Hydrol. Earth Syst. Sci., 16, 3391–3404, 2012
www.hydrol-earth-syst-sci.net/16/3391/2012/
doi:10.5194/hess-16-3391-2012

Hydrology and
Earth System
Sciences

HESS Opinions
"Should we apply bias correction to global and regional climate model data?"

U. Ehret[1], **E. Zehe**[1], **V. Wulfmeyer**[2], **K. Warrach-Sagi**[2], **and J. Liebert**[1]

[1]Institute of Water Resources and River Basin Management, Karlsruhe Institute of Technology KIT, Karlsruhe, Germany
[2]Institute of Physics and Meteorology, University of Hohenheim, Hohenheim, Germany

Correspondence to: U. Ehret (u.ehret@kit.edu)

Received: 5 April 2012 – Published in Hydrol. Earth Syst. Sci. Discuss.: 20 April 2012
Revised: 28 August 2012 – Accepted: 28 August 2012 – Published: 21 September 2012

Abstract. Despite considerable progress in recent years, output of both global and regional circulation models is still afflicted with biases to a degree that precludes its direct use, especially in climate change impact studies. This is well known, and to overcome this problem, bias correction (BC; i.e. the correction of model output towards observations in a post-processing step) has now become a standard procedure in climate change impact studies. In this paper we argue that BC is currently often used in an invalid way: it is added to the GCM/RCM model chain without sufficient proof that the consistency of the latter (i.e. the agreement between model dynamics/model output and our judgement) as well as the generality of its applicability increases. BC methods often impair the advantages of circulation models by altering spatiotemporal field consistency, relations among variables and by violating conservation principles. Currently used BC methods largely neglect feedback mechanisms, and it is unclear whether they are time-invariant under climate change conditions. Applying BC increases agreement of climate model output with observations in hindcasts and hence narrows the uncertainty range of simulations and predictions without, however, providing a satisfactory physical justification. This is in most cases not transparent to the end user. We argue that this hides rather than reduces uncertainty, which may lead to avoidable forejudging of end users and decision makers.

We present here a brief overview of state-of-the-art bias correction methods, discuss the related assumptions and implications, draw conclusions on the validity of bias correction and propose ways to cope with biased output of circulation models in the short term and how to reduce the bias in the long term. The most promising strategy for improved future global and regional circulation model simulations is the increase in model resolution to the convection-permitting scale in combination with ensemble predictions based on sophisticated approaches for ensemble perturbation.

With this article, we advocate communicating the entire uncertainty range associated with climate change predictions openly and hope to stimulate a lively discussion on bias correction among the atmospheric and hydrological community and end users of climate change impact studies.

1 Introduction

Understanding and quantifying the causes and effects of climate change is currently one of the most challenging obstacles in science and of high relevance for society. Today, besides observations, among the best (but certainly not perfect) tools we have to understand Earth's climate dynamics and evolution are global circulation models (GCMs). Confidence in the fidelity of predictions by such models comes from several sources (Randall et al., 2007): firstly, model fundamentals are based on established physical laws, such as conservation of mass, energy and momentum, and process insight comes from a wealth of observations. Secondly, the models are able to simulate important aspects of the current climate, among them many patterns of climate variability observed across a range of time scales such as the seasonal shifts of temperatures, storm tracks or rain belts. Further, the models have proven their ability to reproduce features of past climates and climate changes. Finally, on large spatial and temporal aggregation scales (global, multi-annual) and especially for projections of temperature changes, most models point in the same direction.

However, for most hydrologically relevant variables, GCMs currently do not provide reliable information on scales below about 200 km (Maraun et al., 2010). This is too coarse for a realistic representation of most hydrological processes that act over a large range and down to very fine scales (Blöschl and Sivapalan, 1995; Kundzewicz et al., 2007). This is especially true for the main driver of hydrological processes – precipitation. The resolution of GCMs precludes the simulation of realistic circulation patterns that lead to extreme rainfall events (Kundzewicz et al., 2007), and for hydrological simulations and predictions to become reliable on relevant scales, precipitation input needs to be realistic, not only with respect to the mean but also with respect to intensity (especially extremes), intermittency (Ines and Hansen, 2006) as well as temporal and spatial variability across regions and seasons (Maraun et al., 2010). GCM output is thus currently an inadequate basis for reliable hydrological predictions of climate change impact on scales relevant for decision makers. The same applies to regional agricultural studies (Ines and Hansen, 2006).

One avenue to close this scale gap is stochastic downscaling. Stochastic downscaling establishes a functional relationship between the most robust and reliable fields provided by GCMs such as geopotential height or temperature and locally observed meteorological variables such as precipitation or temperature in a region of interest (e.g. Wójcik and Buishand, 2003; Burger, 1996; Stehlik and Bárdossy, 2002).

A physically more consistent approach to overcome this scale mismatch is dynamical downscaling: a high-resolution (typically 10–50 km) regional circulation model (RCM) is nested into a GCM, which provides the forcing at the boundaries. Due to the higher resolution and a more complete representation of physical processes in RCMs, this can considerably improve simulations and projections of regional-scale climate (Maraun et al., 2010). Applying RCMs has the greatest potential to improve rainfall simulations when the forcing is mainly regional. In the case of large-scale forcing (such as propagation of frontal systems), the quality achievable by the RCM will inevitably be limited by the quality of the boundary conditions provided by the GCM (Wulfmeyer et al., 2011). Often, the output of RCMs is then used in impact models such as hydrological models (HMs).

However, despite considerable progress in recent years, reproduction of hydrologically relevant variables in present-day climate on appropriate scales based on GCM-RCM model chains is still afflicted with systematic errors (bias) to a degree that precludes their direct interpretation or application for simulation and prediction in HMs. This is well known and has been recognized by many authors, e.g. Wilby et al. (2000), Wood et al. (2004), Randall et al. (2007), Piani et al. (2010), Hagemann et al. (2011), Chen et al. (2011), Rojas et al. (2011), Haddeland et al. (2012), Johnson and Sharma (2012). To overcome this problem, post-processing of either GCM or RCM output by correcting with and towards observations has become a standard procedure in climate change impact studies (CCIS). This bias correction (BC) procedure significantly alters the model output and therefore influences the results of all CCIS relying on bias-corrected data.

Based on this, the main question we pursue in this article is whether and when the application of BC methods, which often, unlike the other components of the modelling chain for CCIS (GCMs, RCMs and HMs), lack a sound physical basis, is justified or not. To this end, we start with a definition of bias and present an overview of its causes and typical magnitudes in Sect. 2. We continue in Sect. 3 by presenting approaches to deal with biased model output with a focus on BC and reflect why BC, despite its known deficits, is nevertheless routinely applied. In Sect. 4 we present a brief overview of state-of-the-art BC methods. Based on this, we discuss BC with respect to the assumptions made when applying it and reflect on its implications in Sect. 5. It is a matter of on-going scientific discussion whether these assumptions are really satisfied and thus whether and when the application of BC is justified or not. We complete Sect. 5 by presenting an overview of opinions from current literature and formulate our own reservations with BC. Finally, we propose ways to cope with biased model output from GCMs and RCMs in the short term and how to reduce the bias in the long term in Sect. 6 and draw final conclusions in Sect. 7.

2 Model bias: definition, causes and magnitude

2.1 Definition

When we say bias, what do we mean? The international definition of bias according to WMO (WWRP 2009-1, 2009) is the correspondence between a mean forecast and mean observation averaged over a certain domain and time. According to the recommendation of the Joint Working Group on Forecast Verification Research (JWGFVR), the comparison should be performed between gridded data sets (WWRP 2009-1, 2009), with the grid resolution of the models degraded by a factor of 3–4 to take into account numerical filter effects (see e.g. Bauer et al., 2011).

