Climate Futures for Tasmania: Water and Catchments., Technical Report, Antarctic Climate & Ecosystems Cooperative Research Centre, Hobart, Tasmania, 2010.

Block, P. J., Souza Filho, F. A., Sun, L., and Kwon, H. H.: A Streamflow Forecasting Framework using Multiple Climate and Hydrological Models1, J. Am. Water Resour. As., 45, 828–843, doi:10.1111/j.1752-1688.2009.00327.x, 2009.

Boe, J., Terray, L., Habets, F., and Martin, E.: Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies, Int. J. Climatol., 27, 1643–1655, doi:10.1002/joc.1602, 2007.

Chang, J. C. and Hanna, S. R.: Air quality model performance evaluation, Meteorol. Atmos. Phys., 87, 167–196, doi:10.1007/s00703-003-0070-7, 2004.

Chen, J., Brissette, F. P., and Leconte, R.: Uncertainty of downscaling method in quantifying the impact of climate change on hydrology, J. Hydrol., 401, 190–202, doi:10.1016/j.jhydrol.2011.02.020, 2011.

Chen, J., Brissette, F. P., Chaumont, D., and Braun, M.: Performance and uncertainty evaluation of empirical downscaling methods in quantifying the climate change impacts on hydrology over two North American river basins, J. Hydrol., 479, 200–214, doi:10.1016/j.jhydrol.2012.11.062, 2013.

Christensen, J. H., Boberg, F., Christensen, O. B., and Lucas-Picher, P.: On the need for bias correction of regional climate change projections of temperature and precipitation, Geophys. Res. Lett., 35, L20709, doi:10.1029/2008GL035694, 2008.

Christensen, O. B., Gaertner, M. A., Prego, J. A., and Polcher, J.: Internal variability of regional climate models, Clim. Dyn., 17, 875–887, doi:10.1007/s003820100154, 2001.

Christensen, O. B., Goodess, C. M., Harris, I., and Watkiss, P.: European and Global Climate Change Projections: Discussion of Climate Change Model Outputs, Scenarios and Uncertainty in the EC RTD ClimateCost Project, in The ClimateCost Project. Final Report. Volume 1: Europe, edited by P. Watkiss, Published by the Stockholm Environment Institute, Sweden, available at: http://www.climatecost.cc/images/Policy_brief_1_Projections_05_lowres.pdf (last access: 5 March 2013), 2011.

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., and Hendrickx, F.: Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments, Water Resour. Res., 48, W05552, doi:10.1029/2011WR011721, 2012.

Déqué, M., Rowell, D. P., Lüthi, D., Giorgi, F., Christensen, J. H., Rockel, B., Jacob, D., Kjellström, E., De Castro, M., and van den Hurk, B.: An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections, Clim. Change, 81, 53–70, doi:10.1007/s10584-006-9228-x, 2007.

Deser, C., Phillips, A., Bourdette, V., and Teng, H.: Uncertainty in climate change projections: the role of internal variability, Clim. Dyn., 38, 527–546, doi:10.1007/s00382-010-0977-x, 2012.

Dosio, A., Paruolo, P., and Rojas, R.: Bias correction of the ENSEMBLES high resolution climate change projections for use by impact models: Analysis of the climate change signal, J. Geophys. Res., 117, D17110, doi:10.1029/2012JD017968, 2012.

Eden, J. M., Widmann, M., Grawe, D., and Rast, S.: Skill, Correction, and Downscaling of GCM-Simulated Precipitation, J. Climate, 25, 3970–3984, doi:10.1175/JCLI-D-11-00254.1, 2012.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions "Should we apply bias correction to global and regional climate model data?," Hydrol. Earth Syst. Sci., 16, 3391–3404, doi:10.5194/hess-16-3391-2012, 2012.

Fowler, H. J., Blenkinsop, S., and Tebaldi, C.: Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, Int. J. Climatol., 27, 1547–1578, doi:10.1002/joc.1556, 2007.

Gellens, D. and Roulin, E.: Streamflow response of Belgian catchments to IPCC climate change scenarios, J. Hydrol., 210, 242–258, doi:10.1016/S0022-1694(98)00192-9, 1998.

Giorgi, F.: Regional climate modeling: Status and perspectives, J. Phys. IV, 139, 101–118, doi:10.1051/jp4:2006139008, 2006.

