CrossMark

Bias Correcting Climate Change Simulations - a Critical Review

Douglas Maraun[1] ⬤

Abstract Climate models are our major source of knowledge about climate change. The impacts of climate change are often quantified by impact models. Whereas impact models typically require high resolution unbiased input data, global and regional climate models are in general biased, their resolution is often lower than desired. Thus, many users of climate model data apply some form of bias correction and downscaling. A fundamental assumption of bias correction is that the considered climate model produces skillful input for a bias correction, including a plausible representation of climate change. Current bias correction methods cannot plausibly correct climate change trends, and have limited ability to downscale. Cross validation of marginal aspects is not sufficient to evaluate bias correction and needs to be complemented by further analyses. Future research should address the development of stochastic models for downscaling and approaches to explicitly incorporate process understanding.

Keywords Regional climate modelling · Bias correction · Downscaling · Statistical post-processing · Model output statistics

✉ Douglas Maraun
  douglas.maraun@uni-graz.at

1 Wegener Center for Climate and Global Change, University of Graz, Brandhofgasse 5, 8010 Graz, Austria

Introduction

Many possible impacts of future climate change will be experienced at the regional scale [1]. These impacts may be quantified by impact models, which often require high resolution meteorological input data that are—for present conditions—unbiased compared to observations [2]. Coupled atmosphere ocean general circulation models (GCMs) are our major source of knowledge about future climate change, but they currently neither provide regional-scale nor unbiased information [3]. In particular processes, governing regional- to local-scale extreme events are, if at all, not well represented in GCMs [4]. Regional climate models (RCMs) are becoming popular to bridge this scale-gap at least partly—the typical horizontal resolution at present is about 10–15 km—but also these still have substantial errors, partly inherited from the driving GCMs [5, 6].

Thus, in any instance, many users of climate model data demand some form of bias correction, sometimes called bias adjustment. The origins of bias correction go back to so-called model output statistics (MOS) [7] in numerical weather prediction (NWP), which complements the widely used perfect prognosis statistical downscaling approach [8]. Due to its relative simplicity and low computational demand along with growing data bases of global and regional climate model simulations, bias correction has become very popular in climate impact research. Over recent years, many different methods have been developed [9, 10] and widely applied to post-process climate projections [11–15].

In parallel, a critical debate about downscaling in general [16–18] and bias correction in particular [19–22] has flared up. Bias-corrected climate model data may serve as the basis for real-world adaptation decisions and should thus be plausible, defensible and actionable [18]. The use of bias correction therefore has a distinct ethical dimension.

The aim of this paper is to provide a concise introduction to bias correction. I will not give a comprehensive overview of the technical details of different methods, but rather review the background, conceptual aspects and—also in the light of the ethical dimension—ongoing discussions about the applicability and limitations of bias correction. A detailed presentation of specific bias correction methods can be found in [10], a comparison with classical perfect prognosis statistical downscaling in [9]. For a detailed and comprehensive evaluation of bias correction methods, I refer the reader to the framework developed by the EU COST Action VALUE [23] and the related upcoming special issue.

In the next section, I present some relevant definitions, review the origins of bias correction in weather forecasting and discuss the rationale and associated assumptions of bias correction. In "Methods for Bias Correction", I will give a brief overview of the most commonly used bias correction methods. The ongoing discussion about open questions and limitations of bias correction will be presented in "Recent Discussions and Developments", followed by a discussion of the evaluation and performance of bias correction in "Evaluation and Performance of Bias Correction". I conclude with an outlook of future research.

Conceptual Issues

Definitions

Observed and simulated climate can be thought of as a sample of a time-dependent multivariate probability distribution—multivariate in space, in time and between different climatic variables (see Fig. 1). The unconditional distribution of one variable at one location (and precisely at one time) is called marginal distribution. It can be thought of as the distribution of one variable at one location, ignoring the influence of all other variables (or locations). Marginal aspects of the multivariate climate distribution can be expressed by summary statistics such as mean, variance, or wet day probability. Temporal aspects by, e.g. the lag-one autocorrelation or the mean dry spell length; spatial aspects by, e.g. the correlation between different locations; and multivariable aspects, e.g. by the correlation between different climatic variables.

From a given finite observational time series (or field of time series), one can only derive estimates of the climate distribution and its statistics. Even the estimation of climatic mean values is considerably affected by long term internal variability. The same holds for climate model output, although long simulations or initial condition ensembles help reducing sampling uncertainty.
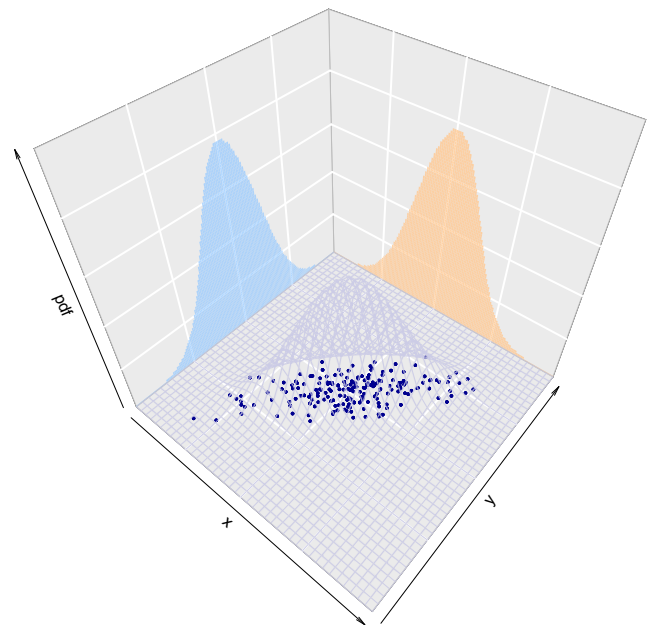

Fig. 1 Two-dimensional distribution (e.g. for two different variables or one variable at two locations or two times). *Surface plot* indicates multivariate probability density function. *Dots* indicate sample from the distribution. *Blue and orange* indicate marginal probability density functions

Transferring the bias concepts from statistics and forecast verification [24, 25] to a climate modelling context, a climate model bias can be defined as the systematic difference between a simulated climate statistic and the corresponding real-world climate statistic. I will follow this definition throughout the manuscript. A model bias derived from model and observational data is—as the statistics it is calculated from—only an estimate of the true model bias and therefore also affected by internal climate variability [9, 26, 27].