However, in the context of CCIS, the definition of bias is not as strict: it varies with the scope of the studies and is often used in a general sense for addressing any deviation of interest (e.g. with respect to the mean, variance, covariance, length of dry spells, etc.) of the model from the corresponding “true” value. Typically, biases are calculated for precipitation or temperature on continental, river basin or model grid scale for annual, seasonal, or monthly aggregations. Unlike weather forecast verification, where atmospheric variables are averaged over short times scales and thus allow the analysis of individual events, climate models cannot be verified for single cases. Instead, their ability to reproduce climate variability is analysed, and typically averaged over the order of ten years. Maraun et al. (2010) give an overview of metrics to validate GCM/RCM output. Chen et al. (2011) and Haerter et al. (2011) define bias as the time-independent component of the model error, i.e. the portion of the error that occurs at all times. However, it should be kept in mind that, as the bias is a result of a dynamic model error chain, it will always be a combination of time-variant errors.

Throughout this text, we will stick to the broad definition of bias established in the CCIS community, i.e. we will use "bias" for any discrepancy of interest between a model (GCM, RCM or HM) output characteristic and the "truth". However, for the future we strongly suggest that the use of "bias" should be narrowed again to the WMO definition (see also Sect. 6.1).

2.2 Causes

The most obvious reasons for biased model output are imperfect model representations of atmospheric physics (Maraun, 2012), incorrect initialization of the model or errors in the parameterization chain: with respect to GCMs, it is currently subject of intense discussion whether better initialization of the state of the oceans and the land surface leads to an improvement of simulations beyond decades. The process chain leading to the model climate depends on the parameterization of various processes of all compartments of the Earth system including the cryosphere, the hydrosphere and the biosphere as well as the atmosphere with its fine-scale, complex turbulent and aerosol-cloud-precipitation microphysics. It is likely that strong deficiencies still exist with respect to the simulation of the cryosphere, the water cycle over the land surface which is controlled by soil and vegetation properties and the corresponding energy balance closure as well as the parameterization of aerosol-cloud-precipitation microphysics (e.g. Doherty et al., 2009; WCRP, 2009).

With respect to RCMs, errors can be introduced by incorrect boundaries provided by reanalyses or GCMs or inconsistencies between the physics of GCMs and RCMs. Furthermore, in spite of the higher resolution of RCMs, several deficiencies remain with respect to the parameterizations. There are strong indications that the main errors in state-of-the-art RCMs are due to incorrect energy balance closure, its feedback to the convective and stable atmospheric boundary layer and the resulting formation of clouds and precipitation, which is strongly controlled by the choice of the microphysical scheme. Furthermore, with respect to precipitation, it is important to consider that the overall bias depends on a time-variant combination of effects leading to precipitation events involving different combinations of model physics.

Within the WWRP projects D-PHASE (Rotach et al., 2009) and COPS (Wulfmeyer et al., 2011), a forcing concept was developed resulting in the following understanding of model errors: if large-scale forcing is present, the main error is driven by GCM boundaries but the fine structure of errors down to the scale of catchments is still influenced by local forcing (land-surface heterogeneity and orography). The importance of local forcing increases from weakly forced conditions (no surface front but upper level instability) to local forcing where convection and precipitation are initiated by orography and/or land-surface heterogeneity. It is clear that the models must be able to simulate the statistics of precipitation depending on the combination of forcing conditions.

Another source of bias that applies to both GCMs and RCMs is climate variability: models are parameterized and evaluated on finite-length time series which may not cover the full range of atmospheric dynamics. This makes them subject to sampling uncertainty or bias. This applies even more to the parameterization of BC methods (Maraun, 2012).

Further, apparent model biases can occur if the reference data sets (the "truth") used for model parameterization and validation are inadequate. On smaller scales, high quality observation-derived data sets such as E-OBS (Haylock et al., 2008) are available, which may be biased due to non-representativeness of the underlying observations. On larger scales, it is mainly only reanalysis data such as the WATCH data set (Weedon et al., 2011), NCEP/NCAR or ERA-interim (Dee et al., 2011) that are available. They are in turn subject to model biases and can significantly deviate from the true weather (Maraun et al., 2010). It is therefore important to develop and validate new high-resolution observation-based reference data sets by exploiting the full range of available observations.

HMs using output from RCMs add further sources of bias: RCMs contain hydrological components to calculate land surface-atmosphere interaction. If the RCM output is used in a HM, an assumption is made on the interchangeability of the two hydrological schemes, i.e. comparability of their land-atmosphere feedback functioning. This is usually not fulfilled (Rojas et al., 2011; see also Sect. 5.1). Also, biases occur if the spatial or temporal resolution of the GCM/RCM input for the HMs is inadequate (Hay et al., 2002). HMs are usually calibrated on interpolated meteorological point observations and observed streamflow. Thus, the models are tuned to reproduce streamflow based on biased input (smooth fields based on sparse data). When changing the input to gridded RCM fields, this model configuration will likely create a biased output, as it still compensates "for the old bias".

For hydrological CCIS, perhaps the most troublesome systematic biases are those in precipitation: "*The biases ordinarily present in hydrological output from GCMs affect all aspects of the intensity spectrum. Simulated precipitation statistics are generally affected by a positive bias in the number of wet days, which is partly compensated by an excessive number of occurrences of drizzle, a bias in the mean, the standard deviation (variability), and the inability to reproduce extreme events*" (Piani et al., 2010). This was also reported by many others, e.g. Stephens et al. (2010), Sun et al. (2006). Specifically for Europe, Christensen et al. (2008) and Dosio and Paruolo (2011) report that wintertime precipitation is generally too abundant. A comprehensive overview of systematic errors in present-day RCMs can be found in Rojas et al. (2011).

2.3 Magnitude

In this section, we will illustrate the magnitude of biases (and with it the magnitude of BC impact by removing them) in the GCM/RCM/HM chain with examples reported in the literature and from own studies. Johnson and Sharma (2012) compared raw output from a GCM (CSIRO Mk3.5) and RCM (MIROC) with observations: in interior Australia, both models over-predicted annual rainfall by up to 200 %, but under-predicted along the coasts. Rojas et al. (2011) found that averaged annual precipitation simulated by the HIRHAM 5 RCM over Europe in the control period 1961–2000 almost doubled the observed measurements. Hagemann et al. (2011) reported, from a study applying three GCMs, two emission scenarios and two global hydrological models (GHMs) that "*for some regions, the impact of the bias correction on the climate change signal may be larger than the signal itself, thereby identifying another level of uncertainty that is comparable in magnitude to the uncertainty related to the choice of the GCM or GHM*." Sun et al. (2011) investigated the influence of BC on the mean and spread of a 39 model ensemble on gridded annual precipitation in the Murray-Darling basin (Australia): BC changed the ensemble mean by 17.7 % and the ensemble spread by 122 % (relative to the observation). Sharma et al. (2007) compared mean monthly rainfall amounts from a GCM (ECHAM4) with spatially interpolated observations on model grid scale: BC changed the correlation between observations and raw GCM output from 0.32 to 0.66, i.e. it caused a relative change of 48 %. Likewise, the root-mean-square error (RMSE) was changed by 56 % (from 3.64 mm to 2.06 mm). This also had a noticeable impact on discharge simulations (Ping river basin, Thailand, 34 453 km^2): the relative RMSE changed by 54 % (from 172 to 93 $\mathrm{m}^3\,\mathrm{s}^{-1}$). On the other hand, the influence of climatic variability seems to be less prominent. Chen et al. (2011) compared the relative contribution of GCM, emission scenario, period for bias correction and inter-annual variability to the uncertainty of hydrological climate impact studies. They concluded that "*the choice of different decadal periods over which to derive the bias correction parameters is a source of comparatively minor uncertainty compared to the choice of GCM, SRES scenario and the natural inter-annual variability*." In the recently conducted study "Flood hazards in a changing climate" (Schädler et al., 2012), climate change impact on flood magnitudes was analysed in a multi-model study including two GCMs, two RCMs, three HMs in three mesoscale catchments in Germany. The GCM/RCM/HM model chain was applied to the reference period 1971–2000, and monthly mean flood magnitudes were calculated. Here we discuss the results at the example of gauge Wetter/Ruhr (3908 km^2). The flood magnitudes were afflicted with strong biases (for scenarios with the RCM "CLM" on average 168 % relative to the observations). To reduce them, BC was applied to precipitation and temperature of the RCM. The effect on the mean monthly flood magnitudes (i.e. the difference in the flood magnitudes with and without bias correction relative to the observed ones) was in the range of 23–181 %, again evaluated in the observation period.

