Bias adjustment and the question of usable climate information: methodological assumptions and value judgements

BAMS Article

Bias Adjustment and the Question of Usable Climate Information: Methodological Assumptions and Value Judgments

Fiona Raphaela Spuler,[a,b] Jakob Benjamin Wessel,[c] Julie Jebeile,[d,e,f] Jakob Zscheischler,[g,h] and Theodore G. Shepherd[a]

KEYWORDS:
Climate change;
Statistical techniques;
Uncertainty;
Climate models;
Climate services;
Decision making

ABSTRACT: Statistical bias adjustment has become a common practice to increase the relevance of climate model outputs for impact studies and other societal applications. However, the application of bias adjustment raises fundamental issues identified in the literature, calling into question the credibility of the adjusted climate information. In the attempt to address the usability gap of climate model output despite these unresolved issues, different approaches to bias adjustment have emerged—from applying a single consistent method across studies, selecting the most suitable method for a given use case, to employing an ensemble of bias adjustment methods. This paper examines how these approaches rest on both methodological assumptions and implicit value judgments about what constitutes usable climate information and for whom it is produced. Building on recent literature in the philosophy of science, we propose a framework for evaluating the usability of climate projections in the context of bias adjustment and apply this framework to evaluate the different approaches to bias adjustment. To evaluate the credibility of the adjusted climate information, the paper provides a detailed discussion of two key methodological assumptions underlying different approaches, the interpretation of performance differences of bias adjustment methods and changes to the climate model trend and ensemble through bias adjustment. Through this perspective, we aim to situate bias adjustment in the discussion around usable climate information and the production of climate services, while offering a practical discussion of assumptions for climate impact researchers and climate service practitioners working with bias adjustment methods.

SIGNIFICANCE STATEMENT: Statistical bias adjustment of climate model output has become common practice but raises fundamental issues unresolved in the literature. Informed by the development of the software package ibicus for the comparison and evaluation of bias adjustment methods, this perspective provides both a technical discussion of methodological assumptions of prevalent approaches to bias adjustment and a philosophical reflection on the associated interpretations of usable climate information. Both of these aspects inform the approach to bias adjustment chosen in practice. We argue that the discussion of both technical assumptions and implicit value judgments conducted here is important to guide future method development and can serve as a practical guide to users of bias adjustment and organizations who aim to provide actionable climate services.

DOI: 10.1175/BAMS-D-25-0022.1
Corresponding author: Jakob Wessel, j.wessel@exeter.ac.uk
Fiona Raphaela Spuler and Jakob Benjamin Wessel contributed equally to this work.
Manuscript received 5 February 2025, in final form 7 November 2025, accepted 24 November 2025

AFFILIATIONS: [a] Department of Meteorology, University of Reading, Reading, United Kingdom; [b] The Alan Turing Institute, London, United Kingdom; [c] Department of Mathematics and Statistics, University of Exeter, Exeter, United Kingdom; [d] Institute of Philosophy, University of Bern, Bern, Switzerland; [e] Oeschger Centre for Climate Change Research, Bern, Switzerland; [f] CNRM UMR 3589, Météo-France/CNRS, Centre National de Recherches Météorologiques, Toulouse, France; [g] Department of Compound Environmental Risks, Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany; [h] Department of Hydro Sciences, TUD Dresden University of Technology, Dresden, Germany

1. Introduction

Global and regional climate models (GCMs and RCMs), based on a numerical implementation of physical laws such as thermodynamics and fluid dynamics, are a primary source of information on the future response of the climate system to anthropogenic greenhouse gas emissions and other forcings. Coordinated in the Coupled Model Intercomparison Project (CMIP), for example, projections by different GCMs provide an envelope of plausible future climate changes and the basis for motivating climate mitigation policies and developing adaptation strategies.

Due to the parameterization of subgrid cell processes and other unmodeled or unresolved processes, both global and regional climate models necessarily suffer from shortcomings in the representation of key climatic processes. These shortcomings manifest in discrepancies between the model statistics and the corresponding observational statistics over the historical period, so-called *biases* (Maraun 2016a). For example, a climate model might have biases in the amount of annual rainfall observed in a particular location due to deficiencies in the representation of the extratropical Atlantic storm track (Maraun et al. 2017; Priestley et al. 2023).