As the climate system and, hence, the climate distribution change with time, also a climate model bias is in general time dependent. In other words, the simulated climate change is in general not correct. Some authors, however, define a bias as the time-invariant component of a model error [28]. But a bias defined this way is not uniquely defined: changing the reference period would in general change the bias. The definition of a time-independent error component is therefore itself time-dependent and arbitrary.

Origins in Weather Forecasting

In NWP, the first statistical post-processing methods have been invented as early as in the late 1950s. The first operational models were too coarse to predict local weather and simulated only few prognostic variables such as pressure

and temperature. To overcome this gap, statistical regression models have been calibrated between the observed large-scale circulation—for those variables that were simulated by the models—and the local scale weather observations of interest, and applied to downscale the actual numerical forecast [8]. For this approach to yield reasonable local forecasts, the large-scale predictor has to be perfectly forecast by the numerical model, hence, the term perfect prognosis (PP) had been introduced. In practice, this assumption is often not met. Therefore, a new approach—so-called Model Output Statistics (MOS)—had been developed [7]: about a decade later, a considerable database of past forecasts had been archived. Instead of calibrating a statistical downscaling model in the real world, it was now possible to infer a relationship between past numerical weather forecasts and corresponding past observations. This approach automatically corrects systematic model biases.

With the growing interest in climate change impacts, MOS-approaches had been adapted for climate modelling. But crucially, and different to numerical weather forecasts, transient climate simulations are not in synchrony with observations. As a result, regression models cannot easily be calibrated. Therefore, researchers explored possibilities to bias correct on the basis of long-term distributions instead of day-to-day relationships [29, 30] (although these studies where still not based on transient climate simulations but reanalysis driven RCMs). To have any credibility, these methods must be homogeneous, i.e. map identical variables onto each other. In numerical weather prediction, MOS systems could employ a range of different predictors; in climate modelling, post-processed temperature is predicted by temperature, precipitation by precipitation. In other words, MOS in climate studies is almost exclusively bias correction. It was indeed shown that reanalysis-simulated precipitation—as a proxy for GCM output—is a skillful predictor for regional-scale observed precipitation [31]. Subsequently, the approach was applied to transient climate change simulations with GCMs [32]. Since then, bias correction has become an essential tool in climate impact studies, in particular after large GCM [33, 34] and RCM [35–37] datasets have become publicly available.

As discussed above, in numerical weather prediction observed and modelled weather sequences are in close synchrony; in climate modelling they are essentially uncorrelated. In addition to the problem of not being able to calibrate regression models, this difference has further important consequences [22]: (1) whereas skill of MOS in numerical weather prediction can be quantified by forecast verification scores, this assessment is essentially impossible for climate change studies. (2) In numerical weather forecasting, prediction lead times are typically too short for the

numerical model to drift into its own biased attractor. Free running climate models, however, are biased on all spatial and temporal scales.

Rationale and Assumptions

Hydrological models—such as other impact models—are typically calibrated to optimally represent the statistics of some desired observed hydrological variable (such as runoff), given observed meteorological input. If the observed input is replaced by model data, the realism of the hydrological model simulation is in general reduced. Bias correcting the input data has been shown to increase the agreement between simulated and observed hydrological data [2]. The first obvious aim of a bias correction is therefore adjusting selected simulated statistics—such as means, variances or wet-day probabilities—to match observations during a present-day calibration period. But several decisions need to be drawn, and the different possibilities may imply different specific assumptions to be fulfilled:

– should a bias correction method be applied that preserves or alters the climate change signal? A trend preserving bias correction is justified under the assumption that the model bias is time invariant; a non-trend preserving method may sensibly be used if it can be assumed that this method captures the time-invariance of the bias, i.e. that it corrects the simulated change.
– is downscaling to higher resolution or even point scales intended? In such a case, one has to assume that the downscaling captures the required local variations at the time scales of interest, as well as the response to climate change.
– which aspects of the climate distribution should be corrected? Most bias correction methods adjust marginal aspects only. Should also spatial, temporal and multivariate aspects be explicitly adjusted? In all cases, the underlying assumption is that the climate change signal of the considered aspects, after bias correction, is plausibly represented.

To what extent these decisions might be sensible will be discussed in "Recent Discussions and Developments".

The most crucial assumption underlying bias correction, however, is that the driving dynamical model skillfully simulates the processes relevant for the output to be corrected [22, 28]. For climate change simulations, this implies that also the changes of these processes are plausibly simulated [22]. Bias correction is a mere statistical post-processing and cannot overcome fundamental mis-specifications of a climate model.