The main point we want to make in this section is that, just as model biases can be on an order of magnitude that precludes the direct use of model output in CCIS, the impact of any BC method that corrects for this bias is of equal magnitude. Hence, BC will have a large influence on the GCM/RCM/HM output in absolute terms and likely also on climate change signals (i.e. the relative change between a control and prediction period). However, this impact of BC is only very rarely explicitly quantified and made transparent in CCIS; also the crucial assumption – stationarity of the BC method under non-stationary conditions – is often not critically discussed.

3 Hiding model bias through bias correction

As discussed in the introduction, the problem of biased GCM/RCM output is well known, and considerable efforts have been made to tackle this problem. We broadly classify them into three approaches.

The first is to *reduce the bias* by improving the models, addressing the deficiencies as outlined in Sect. 2.2. This is the most difficult but, in the long term, the most promising and potentially reliable approach as it is tied directly to the physical model basis. This approach will be discussed in detail in Sect. 6.3.

As a complete removal of bias is likely not possible by a single deterministic model, this step needs to be combined with the development of multi-model ensembles for GCMs, RCMs as well as HMs. The ensemble spread is essential to *quantify the uncertainty* associated with CCIS results. This approach is currently subject to intense research and promises considerable improvement in the mid-term. We discuss this in more detail in Sect. 6.2.

Our focus in this section is on the third approach, namely the correction of model output in a post-processing step. Post-processing can *reduce GCM/RCM bias* and can be regarded as a valid part of the model chain when it meets the requirements we impose on the incorporation of any model component (including the process descriptions inside the models): they should increase consistency (correspondence between model dynamics/output and our judgement), quality (correspondence between model output and observations) and value (benefit of model output to users) of the model (see

Murphy, 1993). Consistency relates to the agreement of the model component with our understanding of the functioning of the system under consideration. For a model component to increase consistency, we should be sure that it is generally applicable, i.e. it should work under the full range of possible boundary conditions and model states. As an example, let us assume a thermometer (our model) that we know has a constant bias of −3 K. Adding a bias correction (add +3 K to the thermometer reading) in a post-processing step would be in full agreement with the three requirements. However, if we had only one pair of model output (the thermometer reading) and the corresponding true value, e.g. 3 °C and 6 °C, we could not be sure whether the correction should be "reading +3 K" or "reading *2". Applying either of the corrections on the single set of reading plus true value would increase quality and value. However, we could not be sure whether this would still hold for other value pairs. The correction would thus not increase consistency and possibly hide (overestimate) the true quality and value of the model. The latter case, in our view, often applies to the way BC methods are currently applied in CCIS, which *hide biases* of the GCM/RCM output from subsequent users. BC methods in this context are usually either applied in combination with a downscaling procedure or on the scale of the model output and are also referred to as model output statistics (MOS). In this paper, we will, in line with the broad definition of "bias" in Sect. 2.1, refer to it as statistical bias correction or simply bias correction (BC). For a good overview and also classification of different approaches, see e.g. Maraun et al. (2010) or Themeßl et al. (2011). Note that, in this paper, we exclude the field of empirical-statistical downscaling (Wilby and Wigley, 1997) as used in perfect prognosis approaches as there the intention is to downscale large-scale data rather than correcting model errors.

A typical modelling chain for hydrological CCIS thus comprises GCM output used in an RCM, whose output is then bias-corrected and applied to a HM. Unlike the other components, most BC methods lack a sound physical basis; they usually do not satisfy conservation laws and are not a model of the physical world in itself (Haerter et al., 2011). This makes their application more questionable than the other components. Why is it used then or has been introduced in the first place? Essentially it is a quick fix that was "*born under the pressure to get answers on the potential impact of climate changes on our society*" (Vannitsem, 2011) and, as a consequence, from the necessity to make biased GCM-RCM output usable for interpretation or further use in HMs.

Compared to the other approaches to tackle the problem of biased model output as described at the beginning of this section, BC has, from the user perspective, several advantages: as BC methods act on model output, they can be developed and applied by any potential user without the need for full insight into the generating model, tailored to the variable and application of interest with manageable effort (compared to the efforts to advance GCMs or RCMs). In line with this, Johnson and Sharma (2012) list a number of reasons that make BC attractive: ease of application, ability to allow future changes in variability (unlike scaling methods), and flexibility to correct the GCM simulations for the parameters of interest. As another advantage, Li et al. (2010) mention the lower computational requirements compared to alternatives based on dynamical models.

In that sense, the range of existing BC methods (see Sect. 4) reflects the range of GCM/RCM model deficiencies in reproducing present-day climate from the user perspective. Many BC methods have therefore been developed more from the perspective of necessity rather than validity.

4 Bias correction methods

BC methods have been developed and applied by many users of GCM/RCM output and for various purposes. The following list of BC methods is far from being complete and should rather be understood as to give the reader a taste of the range and approaches of BC (a more complete overview can be found e.g. in Themeßl et al., 2011): monthly mean correction (Fowler and Kilsby, 2007), delta change method (Hay et al., 2000), multiple linear regression (Hay and Clark, 2003), analog methods (Moron et al., 2008), local intensity scaling (Schmidli et al., 2006), quantile mapping (Wood et al., 2004; Sun et al., 2011), fitted histogram equalization (Piani et al., 2010), and gamma-gamma transformation (Sharma et al., 2007).

In recent years, BC methods have evolved from time-averaged corrections of mean precipitation and temperature towards more advanced methods that correct higher distribution moments (Piani et al., 2010), include further variables such as radiation, humidity and wind (Haddeland et al., 2012), allow for time-dependent model biases (Buser et al., 2009; Li et al., 2010) or correct model output hierarchically on several nested time scales (Haerter et al., 2011; Johnson and Sharma, 2012).

Most BC methods consist of comparable steps which we will briefly present here as the example of the fitted histogram equalization approach as proposed by Piani et al. (2010): after matching the resolution of the model and the reference, outliers are excluded and the remaining values of both the GCM and baseline fields are ordered by magnitude. The obtained probability density function of the model data is then mapped onto that of the observations. This empirical transfer function constitutes the BC and acts on all moments of the distribution. The transfer functions are determined separately for each calendar month, grid point and variable.

The important point here is that BC is carried out separately across time, space and variable, a characteristic shared by most of the current BC approaches. Doing so implies several strong assumptions which affect the applicability of BC.

5 Applicability of bias correction

Here we will discuss which assumptions are taken when applying BC methods and what the related implications are. After this, we will review current literature for statements about the applicability of BC and finally draw our own conclusions.

5.1 Assumptions and implications of bias correction

Due to the variety of existing BC approaches, not all assumptions and implications listed below apply to all methods. Therefore, the list should be seen as a general overview.

- *Reliability*:
  The assumption is, plainly spoken, that a GCM/RCM, with such obvious deficiencies that BC is required, is nevertheless suitable to predict the (sometimes subtle) effects of climate change (see also discussion in Sect. 1).