Glahn, H. R. and Lowry, D. A.: The Use of Model Output Statistics (MOS) in Objective Weather Forecasting, J. Appl. Meteorol., 11, 1203–1211, doi:10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2, 1972.

Graham, L., Andréasson, J., and Carlsson, B.: Assessing climate change impacts on hydrology from an ensemble of regional climate models, model scales and linking methods – a case study on the Lule River basin, Clim. Change, 81, 293–307, doi:10.1007/s10584-006-9215-2, 2007a.

Graham, L., Hagemann, S., Jaun, S., and Beniston, M.: On interpreting hydrological change from regional climate models, Clim. Change, 81, 97–122, doi:10.1007/s10584-006-9217-0, 2007b.

Grotch, S. L. and MacCracken, M. C.: The use of general circulation models to predict regional climatic change, J. Climate, 4, 286–303, doi:10.1175/1520-0442(1991)004<0286:TUOGCM>2.0.CO;2, 1991.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T.: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations - a comparison of methods, Hydrol. Earth Syst. Sci., 16, 3383–3390, doi:10.5194/hess-16-3383-2012, 2012.

Hagemann, S., Chen, C., Haerter, J. O., Heinke, J., Gerten, D., and Piani, C.: Impact of a Statistical Bias Correction on the Projected Hydrological Changes Obtained from Three GCMs and Two Hydrology Models, J. Hydrometeorol., 12, 556–578, doi:10.1175/2011JHM1336.1, 2011.

Hanna, S. R.: Uncertainties in air quality model predictions, Bound.-Lay. Meteorol., 62,, 3–20, doi:10.1007/BF00705545, 1993.

Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in projections of regional precipitation change., Clim. Dyn., 37, 407–418, doi:10.1007/s00382-010-0810-6, 2011.

Ines, A. V. M. and Hansen, J. W.: Bias correction of daily GCM rainfall for crop simulation studies, Agr. Forest Meteorol., 138, 44–53, doi:10.1016/j.agrformet.2006.03.009, 2006.

IPCC: Climate Change 2007: The Physical Science Basis, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, UK, and New York, USA, 2007.

Jacob, D., Bärring, L., Christensen, O., Christensen, J., de Castro, M., Déqué, M., Giorgi, F., Hagemann, S., Hirschi, M., Jones, R., Kjellström, E., Lenderink, G., Rockel, B., Sánchez, E., Schär, C., Seneviratne, S., Somot, S., van Ulden, A., and van den Hurk, B.: An inter-comparison of regional climate models for Europe: model performance in present-day climate, Clim. Change, 81, 31–52, doi:10.1007/s10584-006-9213-4, 2007.

Johansson, B.: Estimation of areal precipitation for hydrological modelling in Sweden, Dissertation, Earth Science Centre, Göteborg University, Report A76. Göteborg, Sweden., 2002.

Johnson, F. and Sharma, A.: Accounting for interannual variability: A comparison of options for water resources climate change impact assessments, Water Resour. Res., 47, W04508, doi:10.1029/2010WR009272, 2011.

Jung, T.: Systematic errors of the atmospheric circulation in the ECMWF forecasting system, Q. J. Roy. Meteorol. Soc., 131, 1045–1073, doi:10.1256/qj.04.93, 2005.

Klein, W. H. and Glahn, H. R.: Forecasting Local Weather by Means of Model Output Statistics, B. Am. Meteorol. Soc., 55, 1217–1227, doi:10.1175/1520-0477(1974)055<1217:FLWBMO>2.0.CO;2, 1974.

Klemeš, V.: Operational testing of hydrological simulation models/Vérification, en conditions réelles, des modèles de simulation hydrologique, Hydrolog. Sci. J., 31, 13–24, doi:10.1080/02626668609491024, 1986.

Leander, R. and Buishand, T. A.: Resampling of regional climate model output for the simulation of extreme river flows, J. Hydrol., 332, 487–496, doi:10.1016/j.jhydrol.2006.08.006, 2007.

Leander, R., Buishand, T. A., van den Hurk, B. J. J. M., and de Wit, M. J. M.: Estimated changes in flood quantiles of the river Meuse from resampling of regional climate model output, J. Hydrol., 351, 331–343, doi:10.1016/j.jhydrol.2007.12.020, 2008.