Methods for Bias Correction

In the following, a simulated present-day model (predictor) time series of length N will be denoted as x_i^p, the corresponding observed (predictand) time series as y_i^p. Following the discussion in "Definitions", both may follow marginal distributions $x_i^p \sim D_{\text{raw}}^p$ and $y_i^p \sim D_{\text{real}}^p$. The mean of the uncorrected model over a chosen present period μ_{raw}^p can be estimated as $\hat{\mu}_{\text{raw}}^p = \bar{x_i^p}$ (where the hat denotes the estimator, the bar averaging in time), the corresponding real mean μ_{real}^p as $\hat{\mu}_{\text{real}}^p = \bar{y_i^p}$. An estimator of the model bias for present conditions is then given as

$$\widehat{Bias}(\mu^p) = \bar{x_i^p} - \bar{y_i^p}. \tag{1}$$

Correspondingly, the relative bias might be estimated as

$$\widehat{Rel.Bias}(\mu^p) = \bar{x_i^p} / \bar{y_i^p}. \tag{2}$$

The quantile for a probability α of a distribution D will be denoted as $q_D(\alpha)$ and is defined as the value which is exceeded with a probability $1 - \alpha$ when sampling from the distribution. The corresponding empirical quantile $\hat{q}_D(\alpha)$ can be obtained by sorting the given time series, say, x_i, and then considering the value at position $\alpha \times N/100$ (also called the rank of the data). The probabilities corresponding to a given quantile $q_D(\alpha)$ (i.e. the cumulative distribution function) are written as $p_D(q) = \alpha$. Future simulations and derived measures will be denoted with a superscript f.

Delta Change vs. Direct Methods

The most simple approach used for bias correction is the so-called delta change approach. It has a long history in climate impact research [38–40]. In fact, this approach is not a bias correction of a climate model, but only employs the model's response to climate change to modify observations. As it is a useful benchmark for bias correction, I will nevertheless discuss this approach. In its most basic application, a time series of future climate is generated as

$$x_{i,\text{corr}}^f = y_i^p + (\bar{x_i^f} - \bar{x_i^p}). \tag{3}$$

That is, an observed time series is taken, and only a model derived climate change signal is added. For variables such as precipitation (which assume positive values only), one would typically consider relative changes, i.e.

$$x_{i,\text{corr}}^f = y_i^p \times \frac{\bar{x_i^f}}{\bar{x_i^p}}. \tag{4}$$

This approach—of course—conserves the observed weather sequence and with it the linear spatial-, temporal- and multi-variable dependence structure. The delta change approach is therefore sensible only when these aspects may be assumed unchanged for the considered future climate.

Mathematically similar, but conceptually different, is a simple mean bias correction. It generates a future time series by subtracting the present-day model bias from the simulated future time series:

$$x_{i,\text{corr}}^f = x_{i,\text{raw}}^f - \widehat{Bias}(\mu^p) = x_{i,\text{raw}}^f - (\bar{x}_{i,\text{raw}}^f - \bar{y_i^p}), \tag{5}$$

or equivalently for precipitation

$$x_{i,\text{corr}}^f = \frac{x_{i,\text{raw}}^f}{\widehat{Rel.Bias}(\mu^p)} = x_{i,\text{raw}}^f \times \frac{\bar{y_i^p}}{\bar{x_i^p}}. \tag{6}$$

The latter formulation is known as linear scaling [31] and adjusts both mean and variance (but keeps their ratio constant). A modified version also adjusts the number of wet days [41]. Such approaches making direct use of the simulated time series are called direct approaches [9, 42]. For a sensible application under future climate, they require time-invariant biases (or relative biases).

Quantile Mapping

More flexible bias correction methods also attempt to adjust the variance of the model distribution to better match the observed variance [10]. A generalisation of all these approaches is quantile mapping, which employs a quantile-based transformation of distributions [43]. In a widely used variant, a quantile of the present day simulated distribution is replaced by the same quantile of the present-day observed distribution (see Fig. 2):

$$x_{i,\text{corr}}^f = q_{D_y^p}\left(p_{D_x^p}(x_{i,\text{raw}}^f)\right). \tag{7}$$

Typically, climate models simulate too many wet days (the so-called drizzle effect [44]). In this situation, quantile mapping automatically adjusts the number of wet days (as the wet-day threshold is a quantile of the distribution)
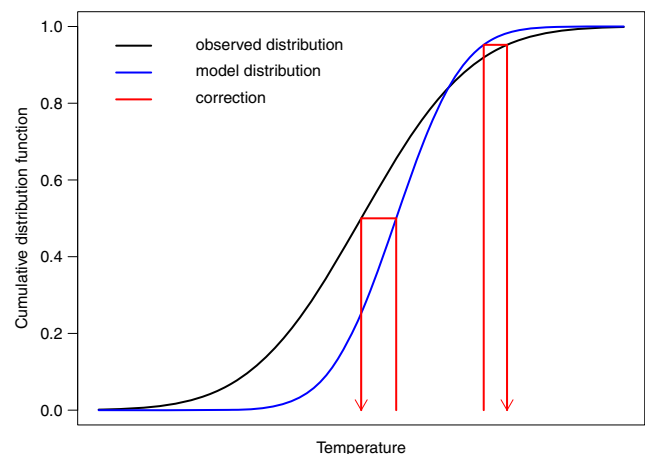
Fig. 2 Quantile mapping. A simulated value, a quantile of the simulated distribution, is replaced by the quantile of the observed distribution corresponding to the same probability

[45]. If a climate model simulates too few wet days, it has been suggested, e.g., to randomly generate low precipitation amounts [46]. The actual formulation of the transfer function depends on the implementation. Some authors consider empirical quantiles, linearly interpolated [47], others employ parametric models such as a normal distribution for temperature and a gamma distribution for precipitation [45, 48]. For extrapolating to unobserved quantiles, a constant transfer function beyond the highest observed quantile has been assumed [49], extrapolation based on the model used for the bulk of the distribution [45, 48], or specific extreme value models [50]. In general, the higher the flexibility of the mapping, the higher the danger of running into over-fitting and implausible applications [22]. In particular for high quantiles, where the sampling noise is very high, non-parametric quantile mapping produces very noisy results and essentially applies random corrections. Quantile mapping is mostly implemented as direct method, but can also be applied as a delta-change-like approach by modifying an observed series individually for different quantiles [51].