- *Effectiveness*:
  The assumption is that the chosen BC method is effective, i.e. that it sufficiently corrects all biases of interest without introducing unwanted side effects (other biases). However, Chen et al. (2011) report that the choice of the BC method may be another source of uncertainty. Along the same line, Haerter et al. (2011) found that "*the consequences of choosing a certain bias-correction method are much more dramatic in the case of precipitation than in the case of temperature*." In one of the few studies applying multiple BC techniques, Teutschbein et al. (2011) found that "*the choice of downscaled precipitation (authors note: based on different BC techniques) time series had a major impact on the streamflow simulations*."

- *Time invariance:*
  The assumption is that the selected BC method, parameterized on a finite period of time for a finite size region, also holds under varying forcing and extreme climate conditions.

  However, this is likely not generally valid: Christensen et al. (2008) report on possible nonlinear characteristics of model biases as a function of increasing temperatures or precipitation amounts. Hagemann et al. (2011) showed that BC can alter the climate change signal for specific locations and months and that BC will lead to changes in the climate change signals if low precipitation amounts (or temperatures) are differently corrected as high amounts or if the distribution between low and high amounts changes in a future climate. Maraun (2012) investigated possible bias non-stationarity in a pseudo-reality approach. He defined different types of bias non-stationarities and distinguished between apparent and real non-stationarities. He could not identify any non-stationarities due to changing relative occurrences of weather types, but found considerable bias changes due to different climate sensitivities, and apparent bias changes due to sampling variability. Similarly, Vannitsem (2011) used artificial reality approaches (scalar systems and a low-order model of moist general circulation) to examine BC properties under transient conditions. For the first, the main finding was that the quality of BC was specific to the system and the model error source, thus precluding the possibility to deduce universal evolution relations. For the latter, the main finding was that "*systematic correction associated with the presence of model errors cannot be straightforwardly transposed from one climate condition to another*." Buser et al. (2009), upon developing a BC method that explicitly allows for the bias to vary with time, stated that "*the problem remains to make assumptions on the nature of the change*" and that "*depending on the assumptions made, the climate change signal may differ considerably*." The authors conclude that "*the aforementioned result is of general interest, as it questions an important implicit assumption of current scenario models, namely that the model bias will not significantly depend upon the climate state*." Finally, Terink et al. (2010) applied reanalysis data to 134 subbasins of the Rhine River and evaluated BC in a split sampling approach. For the validation period, they found that, while temperature was corrected very well, results for precipitation with BC were worse than without.

- *Completeness*:
  Closely connected to the assumption of time invariance as discussed above is the assumption that the finite length control period used to derive BC parameters (e.g. transfer functions) covers the entire spectrum of the variable of interest. However, especially for short control periods, this is not fulfilled. This implies that applying the BC method to predicted values outside the observed range requires an extrapolation of the transfer function beyond the observed range and may lead to bias-correction of GCM/RCM output beyond physical limits. Maraun et al. (2010) present a brief overview on approaches to address this problem.

- *Minor role of spatiotemporal field covariance*:
  BC is in most approaches parameterized and applied individually for finite size regions (e.g. grid cells) of the domain of interest. In general, this alters the spatiotemporal covariance structure of the respective GCM/RCM field and thus impairs the main advantage of dynamic models, which is to create thermodynamic fields with covariance structures that are consistent with atmospheric physics. From a hydrological point of view, changes in the covariance structure may strongly affect hydrological functioning whenever non-linear processes are involved, e.g. surface runoff generation or

macropore flow initiation. Applying BC methods assumes that the effect of spatiotemporal field covariance (e.g. the direction and magnitude of temperature gradients or the length of dry spells) is either not significantly affected by BC or of minor importance, which may not always hold (Johnson and Sharma, 2012).

- *Minor role of feedbacks among variables*:
The assumption is that the links and feedbacks between the meteorological states and fluxes (temperature, humidity, precipitations, evapotranspiration, etc.) are not of key importance, i.e. the resulting fields can be corrected after, not during, modelling the related processes. On this topic, Seneviratne et al. (2006) conclude from a climate change study in Europe that "*the most striking result of our analysis is that land–atmosphere coupling is significantly affected by global warming and is itself a key player for climate change*." Further, they summarize that their "*investigation reveals how profoundly greenhouse gas forcing may affect the functioning of the regional climate system and the role of land-surface processes*." Berg et al. (2009) showed that daily precipitation exhibits some scaling with temperature. Piani et al. (2010) pointed out that "*any bias correction involving multiple fields induces changes in the correlation of such fields and that the relationship between precipitation and temperature depends on the geographical region and the time period and area over which precipitation is averaged*." Furthermore, they conclude that "*the question is not settled whether the statistical relationship can be applied to future changes in global surface temperature*." Along this line, Johnson and Sharma (2012) report from a study conducted in Australia that "*there are clearly significant correlations between temperature and precipitation, particularly at (...) longer time scales*." According to Wood et al. (2004), this may have noticeable impact on processes like evapotranspiration or snowmelt. Haddeland et al. (2012) shed light on the (in addition to precipitation and temperature) significant role of radiation, humidity and wind when simulating the terrestrial water balance especially in energy-limited areas. These variables are all dynamically coupled by various feedback processes.

- *Comparable bias behaviour of GCM/RCM atmospheric fields and fields related to terrestrial hydrology*:
From the output of GCM/RCM systems, usually fields of direct interest and fields required as input for further models (such as HMs) are evaluated and bias corrected. This includes rainfall, temperature, relative humidity, wind, radiation, etc., but rarely fields of terrestrial hydrology, although any GCM/RCM contains land surface models (LSMs) that include terrestrial hydrological processes such as surface and subsurface runoff production. The reason is the usually very simple representation of these processes, resulting in poor agreement with observations (Rojas et al., 2011). This can partly be explained by the fact that the main focus of LSMs in GCMs/RCMs is on the influence of the water balance on surface heat fluxes (and not discharge calculation; van den Hurk et al., 2005), while the focus of HMs is terrestrial water availability and use. LSMs typically solve the water and energy balance, while HMs typically only solve the water balance (Haddeland et al., 2011).

  Thus, if the stationarity of BC methods is tested, this is usually done for meteorological fields, but not so often for discharge, the primary quantity of interest of terrestrial hydrology. It is now possible to imagine that, for meteorological fields, the bias may be found to be sufficiently stationary to make them acceptable for CCIS and that this is extrapolated to fields of terrestrial hydrology. However, due to the strongly non-linear nature of terrestrial hydrological processes, it may well be that small bias instationarities in the meteorological forcing may be amplified to large bias instationarities of terrestrial hydrological variables. This can due to the usually simple representation of runoff-formation processes not being evaluated in the GCM/LSM system itself, but must be done with the output of the HM.

- *No bias due to offsets*:
Many existing BC methods identify bias by comparing model output and observations for identical regions in space and identical points of time during a reference period. This implies that any model deficiency that manifests as spatial or temporal offset is falsely recognized as an offset in magnitude, i.e. a bias (Haerter et al., 2011).

- *Bias can be associated with typical timescales*:
Many existing BC methods determine and correct bias for one (or a few) aggregation times of interest (season, month), thus assuming that bias occurs mainly and can be attributed to effects at this selected time scale. However, Haerter et al. (2011) argue that "*fluctuations on different scales (caused by disparate physical mechanisms) can mix and lead to unexpected and unwanted behavior in the corrected time series and blur the interpretation for future scenario corrections*." In support, they present an example where bias correction based on daily temperature led to an improvement of the day-to-day variance, but the variance of the monthly means in fact became less realistic after performing the bias correction. On the other hand, Rojas et al. (2011) found that BC of temperature based on monthly transfer functions fully preserved observed annual and seasonal statistics.