Aside from generating fundamental insights about the climate system, climate science is also called to deliver *usable* climate information to society (Lemos et al. 2012; Kirchhoff et al. 2013). Climate information is said to be usable if it is simultaneously credible, salient or relevant to the needs of users of climate information, and legitimate, meaning that the production of the information is fair (Cash et al. 2002; Jebeile and Roussos 2023; Jebeile 2024). However, the biases of climate models in representing meteorological variables relevant to societal impacts call into question the credibility and relevance—and, therefore, the usability—of climate model output, especially at the local level. This is particularly important if one is interested not only in evaluating changes relative to the model climatology, i.e., in terms of anomalies, but also in assessing changes in absolute threshold metrics such as frost days, dry days, or wind speed extremes that are important for most societally relevant impacts.

One widespread way of addressing the biases in both global and regional climate model outputs is statistical bias adjustment, hereafter simply referred to as bias adjustment. Bias adjustment has become a near-standard preprocessing step for using the output of climate models across climate service applications (Fung 2018; Kahlenborn et al. 2021), climate impact studies (Jägermeyr et al. 2021; Laux et al. 2021), and extreme event and impact attribution (Philip et al. 2020; Tradowsky et al. 2023).

Bias adjustment can, in the most general manner, be described as an empirical mapping of a climate model statistic onto the corresponding observational statistic, calibrated over the historical period and applied to a future period (Maraun 2016a). Bias adjustment methods implement this empirical mapping in different ways, ranging from simple adjustments to the mean or variance (e.g., linear scaling), to adjustments by quantile (e.g., parametric or nonparametric quantile mapping), to methods that aim to preserve the trend in specific parts of the distribution [e.g., Inter-Sectoral Impact Model Intercomparison Project, Phase 3 Bias-Adjustment and Statistical Downscaling (ISIMIP3BASD)—Lange 2019; quantile delta mapping—Cannon et al. 2015]. Another approach, which is often listed alongside bias adjustment methods, is the so-called delta change method, which adds the mean trend in the climate model to historical observations. In addition to these univariate methods that are implemented at each grid cell independently, more advanced bias adjustment methods exist that also correct multivariable or spatial structure in climate models (see François et al. 2020 for an overview). We refer to Spuler et al. (2024) for a more detailed overview of commonly used bias adjustment methods and key choices that distinguish available methods.

However, all bias adjustment methods described above come with fundamental issues and rely on strong assumptions that have been highlighted and discussed in the literature. Bias adjustment methods are prone to misuse if only calibrated aspects are evaluated and rely on the assumption that biases are stationary over time (Ehret et al. 2012; Maraun et al. 2017). Furthermore, bias adjustment methods can impair the spatiotemporal consistency of variables and modify the climate change trend and multimodel ensemble spread without physical justification. Table 1 provides a brief overview of the most prevalent issues.

Table 1. Fundamental issues with the bias adjustment of climate model output.

Possible overcalibration and evaluation of noncalibrated aspects	Bias adjustment can make any unrelated variable or even random fields look similar to target observations in terms of the marginal (locationwise) statistical structure. Chandel et al. (2024) illustrate this by bias adjusting a random field as a stand-in for GCM output, and Maraun et al. (2017) by bias adjusting daily temperature over the Southern Ocean to daily precipitation over Europe. In both cases, the authors show that the bias-adjusted fields look statistically similar to the target observational field. Therefore, evaluating improvements in the locationwise correspondence of observational and climate model fields—or in general, calibrated aspects—is nonindicative of successful bias adjustment application and cannot detect "misuse," even when the evaluation is out of sample (Maraun and Widmann 2018a). This issue is even more acute for multivariate methods, where more aspects are explicitly calibrated. Despite this, the evaluation of noncalibrated aspects of bias adjustment is not common in current applications.
Modification of the climate change trend and spatiotemporal consistency	All bias adjustment methods—even so-called trend-preserving methods—modify future trends projected by the raw output of GCMs or RCMs, especially in impact-relevant metrics (Casanueva et al. 2020; Chandel et al. 2024; Dosio 2016; Spuler et al. 2024). This modification of the model trend cannot be justified generally and can lead to implausible future trends (Maraun et al. 2017) as well as unrealistic modifications to climate model ensemble spreads (often interpreted as uncertainty). This will be further discussed in section 4b and the example in this section. In addition, bias adjustment will alter spatiotemporal and intervariable relationships in a climate model; Ehret et al. (2012) and Chandel et al. (2024) argue that this impairs the advantage of using a GCM or RCM and removes the physical insights that can be gained from it.
Stationarity assumption and sensitivity to the choice of calibration period	The application of bias adjustment also rests on strong assumptions, such as the stationarity (time invariance) of biases or the minor role of spatiotemporal field covariances that can, in practice, be hard to verify (Ehret et al. 2012) and have been heavily criticized in the literature. In addition, many variables exhibit strong decadal variability, and the bias adjustment can be highly sensitive to the choice of historical calibration and evaluation period (Chen et al. 2015; Hui et al. 2019; Nahar et al. 2017; Van de Velde et al. 2022).
Use for downscaling	Last, the practice of using (deterministic) bias adjustment for downscaling projections is based on the implicit assumption that local-scale variability is governed entirely by the large-scale driving field. This assumption is often not met, and the use of bias adjustment methods for downscaling has, therefore, been criticised in past literature (Maraun 2013, 2016a,b).