The implementation of quantile mapping according to Eq. 7 assumes value-dependent biases: a value $x_{i,raw}$, no matter whether it occurs in present-day or future simulation, will be transformed according to the transfer function $q_{D_y^p}\left(p_{D_x^p}(.)\right)$, but the transfer function might be different for different values. As such, this implementation is in general not trend preserving [11, 21]. Assume, e.g., a correction of temperature values as illustrated in Fig. 3. If the modelled present-day distribution has a negligible mean bias, but is too narrow compared to observations, quantiles close to the mean will only be weakly adjusted, whereas low and high values will be inflated to match the observations. Along with the inflation of the marginal distribution, variability on all time scales including the overall trend will be inflated.

Whether or not such trend modifications are sensible will be discussed in "State Dependent Biases and the Modification of Climate Change Trends". In any case, some authors developed trend preserving variants of quantile mapping.

The simplest version preserves the trend in the mean [13], a more sophisticated variant preserves the additive trend for each quantile [52] (Note that a method preserving trends in quantiles does not necessarily preserve the trend in the mean [53]). Other methods have been developed to preserve variability in the mean for a range of different time-scales [28]. A comparison of how different methods handle the representation of trends can be found in ref. [53].

Recent Discussions and Developments

State Dependent Biases and the Modification of Climate Change Trends

GCMs provide a plausible picture of global climate change, yet crucial phenomena are substantially mis-represented [54]. For instance, key processes governing El Niño/Southern Oscillation, Monsoonal systems or the mid-latitude storm track are biased along with high uncertainties in the representation of changes in these phenomena [22]. Such large-scale errors affect the representation of regional climate [55, 56] and are inherited by downscaling methods [6]. Also at regional scales, both global and regional climate models may mis-represent orography, feedbacks with the land surface [57–59], and sub-grid processes such that local surface climate is considerably biased [5, 60] and uncertainties in projections are high [61]. It is also not a priori clear whether sub-grid parameterisations, tuned to describe present climate, are valid under future climate conditions. For instance, there is evidence that the response of phenomena such as summer convective precipitation is not plausibly captured by operational RCMs at a resolution of 10 km and beyond [62, 63]. In other words, in several cases, simulated climate change trends might be implausible, and biases are expected to be time dependent [26, 64].

In case one has trust in the simulated change, one should employ a trend preserving bias correction. But in case one
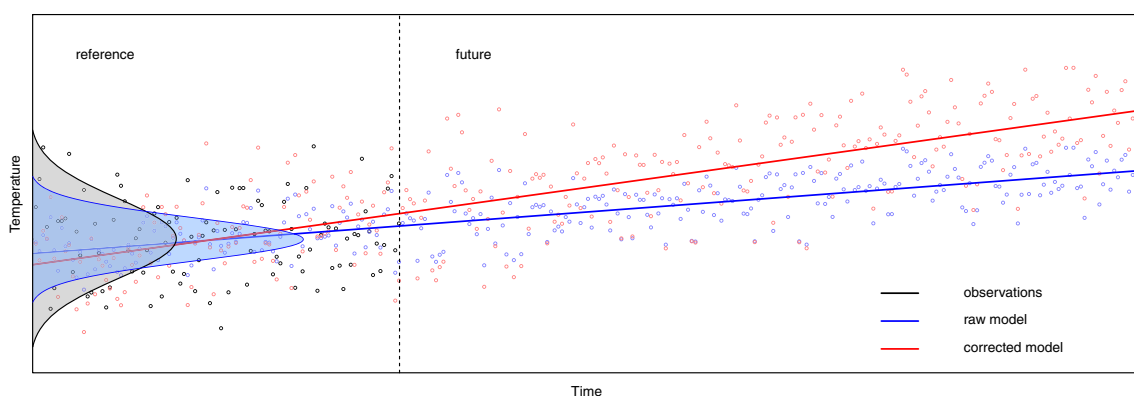

Fig. 3 Modification of trends by standard quantile mapping

suspects implausibly simulated trends, the way forward is more difficult. This is likely the case in the examples listed above, or in presence of circulation biases, even though the simulated changes might be plausible [22, 55, 65].

One solution might be to explicitly modify the simulated trend. Current implementations of quantile mapping do modify the climate change signal. But quantile mapping is calibrated on day-to-day variability, and in general, it is not clear whether the derived transfer function can be sensibly used to modify long-term variability and forced trends [22, 66]. If this transferability between time-scales cannot be established for the given application, one could either clearly communicate that no plausible climate change trends could be provided, or better try to derive an expert guess based on the model results and process understanding. In any case, one should not intransparently communicate the implausible trend as our best knowledge.

The aim of a local bias correction should, at any rate, not be modifying a trend to obtain a (hypothetical) true future value. Assume a GCM with a wrong climate sensitivity (in fact, we do not even know the true one), that simulates a plausible continental-scale response of regional climate in Europe to global warming. However, locally the GCM (or an RCM used to downscale the GCM) might mis-represent important feedbacks and processes, such that local changes are not consistent with global changes. A local bias correction could sensibly aim to correct the local errors, but should not be used to correct the wrong global climate sensitivity.