5.2 Conclusions on the applicability of bias correction

The range of existing BC methods as outlined in Sect. 4 reflects the user perspective of deficits of GCM/RCM models to reproduce present-day and to predict future climate. In general, the biases corrected for are a function of time, space and meteorological variable and spread in a non-uniform way through the entire distribution of the variables. The biases also manifest in the characteristics of spatiotemporal field covariance. In short, the bias structure is complex, which is a direct result of the complex nature of hydro-meteorological atmospheric and land-surface process interactions. The question is then whether or not the application of BC, which is essentially a post-processing step neglecting these complex interactions, is suitable and valid to make GCM/RCM output usable for CCIS. This is increasingly discussed in the scientific community. Hagemann et al. (2011) conclude that "*it is rather difficult to judge whether the impact of the bias correction on the climate change signal leads to a more realistic signal or not*"; Vannitsem (2011) wonders "*whether this type of post processing can still be used in the context of a transient climate, in particular in the context of decadal forecasts. The obvious answer would be no in a strict sense since modifications of external parameters generically imply modifications of the variability of the system*". Haerter et al. (2011) formulate limitations to the application of BC: (i) at every grid box where BC is to be applied, it must be ensured that the model provides a realistic representation of the physical processes involved; (ii) quantitative discrepancies between the modelled and observed probability density function of the quantity at hand must be constant in time; (iii) BC cannot improve the representation of fundamentally misrepresented physical processes; (iv) only when short-term and long-term fluctuations are aligned, the bias correction will lead to improvements on both timescales. Teutschbein and Seibert (2010) generally recommend the application of bias-correction methods but warn that "*the need for bias corrections adds significantly to uncertainties in modelling climate change impacts*."

Let us go back once more to the core of most CCIS, the GCM/RCM/HM model chain: most of the confidence we have in them comes from the fact that the models are based upon established physical-chemical laws, their capability to produce thermodynamic fields with a spatiotemporal correlation structure consistent with atmospheric physics and their inherent consideration of various feedback processes. This is especially important for hydrological considerations, as hydro-meteorological atmospheric and land-surface processes interactions are complex and non-negligible. BC impairs these advantages by altering spatiotemporal field consistency, relations among variables and conservation principles. In addition, it remains doubtful that BC methods parameterized on observed climate will hold under changing climate conditions.

Further we ask what can be gained from advancing BC: let us extrapolate the current evolution of bias correction from simple towards more complex methods (see Sect. 4). If we arrive at the perfect BC method correcting at high spatial and temporal resolution all moments of the variable of interest, assure consistency over many spatial and temporal scales as well as inter-field correlations, discriminate between different weather situations, allow for the bias to be time-transient and include feedback effects, then we inevitably arrive at a complexity of the BC method comparable to the GCM or RCM itself, but still lack the physical justification of the latter. This will limit our confidence in climate change predictions involving BC.

Applying BC on GCM/RCM output (by definition) increases agreement with observations and hence narrows the uncertainty range of simulations and predictions, without however providing a satisfactory physical justification. This is in most cases not transparent to the end user. We argue that this hides rather than reduces uncertainty, which may lead to avoidable forejudging by end users and decision makers.

Our last argument relates to hydrology-related outcomes of CCIS based on GCM/RCM/HM model chains such as future flood or drought characteristics: instead of bias-correcting the meteorological forcing, a logical step would be to simply bias-correct the outcome of the HMs, e.g. discharge simulations and predictions. Applying this "end-of-pipe" bias-correction would be based on the same justification as BC of GCM/RCM output, but we dare say that it would not be accepted by hydrologically educated end users, at least not without an explicit knowledge of the impact of BC on the result.

In short, we conclude that BC is currently often used in an invalid way: it is added to the GCM/RCM model chain without sufficient proof that the consistency of the latter, i.e. the agreement between model dynamics/output, and our judgement and the generality of its applicability increases.

6 Ways forward: proposals on how to use and how to avoid bias correction

Notwithstanding the reservations we have with current BC practice, providing answers on climate change impact remains an urgent task, and the deficiencies of present-day GCMs and RCMs that prepared the grounds for BC in the first place do not vanish by criticizing the shortcomings of BC either. In the following section, we therefore propose ways forward to cope with and to reduce the bias associated with output of GCMs and RCMs for CCIS.

6.1 Proposals for the short term

The first and easiest task to accomplish is to openly communicate to the end user the impact of BC and the uncertainties associated with it by

- providing all results of any impact study for both *bias-corrected AND non-corrected input*, for the hindcast period and the projection, along with a detailed explanation of the BC method. From the spread of the results in the hindcast period and the projection, the impact of

BC must therefore be made comprehensible to any end user. For non-expert end users, it may be better to avoid publication of the bias-corrected results altogether.

- Further, to avoid confusion, we strongly suggest restricting the use of the term "bias" to the definition given by WMO (WWRP 2009-1, 2009) (see Sect. 2.1). Any other discrepancy of interest between a model result and the related observation/reference should be named differently (e.g. mean difference of the variance, etc.).

These steps will not lead to less biased GCM/RCM output; however they will contribute to the quantification of bias and to raising its awareness among end users. Maraun et al. (2010) stated with respect to end user needs for downscaled precipitation that "*as well as the product, the end user might also require a clear statement of the assumptions involved and limitations of the downscaling procedure, a transparent explanation of the method, a description of the driving variables used in the downscaling procedure and their source, a clear statement of the validation method and performance, and some characterization of the uncertainty or reliability of the supplied data*." We agree and suggest that the same also holds for BC methods.

6.2 Proposals for the mid term

The second set of proposals, namely the use of nested GCM/RCM approaches and the use of multi-model ensembles, is already subject of intense research (see also Sect. 3):

- *Nested approaches* (i.e. the use of RCMs to downscale GCM output) have already proven their potential to improve the quality of regional climate simulations and climate change predictions depending upon forcing conditions. Improvements can be attributed to the higher spatial resolution and hence a better description of orographic effects, land/sea contrast, land surface characteristics (Maraun et al., 2010) and especially to the move from parameterized to explicit representation of convection. RCMs also contain (compared to GCMs) better representations of fine-scale physical and dynamical processes including feedback processes which leads to a more realistic regional redistribution of mass, energy and momentum, e.g. in the form of mesoscale circulation patterns which are absent in GCMs (Maraun et al., 2010; Liang et al., 2008).

- *Multi-model ensembles* provide an ensemble of simulations and predictions either by the use of several models for some or all components of the modelling chain (GCM/RCM/HM) and/or by using ensembles of perturbed initial conditions or model parameterizations. Ensemble approaches help to quantify uncertainty of CCIS through the ensemble spread (e.g. Knutti, 2008). They are also useful to attribute uncertainty to different components of the modelling chain and natural variability (Maraun et al., 2010; Teutschbein and Seibert, 2010). With respect to uncertainty quantification, many projects such as ENSEMBLES (Christensen et al., 2008), PRUDENCE (Christensen and Christensen, 2007) and among many others, Wilby (2010), Ott et al. (2012), Schädler et al. (2012) or Sun et al. (2011) promote the use of model ensembles to avoid non-representativeness of the sample. Currently within the **Co**ordinated **R**egional climate Downscaling **Ex**periment (CORDEX) (Giorgi et al., 2009), high-resolution (50 km, 25 km and – for Europe – 11 km) ensembles and comparisons of regional climate simulations are underway for all continents, forced with the most recent re-analysis data (ERA-interim) and GCM data from CMIP5 for the IPCC-AR5 report (e.g. Warrach-Sagi et al., 2012). Haddeland et al. (2011) highlighted that ensemble approaches should also include HMs as they contribute considerably to overall impact uncertainty. It is interesting that, with respect to the ensemble mean, Jacob et al. (2007) pointed out that "*when many RCMs are used in a coordinated way, ... the ensemble mean nearly always is in better agreement with observed climatology than any individual model*." Similar findings were reported e.g. by Ines and Hansen (2006), Gleckler et al. (2008), Dosio and Paruolo (2011) or Nikulin et al. (2012). It should be kept in mind, however, that, just as with the application of BC methods, averaging across an ensemble invariably compromises physical consistency among fields.