Lenderink, G., Buishand, A., and van Deursen, W.: Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach, Hydrol. Earth Syst. Sci., 11, 1145–1159, doi:10.5194/hess-11-1145-2007, 2007.

Lettenmaier, D. P., Wood, A. W., Palmer, R. N., Wood, E. F., and Stakhiv, E. Z.: Water Resources Implications of Global Warming: A U.S. Regional Perspective, Clim. Change, 43, 537–579, doi:10.1023/A:1005448007910, 1999.

Li, C. Z., Zhang, L., Wang, H., Zhang, Y. Q., Yu, F. L., and Yan, D. H.: The transferability of hydrological models under nonstationary climatic conditions, Hydrol. Earth Syst. Sci., 16, 1239–1254, doi:10.5194/hess-16-1239-2012, 2012.

Liepert, B. G. and Previdi, M.: Inter-model variability and biases of the global water cycle in CMIP3 coupled climate models, Environ. Res. Lett., 7, 014006, doi:10.1088/1748-9326/7/1/014006, 2012.

Majda, A. J. and Gershgorin, B.: Quantifying uncertainty in climate change science through empirical information theory, Proc. Natl. Acad. Sci. USA, 107, 14958–14963, doi:10.1073/pnas.1007009107, 2010.

Maraun, D.: Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums, Geophys. Res. Lett., 39, L06706, doi:10.1029/2012GL051210, 2012.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M.,

Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 48, RG3003, doi:10.1029/2009RG000314, 2010.

Ménard, R.: Bias Estimation, in: Data Assimilation, edited by: Lahoz, W., Khattatov, B., and Menard, R., 113–135, Springer Berlin Heidelberg, available at: http://dx.doi.org/10.1007/978-3-540-74703-1_6, 2010.

Middelkoop, H., Daamen, K., Gellens, D., Grabs, W., Kwadijk, J. C. J., Lang, H., Parmet, B. W. A. H., Schädler, B., Schulla, J., and Wilke, K.: Impact of Climate Change on Hydrological Regimes and Water Resources Management in the Rhine Basin, Clim. Change, 4, 105–128, doi:10.1023/A:1010784727448, 2001.

Moore, K., Pierson, D., Pettersson, K., Schneiderman, E., and Samuelsson, P.: Effects of warmer world scenarios on hydrologic inputs to Lake Mälaren, Sweden and implications for nutrient loads, Hydrobiologia, 599, 191–199, doi:10.1007/s10750-007-9197-8, 2008.

Mpelasoka, F. S. and Chiew, F. H. .: Influence of rainfall scenario construction methods on runoff projections, J. Hydrometeorol., 10, 1168–1183, doi:10.1175/2009JHM1045.1, 2009.

Muerth, M. J., Gauvin St-Denis, B., Ricard, S., Velázquez, J. A., Schmid, J., Minville, M., Caya, D., Chaumont, D., Ludwig, R., and Turcotte, R.: On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff, Hydrol. Earth Syst. Sci., 17, 1189–1204, doi:10.5194/hess-17-1189-2013, 2013.

Palmer, T. N., Shutts, G. J., Hagedorn, R., Doblas-Reyes, F. J., Jung, T., and Leutbecher, M.: Representing Model Uncertainty In Weather And Climate Prediction, Annu. Rev. Earth Pl. Sc., 33, 163–C–5, doi:10.1146/annurev.earth.33.092203.122552, 2005.

Piani, C., Haerter, J. O., and Coppola, E.: Statistical bias correction for daily precipitation in regional climate models over Europe, Theor. Appl. Climatol., 99, 187–192, doi:10.1007/s00704-009-0134-9, 2010.

Räisänen, J. and Räty, O.: Projections of daily mean temperature variability in the future: cross-validation tests with ENSEMBLES regional climate simulations, Clim. Dyn., 1–16, doi:10.1007/s00382-012-1515-9, 2012.

Rasmussen, J., Sonnenborg, T. O., Stisen, S., Seaby, L. P., Christensen, B. S. B., and Hinsby, K.: Climate change effects on irrigation demands and minimum stream discharge: impact of bias-correction method, Hydrol. Earth Syst. Sci., 16, 4675–4691, doi:10.5194/hess-16-4675-2012, 2012.

Rojas, R., Feyen, L., Dosio, A., and Bavera, D.: Improving pan-european hydrological simulation of extreme events through statistical bias correction of RCM-driven climate simulations, Hydrol. Earth Syst. Sci., 15, 2599–2620, doi:10.5194/hess-15-2599-2011, 2011.