Bias Correction and Downscaling

As discussed in "Rationale and Assumptions", often a major aim of bias correction is to spatially downscale to point data. Downscaling itself can have several aims: (1) the provision of systematic spatial variations such as the variation of climatological temperature with elevation, or of climatological precipitation from the windward-side to the rain shadow of a mountain. (2) The provision of day-to-day variations in space such as the occurrence of localised rainfall events or temperature inversions between valleys and nearby mountains. For climate change applications, one of course also expects a sensible downscaling approach to provide a plausible future change of these variations. Almost trivially, bias correction can fulfill the first aim for present climate conditions—simply by calibration to high resolution observations.

Whether the second aim can be fulfilled depends strongly on the variable and region of interest. Consider two examples. Subgrid-variations of daily precipitation have a strong random component. As almost all state-of-the-art bias correction methods are deterministic, they cannot generate this random variability; variance correcting methods such as quantile mapping may inflate extreme events [21].

Similarly, temperature inversions in sub-grid scale valleys cannot be generated [22]. For some local phenomena, such as valley breezes, the grid-scale average might not be representative of local variations at all. Thus, bias correction should be used for downscaling only if sub-grid variations are very smooth in space.

Whether plausible sub-grid changes can be provided depends again on the variable and region of interest; the discussion is identical to the one in "State Dependent Biases and the Modification of Climate Change Trends". As an example, local effects such as the snow-albedo feedback might modulate the climate change signal at sub-grid scales, such that the large-scale simulated change is not a plausible representation of local changes. As a result, a trend preserving bias correction missed about one third of the expected springtime warming in the Sierra Nevada (USA) [22]. Also, local wind phenomena such as valley breezes might respond differently to climate change than the grid-scale average. Bias correction cannot plausibly improve these changes and should therefore not be used in such situations. These problems are not only relevant for downscaling to sub-grid scales, but also at the grid scale.

Correction of Multivariate Aspects

All bias correction methods discussed so far only modify marginal distributions and thus leave other aspects largely unchanged [67]. But as spatial, temporal and multi-variable aspects are often misrepresented by climate models, methods have been developed for multivariate bias correction, that also correct the dependence between variables [68, 69].

As discussed above, any bias correction introduces inconsistencies with the driving model. When adjusting joint (i.e. non-marginal) aspects, these inconsistencies might become rather complex and affect other aspects. Take for instance a correction of the dependence between precipitation at two locations. Assume that the probability of joint dry days is too high. Thus, a correction needs to replace some joint dry days at one location by a wet day. As a consequence, also the temporal dependence is altered. In the most extreme case, the simulated multivariate dependence is completely ignored and replaced by the observed [69]. As a consequence, also the sequence of events is taken from observations, the method is essentially a multivariate delta change approach. If one aims to apply such methods for climate change projections, one implicitly assumes that the resulting changes in the adjusted—and indirectly affected—aspects are plausible. The stronger the modification, the more this assumption is questionable. In particular the frequency and duration of long dry, hot and cold spells are still not well simulated by global climate models [54] (and these errors are inherited by regional climate models [6]). Moreover, these phenomena are strongly governed by

atmospheric dynamics and our confidence in projected changes is generally low [56]. In particular substantial changes which break consistency with major physical processes should be avoided, such as adjusting the diurnal cycle of precipitation or the onset of the Monsoon season. In any case, a decision thus has to be drawn about which aspects should be adjusted, and which inconsistencies might be tolerable.

Added Value

All these cases highlight the value of RCMs. Directly bias correcting a GCM is of course much cheaper than having an intermediate dynamical downscaling step. Moreover, it is difficult to demonstrate added value of an RCM, after both GCM and RCM have been bias corrected [70]. But the RCM resolution is typically five times higher than that of the GCM, i.e. much of the regional-scale variability resolved by the RCM is not represented by the GCM. Therefore, also the risk is high that crucial processes and as a result the climate change signal are represented much worse in the GCM than in an RCM—as illustrated in the case of snow albedo feedbacks [22, 71].

Evaluation and Performance of Bias Correction

The previous discussions have highlighted the need for a careful evaluation. As stated in "Origins in Weather Forecasting", the performance of MOS used in numerical weather forecasting can be evaluated by classical forecast verification, which are based on a pairwise comparison of predicted and actually observed values. To eliminate artificial skill from overly complex statistical post-processing, cross validation is applied: the statistical model is calibrated on a subset of data spanning the calibration period, and then used to predict the remaining data from the validation period [24, 25]. To reduce variability, a k-fold cross validation might be applied: the data is separated into k non-overlapping subsets, and the model is calibrated on all k permutations of $k - 1$ blocks, the withheld blocks are used as validation periods. The resulting k predictions are concatenated to one cross-validated time series which can be compared with the observed time series for model evaluation. Cross validation has widely been applied to evaluate bias correction methods [10, 48].

Because pairwise correspondence between predictors and predictands is generally missing in a climate modelling context, an evaluation can only compare modelled and observed long-term distributions. But these typically change slowly, such that a cross validation is not sufficient to identify artificial skill and unskillful bias correction [22]. The limitations of cross validation are particularly severe, because most evaluation studies of bias correction evaluate marginal aspects only—but these are typically calibrated to match observations. The significance of a positive result in such an evaluation is therefore limited. To minimise the dangers of not identifying artificial skill and unskillful bias correction, one should therefore evaluate aspects which have not been calibrated, in particular non-marginal aspects. Considering, e.g. diagnostics of inter-annual variability or spell length distributions helps to uncover bias correction problems [22].