Despite these fundamental issues, the application of bias adjustment remains widespread in current practice to produce information that is seen as relevant for a user or impact study. For example, and importantly, the majority of workflows using climate model projections to run impact models require the use of bias-adjusted climate model data. Impact models, coordinated, for example, in the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP; Frieler et al. 2017), are used to assess the impacts of climate change on wildfires, crop yields, biodiversity, or the water cycle. They provide the basis for a large body of scientific literature and research, the IPCC Working Group (IPCC WG II), and decision-relevant outputs such as National Adaptation Plans and central banks' climate risk scenarios [Network for Greening the Financial System (NGFS) 2021]. Recent studies, however, have shown a large sensitivity of relevant results to the choice of bias adjustment method (e.g., Chae and Chung 2024; Chen et al. 2013a,b; Iizumi et al. 2017; Laux et al. 2021; Padulano et al. 2025; Teutschbein and Seibert 2012).

In this tension between the widespread use of bias adjustment to improve the relevance of climate information and fundamental issues that call into question its credibility, different approaches to bias adjustment have developed in current practice. Each of these approaches is based on a different way of addressing the fundamental assumptions and issues that come with bias adjustment, as well as a different perspective on what usable climate information is and who it is generated for.

In this paper, we use the concept of usability as a framework to evaluate these prevalent approaches to bias adjustment. Focusing primarily on two aspects of usability, credibility and salience, we examine how both methodological assumptions and implicit value judgments, as recommended notably by Pulkkinen et al. (2022), shape current approaches to bias adjustment. While most literature on bias adjustment has focused largely on aspects of credibility, we argue that it is the interplay between credibility and salience that not only influences current practice but can also inform future research and development. Through this analysis, we aim to situate bias adjustment in the discussion around usable climate information and the production of climate services, while offering a practical discussion of assumptions for climate impact researchers and climate service practitioners working with bias adjustment methods.

The two lead authors of the paper have developed the *ibicus* software package (Spuler et al. 2024) for the bias adjustment of climate model output and associated evaluation, which is now used by various researchers and organizations. This perspective paper is informed by resulting engagements with users of bias adjustment, method developers, and impact modelers over the past 3 years.

The remainder of the paper is structured as follows. We first present different approaches to bias adjustment in section 2. Section 3 introduces usability as an evaluative framework to investigate bias adjustment, and section 4 examines, from a methodological perspective, two fundamental issues around bias adjustment, which forms the basis for the subsequent discussion of usability. We then investigate the four approaches to bias adjustment in terms of their methodological assumptions and interpretation of usability in section 5 and provide a discussion and conclusion in section 6.

2. Background: Approaches to bias adjustment in current practice

When selecting the bias adjustment strategy for a given project, researchers and practitioners alike are often guided by a methodological understanding of bias adjustment alongside practical considerations. In the following, we propose a categorization of different approaches to bias adjustment prevalent in both academic and climate service contexts. This categorization is based on a review of existing literature as well as extended engagement with users of bias adjustment methods and is intended to provide a starting point for discussing the assumptions underlying these strategies in the next sections.

The most widely used approach to bias adjustment in current practice is what in this paper is called the **consistency approach**, that is, to use a single bias adjustment method consistently across regions and impacts studied. This is implemented either by working with a published bias-adjusted dataset or adhering to an organizational policy regarding the choice of bias adjustment method.