Salathé Jr., E. P.: Comparison of various precipitation downscaling methods for the simulation of streamflow in a rainshadow river basin, Int. J. Climatol., 23, 887–901, doi:10.1002/joc.922, 2003.

Schmidli, J., Frei, C., and Vidale, P. L.: Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods, Int. J. Climatol., 26, 679–689, doi:10.1002/joc.1287, 2006.

Seiller, G., Anctil, F., and Perrin, C.: Multimodel evaluation of twenty lumped hydrological models under contrasted

climate conditions, Hydrol. Earth Syst. Sci., 16, 1171–1189, doi:10.5194/hess-16-1171-2012, 2012.

Sennikovs, J. and Bethers, U.: Statistical downscaling method of regional climate model results for hydrological modelling, in: 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, edited by: Anderssen, R. S., Braddock, R. D., and Newham, L. T. H., 3962–3968, Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, Cairns, Australia, available at: http://www.mssanz.org.au/modsim09/I13/sennikovs.pdf (last access: 21 December 2012), 2009.

Shabalova, M. V., van Deursen, W. P., and Buishand, T. A.: Assessing future discharge of the river Rhine using regional climate model integrations and a hydrological model, Clim. Res., 23, 233–246, doi:10.3354/cr023233, 2003.

Sun, F., Roderick, M. L., Lim, W. H., and Farquhar, G. D.: Hydroclimatic projections for the Murray-Darling Basin based on an ensemble derived from Intergovernmental Panel on Climate Change AR4 climate models, Water Resour. Res., 47, W00G02, doi:10.1029/2010WR009829, 2011.

Terink, W., Hurkmans, R. T. W. L., Torfs, P. J. J. F., and Uijlenhoet, R.: Bias correction of temperature and precipitation data for regional climate model application to the Rhine basin, Hydrol. Earth Syst. Sci. Discuss., 6, 5377–5413, doi:10.5194/hessd-6-5377-2009, 2009.

Terink, W., Hurkmans, R. T. W. L., Torfs, P. J. J. F., and Uijlenhoet, R.: Evaluation of a bias correction method applied to downscaled precipitation and temperature reanalysis data for the Rhine basin, Hydrol. Earth Syst. Sci., 14, 687–703, doi:10.5194/hess-14-687-2010, 2010.

Teutschbein, C. and Seibert, J.: Regional Climate Models for Hydrological Impact Studies at the Catchment Scale: A Review of Recent Modeling Strategies, Geogr. Comp., 4, 834–860, doi:10.1111/j.1749-8198.2010.00357.x, 2010.

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

Teutschbein, C., Wetterhall, F., and Seibert, J.: Evaluation of different downscaling techniques for hydrological climate-change impact studies at the catchment scale, Clim. Dynam., 37, 2087–2105, doi:10.1007/s00382-010-0979-8, 2011.

Themeßl, M. J., Gobiet, A., and Leuprecht, A.: Empirical statistical downscaling and error correction of daily precipitation from regional climate models, Int. J. Climatol., 31, 1530–1544, doi:10.1002/joc.2168, 2011.

Tramblay, Y., Ruelland, D., Somot, S., Bouaicha, R., and Servat, E.: High-resolution Med-CORDEX regional climate model simulations for hydrological impact studies: a first evaluation of the ALADIN-Climate model in Morocco, Hydrol. Earth Syst. Sci., 17, 3721–3739, doi:10.5194/hess-17-3721-2013, 2013.

Van der Linden, P. and Mitchell, J. F. B.: ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project, Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK, available at: http://ensembles-eu.metoffice.com/docs/Ensembles_final_report_Nov09.pdf (last access: 21 December 2012), 2009.

Vannitsem, S.: Dynamical Properties of MOS Forecasts: Analysis of the ECMWF Operational Forecasting System, Weather Forecast., 23, 1032–1043, doi:10.1175/2008WAF2222126.1, 2008.

Varis, O., Kajander, T., and Lemmelä, R.: Climate and water: from climate models to water resources management and vice versa, Clim. Change, 66, 321–344, doi:10.1023/B:CLIM.0000044622.42657.d4, 2004.

Von Storch, H.: On the Use of "Inflation" in Statistical Downscaling, J. Climate, 12, 3505–3506, 1999.