In short, nested approaches can help to reduce the bias; multi-model ensembles can help to quantify the uncertainty associated with CCIS results. Implementing any of these approaches requires considerable expertise across a range of models as well as extensive data handling and computing power. Establishing full multi-model ensembles as a standard will therefore be more likely to happen in the mid- rather than the short term.

6.3 Proposals for the long term

The most challenging, time-consuming but ultimately most promising and satisfying approach to reduce the bias in GCM/RCM/HM model chains is to improve the models themselves. Present-day GCMs, RCMs and HMs are far from being perfect, and issues such as truncation of scales, violation of scaling laws, collapsing physical processes to their mean, lack of feedback from regional to global scales, etc. still compromise the physical foundation of the models. However, they are the only basis on which we can, by and by, add new insights in the functioning of the coupled atmosphere-land-ocean system.

This can be achieved in several ways:

- *Improved process descriptions*: besides improvements as a result of deeper insight into meteorological processes based on novel experiments and observations, especially the explicit representation of convection in RCMs, also GCMs have the potential to substantially enhance model accuracy (Maraun et al., 2010). Explicit incorporation of convection adds process knowledge to the model and allows for small-scale land-atmosphere feedback processes. Convection-permitting approaches partially alleviate the wet-day bias and underestimation of precipitation extremes present in most GCMs/RCMs (see Sect. 2.2) (Stephens et al., 2010; Maraun et al., 2010; Warrach-Sagi et al., 2012). Recent results from campaigns and modeling activities within projects of the World Weather Research Program (WWRP) demonstrate advanced model performance if the models are operated on the convection-permitting scale, i.e. grid resolutions of about 4 km (Rotach et al., 2009; Bauer et al., 2011; Wulfmeyer et al., 2011).

- An indispensable prerequisite for the move from parameterized to *explicit representation of deep convection* is *increased spatiotemporal resolution*. This is computationally expensive and currently restricts convection-permitting approaches mainly to RCMs. However, first tests with the global Nonhydrostatic ICosahedral Atmospheric Model (NICAM) (Satoh et al., 2008) at convection-permitting resolution (e.g. Fudeyasu et al., 2008) show encouraging results.

- *Improved ensemble prediction systems (EPS) by suitable perturbations*: extensive research is required on the development of multi-model or multi-physics EPS. It is not clear yet what is the most promising approach. In any case, it is also necessary to perturb the land-surface model.

- *Integration of state-of-the-art hydrological models in GCMs/RCMs*: as described in Sect. 5.1, terrestrial hydrological processes in GCMs and RCMs are usually represented in a way which precludes their direct use for hydrological problems. Instead, HMs are successively applied at the expense of losing the possibility for direct land-atmosphere feedback. The way forward is then to integrate state-of-the-art hydrological models, capable of closing the energy, mass and momentum balance of the atmospheric model components while at the same time operating at acceptable computation times (e.g. Van den Hurk et al., 2005). Given the importance of land-atmosphere interaction (especially related to water availability on the ground and its partitioning into evapotranspiration and runoff) for local heat fluxes and convection initiation (Betts, 2009; Van den Hurk et al., 2005), this has the potential to substantially improve the reliability of climate simulations and predictions.

Have the research activities conducted to develop and test BC methods then, after all, been a waste of time? Surely not. Despite our opinion that BC should not be applied in the way it is currently often done, analysing the nature and quantifying the magnitude of model biases associated with research on BC or post-processing in general has greatly improved the identification of model deficiencies (e.g. Vannitsem and Nicolis, 2008; Vannitsem, 2008; Eden et al. 2012). In that sense, the methods of BC can be seen as model diagnostic tools, for instance for problems associated with model resolution (e.g. Giorgi and Marinucci, 1996) or coupling of climate system components (e.g. Gupta et al., 2012).

Knowledge of the spatio-temporal patterns of bias thus helps to identify specific model deficits and offers the possibility of targeted improvement of GCM/RCM/HM process formulations, resolution and parameterization.

7 Summary and conclusions

In this article, we have argued that bias correction as currently used to correct the output of global or regional circulation models (GCM/RCM) in climate change impact studies (CCIS) is often not a valid procedure. To motivate this, we started with a definition of bias and presented an overview of its causes. We have demonstrated that biases of present-day circulation models are substantial and that, as a consequence, removing them through bias correction (BC) influences the results of CCIS in a non-negligible way. We have presented approaches to deal with biased model output with a focus on BC. We argue that the range of existing BC methods reflects the range of circulation model deficiencies from the user perspective and that they have been developed more from the perspective of necessity rather than validity. Based on a brief overview of state-of-the-art BC methods, we discussed the related assumptions and implications and concluded that BC is currently often used in an invalid way: it is added to the GCM/RCM model chain without sufficient proof that the consistency of the latter, i.e. the agreement between model dynamics/output, and our judgement and the generality of its applicability increases. BC methods often impair the advantages of circulation models by altering spatiotemporal field consistency, relations among variables and by violating conservation principles. BC largely neglects feedback mechanisms, and it is unclear whether BC methods are time-invariant under climate change conditions. Applying BC increases agreement of GCM/RCM output with observations and hence narrows the uncertainty range of simulations and predictions, often without providing a satisfactory physical justification. This is in most cases not transparent to the end user. We argued that this hides rather than reduces uncertainty, which may lead to avoidable forejudging by end users and decision makers. Finally, we proposed ways to cope with biased output of circulation models in the short term and how to reduce the bias in the long term.

The most promising strategy for improved future GCM and RCM simulations is the increase in model resolution at the convection-permitting scale in combination with ensemble predictions based on sophisticated approaches for ensemble perturbation.

With this article, we advocate openly communicating the entire uncertainty range associated with climate change predictions and hope to stimulate a lively discussion on BC among the atmospheric and hydrological community and end users of CCIS.

Acknowledgements. Uwe Ehret would like to thank HESS Editor Stan Schymanski and HESS Executive Editors Hubert Savenije and Murugesu Sivapalan for inviting him to write this commentary.

Uwe Ehret and Erwin Zehe thank Hoshin V. Gupta for a stimulating discussion on the topic.

The service charges for this open access publication
have been covered by a Research Centre of the
Helmholtz Association.

Edited by: M. Sivapalan

References

Bauer, H.-S., Weusthoff, T., Dorninger, M., Wulfmeyer, V., Schwitalla, T., Gorgas, T., Arpagaus, M., and Warrach-Sagi, K.: Predictive skill of a subset of models participating in D-PHASE in the COPS region, Q. J. Roy. Meteorol. Soc., 137, 287–305, doi:10.1002/qj.715, 2011.

Berg, P., Haerter, J. O., Thejll, P., Piani, C., Hagemann, S., and Christensen, J. H.: Seasonal characteristics of the relationship between daily precipitation intensity and surface temperature, J. Geophys. Res.-Atmos., 114, D18102, doi:10.1029/2009jd012008, 2009.

Betts, A. K.: Land-Surface-Atmosphere Coupling in Observations and Models, J. Adv. Model. Earth Syst., 1, 18 pp., doi:10.3894/james.2009.1.4, 2009.