For the design of sensible evaluation approaches, two issues have to be addressed: (1) does the bias correction method itself perform well under present and future conditions? and (2) does the climate model provide plausible input for a bias correction, both under present and future conditions? The EU COST Action VALUE developed the first comprehensive evaluation framework for downscaling methods including bias correction [23]. Different experiments need to be defined to fully address the two issues: (1) bias correction of reanalyis data or reanalysis driven RCMs to quantify the performance of a given bias correction method under present climate conditions. Here, modelled and simulated weather sequence is weakly synchronised, and an assessment of classical forecast skill is possible at least with seasonally aggregated data [57]. (2) Bias correction of GCMs (or GCM-driven RCMs) under present conditions to evaluate the plausibility of the GCM simulation. And (3) pseudo reality experiments to assess the performance of a given bias correction method under climate change conditions [26]. The latter experiment can also be carried out informally, e.g. by comparing the change given by a bias corrected model with changes simulated by a higher resolution model simulation, which is considered to simulate plausible changes [22, 72]. The assessment whether the driving model simulates plausible changes inherently relies on expert knowledge. One should, at least based on a literature review, assess whether the model plausibly simulates the processes relevant for the variable and region of interest.

For a detailed evaluation of the performance of different bias correction methods, also in comparison to classical perfect prognosis statistical downscaling methods, please refer to the forthcoming special issue of VALUE in the International Journal of Climatology (currently in preparation/under review). Generally, bias correction methods used to post-process reanalysis data or reanalysis-driven RCMs improve the marginal aspects of the raw model [73], conserve or improve (indirectly by correcting the marginal aspects) temporal [74] and spatial aspects [75]. As spatial and temporal aspects are not explicitly post-processed, a bias corrected RCM driven by reanalysis data typically performs better for these aspects than the directly bias corrected reanalysis data. In general bias correction performs at least

as well as perfect prognosis. Thus, if a dynamical model is available that provides skillful input with a plausible climate change signal for a given user problem, bias correction is a defensible and potentially powerful approach.

Discussion and Future Research

Bias correction is widely used in climate impact modelling. It first of all aims to adjust selected statistics of a climate model simulation to better match observed statistics over a present-day reference period. Bias correction may or may not involve a downscaling step; it may or may not modify the simulated climate change; and it may adjust marginal aspects only, or also spatial, temporal and multi-variable aspects. A fundamental assumption of bias correction is that the chosen climate model produces skillful input for a bias correction, including a plausible representation of climate change. Bias correction cannot fix fundamental problems of a climate model.

Current approaches do not apply physical knowledge to modify the climate change signal. If one can trust the simulated change, a trend-preserving correction is the method of choice. Standard quantile mapping does in general not modify trends in a physically plausible way. Current bias correction methods have a limited ability to further downscale the model output. Sub-grid day-to-day variability cannot be generated, and feedbacks altering the sub-grid climate change signal cannot be represented. Any modifications of spatial, temporal or multi-variable aspects may strongly break the consistency with the driving model, and affect other aspects than the desired ones. This holds in particular for major modifications of the temporal structure. Cross validation of marginal aspects is not sufficient to identify problems of bias correction and needs to be complemented by an evaluation of multivariate aspects. The evaluation should be carried out in a perfect boundary setting as well as in a transient setting; ideally also the simulated climate change should be analysed, e.g. in a pseudo reality.

Two major issues should be addressed by bias correction research. First, the development of bias correction methods that are explicitly designed for downscaling. Stochastic approaches should be developed to downscale either based on regression models [76] or disaggregation approaches [9]. They might be used in conjunction with quantile mapping to first bias correct and then downscale climate model data [77]. If downscaling is not needed, but only station data are available for bias correction, one could upscale the station statistics using Taylor's hypothesis of frozen turbulence [78]. Second, the development of approaches that explicitly incorporate process knowledge, to generate a plausible local response to climate change. Such approaches might be statistical convection emulators or statistical models that

represent sub-grid feedbacks [79]. Also, the use of emergent constraints [80] should be considered to bias correct and constrain the climate change signal.

In any case it should be acknowledged that a successful bias correction relies on a sound understanding not only of the statistical model, but also the relevant climatic processes and their representation of the considered climate model.

Compliance with Ethical Standards

References

1. Barros V, Field C, Dokken D, Mastrandrea M, Mach K, Bilir T, Chatterjee M, Ebi K, Estrada Y, Genova R, Girma B, Kissel E, Levy A, MacCracken S, Mastrandrea P, White L, (eds). Climate change 2014: impacts, adaptation, and vulnerability. Part B: regional aspects. Contribution of working group II to the fifth assessment report of the intergovernmental panel on climate change. Cambridge: Cambridge University Press; 2014.
2. Wilby R, Hay L, Gutowski W, Arritt R, Takle E, Pan Z, Leavesley G, Clark M. Geophys Res Lett. 2000;27(8):1199.
3. Collins M, Knutti R, Arblaster J, Dufresne JL, Fichefet T, P-Friedlingstein X, Gutowski W, Johns T, Krinner G, Shongwe M, Tebaldi C, Weaver A, Wehner M. Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge: Cambridge University Press; 2013. Long-term Climate Change: Projections, Committments and Irreversibility.
4. Volosciuk C, Maraun D, Semenov V, Park W. J Climate. 2015;28(3):1184.
5. Rummukainen M. Wiley Int Rev Clim Change. 2010;1:82. doi:10.1002/wcc.8.
6. Hall A. Science. 2014;346:1461. This paper illustrates the role of large-scale circulation errors for downscaling and highlights the garbage-in garbage-out problem.
7. Glahn HR, Lowry DA. J Appl Meteorol. 1972;11:1203.