Examples of this approach include the publication of a bias-adjusted dataset for specific regions (e.g., Dumitrescu et al. 2020; Gergel et al. 2024; Lavoie et al. 2024; Mishra et al. 2020; Navarro-Racines et al. 2020; Xu et al. 2021) or global datasets often published by larger research groups or institutional actors: e.g., NASA Earth Exchange Global Daily Downscaled Projections (NEX-GDDP) (Thrasher et al. 2022), Carbonplan Deep Learning for Statistical Downscaling (DeepSD) (Vandal et al. 2017), and Ensemble Generalized Analog Regression Downscaling (En-GARD) (Gutmann et al. 2022). The adherence to a single bias adjustment method for impact model intercomparison, such as that conducted in ISIMIP based on the ISIMIP3BASD bias adjustment method (Lange 2019, 2021), as well as sister projects such as Fire Modeling Intercomparison Project (FireMIP) (Rabin et al. 2017) or Agricultural Model Intercomparison and Improvement Project (AgMIP) (Rosenzweig et al. 2013), is another example of the consistency approach. Published datasets are often used in hundreds of publications, looking at different regions and using bias-adjusted data for different applications.

In most cases, this consistency approach is justified based on a general-purpose evaluation of the chosen method. This means that the method is determined to be fit for purpose by the provider of the dataset based on an evaluation of one or a selected few applications. Often, this evaluation consists of a published case study and subsequent evaluation of global average properties [see e.g., the published evaluation in Lange (2021)]. In most cases, no further evaluation of the bias-adjusted results is conducted when the dataset is applied by the user, which could be a researcher or climate service practitioner.

However, recent comparison and evaluation of bias adjustment methods show that the performance of a bias adjustment method depends on the use case, i.e., region, climate model variable and impact studied (Casanueva et al. 2020; Chandel et al. 2024; Chen et al. 2013a,b; Lafon et al. 2013; Maraun 2016a; Padulano et al. 2025; Spuler et al. 2024; Teutschbein and Seibert 2012; Zscheischler et al. 2019), which will be discussed in more detail in section 4. In light of this finding, several publications advocate for a use-case-specific evaluation of multiple methods and selection of the bias adjustment method that performs best in terms of reducing biases of calibrated aspects and preserving or modifying trends in line with the assumptions of the case study. We call this approach to bias adjustment the **use-case-specific evaluation approach**. In this approach, the choice of bias adjustment method depends on a specific-purpose evaluation, that is, the variable and impact studied, as well as the physical source of bias in the climate model (Addor et al. 2016; Eden et al. 2012; Gudmundsson et al. 2012; Maraun 2016a; Maraun et al. 2017), which can include so-called process-based bias adjustment (Maraun et al. 2017). This recommendation has been followed in a range of publications applying bias adjustment in practice, for example, Olschewski et al. (2023) and Tootoonchi et al. (2023).

Another approach to interpreting performance differences of bias adjustment methods is what is here called the **ensembling approach**. This approach is motivated by recent findings that show that the choice of bias adjustment method can lead to large differences in the resulting ensemble of bias-adjusted climate model projections (e.g., Ho et al. 2012; Lafferty and Sriver 2023). In the ensembling approach, these differences are understood as an additional source of uncertainty in local-scale future climate projections. The proposed response based on this understanding is to “sample" the additional uncertainty by applying several bias adjustment methods to the ensemble of climate models used and interpret these results probabilistically. This approach has so far primarily been proposed in academic publications such as Laux et al. (2021), Lafferty and Sriver (2023), Chen et al. (2013b), Iizumi et al. (2017), Liess et al. (2022), Wootten et al. (2017), Senatore et al. (2022), and others.

Given the fundamental issues with currently available bias adjustment, another possible strategy is to attempt to circumvent the use of bias adjustment methods altogether, called the **no bias adjustment approach** in this paper. Available strategies for circumventing bias adjustment include working only with climate change trends and anomalies (the difference between projected and historical values). This can be considered a form of implicit bias adjustment; however, challenges arise when absolute values are required to drive impact models. Calculated trends can be reported as such, used to drive downstream models such as (change factor) weather generators (Maraun et al. 2010; Maraun and Widmann 2018b), or in principle integrated into different approaches that start from a specific impact or vulnerability and assess plausible futures in a scenario-based or scenario-neutral approach (Guo et al. 2017; Prudhomme et al. 2010; Shepherd et al. 2018; Wilby and Dessai 2010). Another group of methods explicitly links present-day biases in climate models to future uncertainty, formulating statistically coherent models to capture (and constrain) future uncertainty independently of possible model biases. This encompasses statistical work on multimodel ensembles (Chandler 2013; Sansom et al. 2021) as well as approaches based on emergent constraints (Hall et al. 2019; Williamson and Sansom 2019). Similarly, model weighting that is in some way based on present-day performance can, depending on the specific method, be interpreted as an implicit bias adjustment (Knutti et al. 2017; Sippel et al. 2016).