Blöschl, G. and Sivapalan, M.: Scale issues in hydrological modeling – a review, Hydrol. Process., 9, 251–290, doi:10.1002/hyp.3360090305, 1995.

Burger, G.: Expanded downscaling for generating local weather scenarios, Clim. Res., 7, 111–128, doi:10.3354/cr007111, 1996.

Buser, C. M., Kunsch, H. R., Luthi, D., Wild, M., and Schar, C.: Bayesian multi-model projection of climate: bias assumptions and interannual variability, Clim. Dynam., 33, 849–868, doi:10.1007/s00382-009-0588-6, 2009.

Chen, C., Haerter, J. O., Hagemann, S., and Piani, C.: On the contribution of statistical bias correction to the uncertainty in the projected hydrological cycle, Geophys. Res. Lett., 38, L20403, doi:10.1029/2011gl049318, 2011.

Christensen, J. H. and Christensen, O. B.: A summary of the PRUDENCE model projections of changes in European climate by the end of this century, Climatic Change, 81, 7–30, doi:10.1007/s10584-006-9210-7, 2007.

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.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteorol. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Doherty, S., Bojinski, S., Goodrich, D., Henderson-Sellers, A., Noone, K., Bindoff, N., Church, J., Hibbard, K., Karl, T., Kajfez-Bogataj, L., Lynch, A., Parker, D., Thorne, P., Prentice, I., Ramaswamy, V., Saunders, R., Smith, M., Steffen, K., Stocker, T., Trenberth, K., Verstraete, M., and Zwiers, F.: Lessons learned from IPCC AR4: Scientific developments needed to understand, predict, and respond to climate change, B. Am. Meteorol. Soc., 90, 497–513, doi:10.1175/2008bams2643.1, 2009.

Dosio, A. and Paruolo, P.: Bias correction of the ENSEMBLES high-resolution climate change projections for use by impact models: Evaluation on the present climate, J. Geophys. Res.-Atmos., 116, D16106, doi:10.1029/2011jd015934, 2011.

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.

Fowler, H. and Kilsby, C.: Using regional climate model data to simulate historical and future river flows in northwest England, Climatic Change, 80, 337–367, doi:10.1007/s10584-006-9117-3, 2007.

Fudeyasu, H., Wang, Y. Q., Satoh, M., Nasuno, T., Miura, H., and Yanase, W.: Global cloud-system-resolving model NICAM successfully simulated the lifecycles of two real tropical cyclones, Geophys. Res. Lett., 35, L22808, doi:10.1029/2008gl036003, 2008.

Giorgi, F. and Marinucci, M. R.: An investigation of the sensitivity of simulated precipitation to model resolution and its implications for climate studies, Mon. Weather Rev., 124, 148–166, doi:10.1175/1520-0493(1996)124<0148:aiotso>2.0.co;2, 1996.

Giorgi, F., Jones, C., and Asrar, G.: Addressing climate information needs at the regional level: The cordex framework, WMO Bull., 58, 175–183, 2009.

Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, J. Geophys. Res.-Atmos., 113, D06104, doi:10.1029/2007jd008972, 2008.

Gupta, A. S., Muir, L. C., Brown, J. N., Phipps, S. J., Durack, P. J., Monselesan, D., and Wijffels, S. E.: Climate Drift in the CMIP3 Models, J. Climate, 25, 4621–4640, doi:10.1175/jcli-d-11-00312.1, 2012.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., and Yeh, P.: Multimodel Estimate of the Global Terrestrial Water Balance: Setup and First Results,

J. Hydrometeorol., 12, 869–884, doi:10.1175/2011jhm1324.1, 2011.

Haddeland, I., Heinke, J., Voß, F., Eisner, S., Chen, C., Hagemann, S., and Ludwig, F.: Effects of climate model radiation, humidity and wind estimates on hydrological simulations, Hydrol. Earth Syst. Sci., 16, 305–318, doi:10.5194/hess-16-305-2012, 2012.

Haerter, J. O., Hagemann, S., Moseley, C., and Piani, C.: Climate model bias correction and the role of timescales, Hydrol. Earth Syst. Sci., 15, 1065–1079, doi:10.5194/hess-15-1065-2011, 2011.

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.

Hay, L. E. and Clark, M. P.: Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western United States, J. Hydrol., 282, 56–75, doi:10.1016/s0022-1694(03)00252-x, 2003.

Hay, L. E., Wilby, R. J. L., and Leavesley, G. H.: A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States, J. Am. Water Resour. Assoc., 36, 387–397, doi:10.1111/j.1752-1688.2000.tb04276.x, 2000.

Hay, L. E., Clark, M. P., Wilby, R. L., Gutowski, W. J., Leavesley, G. H., Pan, Z., Arritt, R. W., and Takle, E. S.: Use of regional climate model output for hydrologic simulations, J. Hydrometeorol., 3, 571–590, doi:10.1175/1525-7541(2002)003<0571:uorcmo>2.0.co;2, 2002.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006, J. Geophys. Res., 113, D20119, doi:10.1029/2008jd010201, 2008.

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.

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, Climatic Change, 81, 31–52, doi:10.1007/s10584-006-9213-4, 2007.

Johnson, F. and Sharma, A.: A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations, Water Resour. Res., 48, W01504, doi:10.1029/2011wr010464, 2012.

Knutti, R.: Should we believe model predictions of future climate change?, Philos. T. Roy. Soc. A, 366, 4647–4664, doi:10.1098/rsta.2008.0169, 2008.

Kundzewicz, Z. W., Mata, L. J., Arnell, N. W., Döll, P., Kabat, P., Jiménez, B., Miller, K. A., Oki, T., Sen, Z., and Shiklomanov, I. A.: Freshwater resources and their management, in: Climate Change 2007: Impacts, Adaptation and Vulnerability – Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK and New York, NY, USA, 2007.

Li, H. B., Sheffield, J., and Wood, E. F.: Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching, J. Geophys. Res.-Atmos., 115, D10101, doi:10.1029/2009jd012882, 2010.

Liang, X. Z., Kunkel, K. E., Meehl, G. A., Jones, R. G., and Wang, J. X. L.: Regional climate models downscaling analysis of general circulation models present climate biases propagation into future change projections, Geophys. Res. Lett., 35, L08709, doi:10.1029/2007gl032849, 2008.

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, Reviews of Geophysics, 48, Rg3003, doi:10.1029/2009rg000314, 2010.

Moron, V., Robertson, A. W., Ward, M. N., and Ndiaye, O.: Weather types and rainfall over Senegal, Part II: Downscaling of GCM simulations, J. Climate, 21, 288–307, doi:10.1175/2007jcli1624.1, 2008.

Murphy, A. H.: What is a good forecast – An essay on the nature of goodness in weather forecasting, Weather Forecast., 8, 281–293, doi:10.1175/1520-0434(1993)008<0281:wiagfa>2.0.co;2, 1993.

Nikulin, G., Jones, C., Samuelsson, P., Giorgi, F., Sylla, M. B., Asrar, G., Büchner, M., Cerezo-Mota, R., Christensen, O. B., Dequè, M., Fernández, J., Hänsler, A., van Meijgaard, E., and Sushama, L.: Precipitation Climatology in an Ensemble of CORDEX-Africa Regional Climate Simulations, J. Climate, in press, 2012.

Ott, I., Düthmann, D., Liebert, J., Berg, P., Feldmann, H., Ihringer, J., Kunstmann, H., Merz, B., Schädler, G., and Wagner, S.: Climate change impact on medium and small sized river catchments in Germany: An ensemble assessment, J. Hydrol., submitted, 2012.

Piani, C., Weedon, G. P., Best, M., Gomes, S. M., Viterbo, P., Hagemann, S., and Haerter, J. O.: Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models, J. Hydrol., 395, 199–215, doi:10.1016/j.jhydrol.2010.10.024, 2010.