8. Klein WH, Lewis BM, Enger I. J Meteorol. 1959;16:672.
9. Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienen S, Rust HW, Sauter T, Themeßl M, Venema VKC, Chun KP, Goodess CM, Jones RG, Onof C, Vrac M, Thiele-Eich I. Rev Geophys. 2010;48:RG3003.
10. Teutschbein C, Seibert J. J Hydrol. 2012;456:12.
11. Hagemann S, Chen C, Haerter J, Heinke J, Gerten D, Piani C. J Hydrometeorol. 2011;12(4):556.
12. Dosio A, Paruolo P, Rojas R, Geophys J. Res Atmos. 2012;117(D17).
13. Hempel S, Frieler K, Warszawski L, Schewe J, Piontek F. Earth Syst Dynam. 2013;4:219.
14. Maurer E, Brekke L, Pruitt T, Thrasher B, Long J, Duffy P, Dettinger M, Cayan D, Arnold J. Bull. Amer Meteorol Soc. 2014;95(7):1011.
15. Harding R, Weedon G, van Lanen H, Clark D. J Hydol. 2014;518:186.
16. Pielke R, Wilby R. EOS. 2012;93(5):52.
17. Barsugli J, Guentchev G, Horton RM, Wood A, Mearns L, Liang XZ, Winkler J, Dixon K, Hayhoe K, Rood R, Goddard L, Ray A, Buja L, Ammann C. EOS. 2013;94(46):424. This paper nicely discusses the problems users face when choosing regional climate change projections.
18. Hewitson B, Daron J, Crane R, Zermoglio M, Jack C. Clim Change. 2014;122:539. Starting from the ethical responsibility of climate data providers, this paper discusses key requirements for statistical downscaling methods.
19. Vannitsem S. Nonlin Proc Geophys. 2011;18:911.
20. Ehret U, Zehe E, Wulfmeyer V, Warrach-Sagi K, Liebert J. Hydrol Earth Syst Sci. 2012:16.
21. Maraun D. J Climate. 2013;26:2137. This paper reveals that bias correction methods cannot add missing unresolved local variability. Any attempt to represent this will produce artefacts.
22. Maraun D, Shepherd T, Widmann M, Zappa G, Walton D, Hall A, Gutierrez JM, Hagemann S, Richter I, Soares P, Mearns L. submitted. 2016. This paper illustrates problems and artefacts that may occur when applying bias correction without a thorough understanding of the underlying processes.
23. Maraun D, Widmann M, Gutierrez J, Kotlarski S, Chandler R, Hertig E, Wibig J, Huth R, Wilcke RAI. Earth's Future. 2015;3:1. This paper presents a comprehensive framework for the evaluation of downscaling and bias correction approaches.
24. von Storch H, Zwiers FW. Statistical analysis in climate research. Cambridge: Cambridge University Press; 1999.
25. Wilks DS. Statistical Methods in the Atmospheric Sciences, 2nd ed.: Academic Press/Elsevier; 2006.
26. Maraun D. Geophys Res Lett. 2012;39:L06706.
27. Teutschbein C, Seibert J. Hydrol Earth Syst Sci. 2013;17(12):5061.
28. Haerter J, Hagemann S, Moseley C, Piani C. Hydrol Earth Syst Sci. 2011;15(3):1065.
29. Hay L, Clark M, Wilby R, Gutowski W, Leavesley G, Pan Z, Arritt R, Takle E. J Hydrometeorol. 2002;3(5):571.
30. Wood A, Maurer E, Kumar A, Lettenmaier D, Geophys J. Res Atmos. 2002;107(D20).
31. Widmann M, Bretherton CS, Salathe EP. J Climate. 2003;16(5):799.
32. Salathé EP. Int J Climatol. 2005;25(4):419.
33. Meehl G, Covey C, Delworth T, Latif M, McAvaney B, Mitchell J, Stouffer R, Taylor K. Bull. Amer Meteorol Soc. 2007;88:1383.
34. Taylor KE, Stouffer RJ, Meehl GA. A summary of the CMIP5 experiment design. 2009. http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor_CMIP5_design.pdf.
35. Hewitt CD. EGGS Newsletter. 2005;13:22.
36. van der Linden P, Mitchell JFB. ENSEMBLES: climate change and its impacts: summary of research and results from the ENSEMBLES project. Tech. rep., Met Office Hadley Centre. 2009.
37. Giorgi F, Jones C, Asrar G. WMO Bull. 2009;58(3):175.
38. Rosenzweig C. Clim Change. 1985;7:367.
39. Santer B. Clim Change. 1985;7:71.
40. Gleick P. J Hydrol. 1986;88(1):97.
41. Schmidli J, Frei C, Vidale PL. Int J Climatol. 2006;26:679.
42. Déqué M, Rowell DP, Luthi D, Giorgi F, Christensen JH, Rockel B, Jacob D, Kjellström E, de Castro M, van den Hurk B. Clim Change. 2007;81:53.
43. Panofsky HW, Brier GW. Some applications of statistics to meteorology: The Pennsylvania State University Press; 1968, p. 224.
44. Gutowski W, Decker S, Donavon R, Pan Z, Arritt R, Takle E. J Climate. 2003;16:3841.
45. Hay LE, Clark MP. J Hydrol. 2003;282:56.
46. Themeßl MJ, Gobiet A, Heinrich G. Clim Change. 2012;112:449.
47. Themeßl MJ, Gobiet A, Leuprecht A. Int J Climatol. 2011;31:1530.
48. Piani C, Haerter JO, Coppola E. Theor Appl Climatol. 2010;99(1-2):187.
49. Boé J, Terray L, Habets F, Martin E. Int J Climatol. 2007;27:1643.
50. Michelangeli PA, Vrac M, Loukos H. Geophys Res Lett. 2009;36(11).
51. Willems P, Vrac M. J Hydrol. 2011;402(3):193.
52. Li H, Sheffield J, Wood E. J Geophys Res. 2010;115:D10101.
53. Pierce D, Cayan D, Maurer E, Abatzoglou J, Hegewisch K. J Hydrometeorol. 2015;16(6):2421. This paper discusses, how different implementations of quantile-mapping or similar approaches alter the climate change signal.
54. Flato G, Marotzke J, Abiodun B, Braconnot P, Chou S, Collins W, Cox P, Driouech F, Emori S, Eyring V, Forest C, Gleckler P, Guilyardi E, Jakob C, Kattsov V, Reason C, Rummukainen M. Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, in press, 2013), chap. Evaluation of climate models.
55. Eden J, Widmann M, Grawe D, Rast S. J Climate. 2012;25:3970.
56. Shepherd T. Nat Geosci. 2014;7:703. This paper discusses that we have confidence mainly in thermodynamic changes of climate. Dynamical changes, which strongly control regional climate change, are much more uncertain.
57. Maraun D, Widmann M. Hydrol Earth Syst Sci. 2015;19:3449.
58. Christensen JH, Boberg F, Christensen OB, Lucas-Picher P. Geophys Res Lett. 2008;35:L20709.
59. Bellprat O, Kotlarski S, Lüthi D, Schär C. Geophys Res Lett. 2013;40:4042. This paper demonstrates the limitations of standard bias correction techniques in the presence of non-resolved feedback mechanisms.
60. Kotlarski S, Keuler K, Christensen O, Colette A, Déqué M, Gobiet A, Goergen K, Jacob D, Lüthi D, van Meijgaard E, Nikulin G, Schär C, Teichmann C, Vautard R, Warrach-Sagi K, Wulfmeyer V. Geosci Model Dev Discuss. 2014;7:217. This paper presents a comprehensive analysis of biases in the state-of-the-art European regional climate model ensemble.
61. Jacob D, Petersen J, Eggert B, Alias A, Christensen O, Bouwer L, Braun A, Colette A, Déqué M, Georgievski G, Georgopoulou E, Gobiet A, Nikulin G, Haensler A, Hempelmann N, Jones C, Keuler K, Kovats S, Kröner N, Kotlarski S, Kriegsmann A, Martin E, E EV, Moseley C, Pfeifer S, Preuschmann S, Radtke K, Rechid D, Rounsevell M, Samuelsson P, Somot S, Soussana JF, Teichmann C, Valentini R, Vautard R, Weber B. Reg Environ Change. 2014;14:563.
62. Kendon E, Roberts N, Fowler H, Roberts M, Chan S, Senior C. Nat Clim Change. 2014;4:570. This paper demonstrates