Randall, D. A., Wood, R. A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R. J., Sumi, A., and Taylor, K. E.: Climate Models and Their Evaluation, in: Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change Cambridge University Press, Cambridge, UK and New York, NY, USA, 2007.

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.

Rotach, M. W., Ambrosetti, P., Ament, F., Appenzeller, C., Arpagaus, M., Bauer, H. S., Behrendt, A., Bouttier, F., Buzzi, A., Corazza, M., Davolio, S., Denhard, M., Dorninger, M., Fontannaz, L., Frick, J., Fundel, F., Germann, U., Gorgas, T., Hegg, C., Hering, A., Keil, C., Liniger, M. A., Marsigli, C., McTaggart-Cowan, R., Montaini, A., Mylne, K., Ranzi, R., Richard, E., Rossa, A., Santos-Munoz, D., Schar, C., Seity, Y., Staudinger, M., Stoll, M., Volkert, H., Walser, A., Wang, Y., Werhahn, J., Wulfmeyer, V., and Zappa, M.: MAP D-PHASE Real-Time Demonstration of Weather Forecast Quality in the Alpine Region, B. Am. Meteorol. Soc., 90, 1321–1336, doi:10.1175/2009bams2776.1, 2009.

Satoh, M., Matsuno, T., Tomita, H., Miura, H., Nasuno, T., and Iga, S.: Nonhydrostatic icosahedral atmospheric model (NICAM) for global cloud resolving simulations, J. Comput. Phys., 227, 3486–3514, doi:10.1016/j.jcp.2007.02.006, 2008.

Schädler, G., Berg, P., Düthmann, D., Feldmann, H., Ihringer, J., Kunstmann, H., Liebert, J., Merz, B., Ott, I., and Wagner, S.: Flood hazards in a changing climate, Project Report, Center for Disaster Management and Risk Reduction Technology (CEDIM), http://www.cedim.de/download/Flood_Hazards_in_a_Changing_Climate.pdf (last access: 20 September 2012), 2012.

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

Seneviratne, S. I., Luthi, D., Litschi, M., and Schar, C.: Land-atmosphere coupling and climate change in Europe, Nature, 443, 205–209, doi:10.1038/nature05095, 2006.

Sharma, D., Das Gupta, A., and Babel, M. S.: Spatial disaggregation of bias-corrected GCM precipitation for improved hydrologic simulation: Ping River Basin, Thailand, Hydrol. Earth Syst. Sci., 11, 1373–1390, doi:10.5194/hess-11-1373-2007, 2007.

Stehlik, J. and Bardossy, A.: Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation, J. Hydrol., 256, 120–141, doi:10.1016/s0022-1694(01)00529-7, 2002.

Stephens, G. L., L'Ecuyer, T., Forbes, R., Gettlemen, A., Golaz, J. C., Bodas-Salcedo, A., Suzuki, K., Gabriel, P., and Haynes, J.: Dreary state of precipitation in global models, J. Geophys. Res.-Atmos., 115, D24211, doi:10.1029/2010jd014532, 2010.

Sun, F. B., 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.

Sun, Y., Solomon, S., Dai, A., and Portmann, R. W.: How often does it rain?, J. Climate, 19, 916–934, doi:10.1175/jcli3672.1, 2006.

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. Compass, 4, 834–860, doi:10.1111/j.1749-8198.2010.00357.x, 2010.

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.

Van den Hurk, B., Hirschi, M., Schar, C., Lenderink, G., Van Meijgaard, E., Van Ulden, A., Rockel, B., Hagemann, S., Graham, P., Kjellstrom, E., and Jones, R.: Soil control on runoff response to climate change in regional climate model simulations, J. Climate, 18, 3536–3551, doi:10.1175/jcli3471.1, 2005.

Vannitsem, S.: Dynamical properties of MOS forecasts: Analysis of the ECMWF operational forecasting system, Weather Forecast., 23, 1032–1043, 2008.

Vannitsem, S.: Bias correction and post-processing under climate change, Nonlin. Processes Geophys., 18, 911–924, doi:10.5194/npg-18-911-2011, 2011.

Vannitsem, S. and Nicolis, C.: Dynamical properties of Model Output Statistics forecasts, Mon. Weather Rev., 136, 405–419, 2008.

Warrach-Sagi, K., Schwitalla, T., Wulfmeyer, V., and Bauer, H.-S.: Evaluation of a CORDEX-Europe simulation with WRF: precipitation in Germany, Clim. Dynam., submitted, 2012.

Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Osterle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: Creation of the WATCH Forcing Data and Its Use to Assess Global and Regional Reference Crop Evaporation over Land during the Twentieth Century, J. Hydrometeorol., 12, 823–848, doi:10.1175/2011jhm1369.1, 2011.

Wilby, R. L.: Evaluating climate model outputs for hydrological applications, Hydrol. Sci. J.-J. Sci. Hydrol., 55, 1090–1093, doi:10.1080/02626667.2010.513212, 2010.

Wilby, R. L. and Wigley, T. M. L.: Downscaling general circulation model output: a review of methods and limitations, Prog. Phys. Geogr., 21, 530–548, doi:10.1177/030913339702100403, 1997.

Wilby, R. L., Hay, L. E., Gutowski, W. J., Arritt, R. W., Takle, E. S., Pan, Z. T., Leavesley, G. H., and Clark, M. P.: Hydrological responses to dynamically and statistically downscaled climate model output, Geophys. Res. Lett., 27, 1199–1202, doi:10.1029/1999gl006078, 2000.

Wojcik, R. and Buishand, T. A.: Simulation of 6-hourly rainfall and temperature by two resampling schemes, J. Hydrol., 273, 69–80, doi:10.1016/s0022-1694(02)00355-4, 2003.

Wood, A. W., Leung, L. R., Sridhar, V., and Lettenmaier, D. P.: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, Climatic Change, 62, 189–216, doi:10.1023/B:CLIM.0000013685.99609.9e, 2004.

WCRP – World Climate Research, P., World Meteorological, O., Intergovernmental Oceanographic, C., and International Council of Scientific, U.: WCRP implementation plan 2010–2015, World Meteorological Organization, Geneva, 2009.

Wulfmeyer, V., Behrendt, A., Kottmeier, C., Corsmeier, U., Barthlott, C., Craig, G. C., Hagen, M., Althausen, D., Aoshima, F., Arpagaus, M., Bauer, H. S., Bennett, L., Blyth, A., Brandau, C., Champollion, C., Crewell, S., Dick, G., Di Girolamo, P., Dorninger, M., Dufournet, Y., Eigenmann, R., Engelmann, R., Flamant, C., Foken, T., Gorgas, T., Grzeschik, M., Handwerker, J., Hauck, C., Holler, H., Junkermann, W., Kalthoff, N., Kiemle, C., Klink, S., Konig, M., Krauss, L., Long, C. N., Madonna, F., Mobbs, S., Neininger, B., Pal, S., Peters, G., Pigeon, G., Richard, E., Rotach, M. W., Russchenberg, H., Schwitalla, T., Smith, V., Steinacker, R., Trentmann, J., Turner, D. D., van Baelen, J., Vogt, S., Volkert, H., Weckwerth, T., Wernli, H., Wieser, A., and Wirth, M.: The Convective and Orographically-induced Precipitation Study (COPS): the scientific strategy, the field phase, and research highlights, Q. J. Roy. Meteorol. Soc., 137, 3–30, doi:10.1002/qj.752, 2011.

WWRP 2009-1: Recommendations for the Verification and Intercomparison of QPFs and PQPFs from Operational NWP Models, WMO/TD – No. 1485, World Meteorological Organization, available at: www.wmo.int (last access: 20 September 2012), 2009.