that very high resolution simulations might be necessary to represent changes in summer extreme rainfall - which illustrates the relevance of having a sufficient resolution of input for bias correction.

63. Meredith E, Maraun D, Semenov V, Park W, Geophys J. Res Atmos. 2015;120(12):500.
64. Buser C, Künsch H, Lüthi D, Wild M, Schär C. Clim Dyn. 2009;33:849.
65. Addor N, Rohrer M, Furrer R, Seibert J. J Geophys Res. 2016. This paper nicely demonstrates the consequences of circulation errors for bias correction, which should be considered when correcting biases originating in large-scale GCM fields.
66. Maurer E, Pierce D. Hydrol Earth Syst Sci. 2014;18(3):915.
67. Wilcke R, Mendlik T, Gobiet A. Clim Change. 2013; 120(4):871.
68. Piani C, Haerter J. Geophys Res Lett. 2012;39(20):L20401.
69. Vrac M, Friederichs P. J Climate. 2015;28(1):218.
70. Eden J, Widmann M, Maraun D, Vrac M. J Geophys Res. 2014;119(11):040.
71. Salathe E, Steed R, Mass C, Zahn P. J Climate. 2008; 21:5708.
72. Dixon KW, Lanzante J, Nath M, Hayhoe K, Stoner A, Radhakrishnan A, Balaji V, Gaitán C. Clim Change. 2016; 135(3):395.
73. Gutiérrez J et al. Int J Climatol. 2016. submitted.
74. Maraun D, Huth R, Gutierrez J, San Martin D, Dubrovsky M, Fischer A, Hertig E, Soares P, Bartholy J, Pongracz R, Widmann M, Casado M, Ramos P. Int J Climatol. 2016. subm.
75. Widmann M et al. Int J Climatol. 2016. submitted.
76. Wong G, Maraun D, Vrac M, Widmann M, Eden J, Kent T. J Climate. 2014;27:6940.
77. Volosciuk C, Maraun D, Vrac M, Widmann M. 2016. submitted.
78. Haerter J, Eggert B, Moseley C, Piani C, Berg P. Geophys Res Lett. 2015;42:1919. This paper provides a nice idea to compare data at different spatial and temporal scales. It might prove useful to avoid the scale gap between model and observational data for evaluation.
79. Walton D, Sun F, Hall A, Capps S. J Climate. 2015;28(12):4597. This paper presents one of the first statistical post-processing approaches that deliberately modify the climate change signal based on process understanding.
80. Collins M, Chandler R, Cox P, Huthnance J, Rougier J, Stephenson D. Nat Clim Change. 2012;2(6):403.