3. Usability as an evaluative framework

Usability has been conceptualized in different ways in the climate services literature as well as the environmental social science and philosophy of science literature (e.g., Bremer et al. 2019; Kirchhoff et al. 2013; Maraun et al. 2010; Maraun and Widmann 2018c; Skelton et al. 2017). Here, we define usable climate information as information that simultaneously meets the requirements of *credibility*, *salience*, and *legitimacy* (following Cash et al. 2002, 2003). Credibility requires that the knowledge production process, in particular modeling, meets the standards of rationality and scientific plausibility as defined by the scientific community. Salience, in turn, means that information meets the needs of the downstream users. Finally, information is legitimate if its production is unbiased, i.e., does not serve the interests of some groups only, and if its production is fair, i.e., justice oriented, and, thereby, addresses the needs and the values of different stakeholders; this carries an ethical component.

The following analysis will focus primarily on the aspects of salience and credibility. Concerns regarding the production of legitimate climate information, especially regarding the underrepresentation of researchers and stakeholders from the Global South (Dike et al. 2018; Rodrigues 2021; Schipper et al. 2021; Tandon 2021), require further attention from the climate research community and are discussed in the conclusion of this paper in so far as they relate to issues distinguishing different approaches to bias adjustment.

To apply the concept of usability for the evaluation of bias adjustment, we further refine the concepts of credibility and salience, as different aspects of both concepts can be more or less important in different contexts. Epistemic and nonepistemic values that are part of credibility and salience will be italicized throughout the paper. We also note that a clear separation between credibility as relating to the knowledge production process and salience as relating to the user considerations is to some extent artificial. As we will discuss later, the two concepts are interdependent to some degree, meaning that credibility in a specific application context can be dependent on what is considered relevant in this context, i.e., conditional on salience criteria.

In the context of bias adjustment, credibility covers our *justificatory capacity* of the method. This means that the assumptions underlying many bias adjustment methods require justification and testing for the method to be seen as credible. Related is also the notion of *physical interpretability*, which is a necessary condition for, for example, using future climate projections beyond the observed period. The concept of credibility also covers *representational* and *empirical accuracy* of the method. Empirical accuracy encompasses the match of individual variables or their interdependence to observations. Representational accuracy covers the fidelity of the bias adjustment method to the understanding of the sources of bias, as well as the faithfulness to the laws of physics of both the bias-adjusted variable and its relationship to large-scale drivers or other variables. Credibility also comes with the possibility of clearly delimiting the domain of validity of the method and, therefore, the *context specificity* of the method application. It can also include *reproducibility* to allow independent verification of a method. *Intercomparability* and *standardization* across applications are also often seen as implying credibility: If one uses the same method as everyone else, one can appear scientifically more credible. But social conformism within science is certainly not a genuine mark of scientificity. Finally, *capturing the full uncertainty* could also be interpreted as enhancing scientific credibility through sampling.

Salience, for one, requires that data are *relevant for users* in their application context. For example, climate model output that is biased in terms of extreme precipitation occurrence will in general—without further processing such as bias adjustment—be inadequate to study changes in flooding probabilities. On the other hand, data that have been bias adjusted in such a way that the physical consistency between temperature and precipitation is broken will not provide relevant or adequate information for studying hot and dry compound extremes. Salience thus requires that information meets users' demands on adequacy for specific use cases.

Salience, as defined in this paper, also requires *ease of use* of the method or information, meaning that it can be practically understood and applied by the users. Ease of use encompasses several components. It includes the *rapidity* of the provided method or information, that is, how fast the data can be downloaded or how much computational resources the method takes up and how fast it runs. These considerations can be crucial for determining whether the information or method can be used in specific applications or to answer certain impact-related research questions. It may also include the possible *opacity* of the method to the users. Opacity here refers to the extent to which the internal workings or underlying mechanisms of a method are hidden from, or not directly understandable by, the user. While epistemic opacity of simulations (Humphreys 2009) is often considered as epistemically detrimental as it hampers scientists' understanding of their models, and thereby of the target phenomena, epistemic opacity may still be in practice beneficial for the users that can work more efficiently with the model as a black box and do not have to spend too much time engaging themselves with the details of the model (Dowling 1999). Ease of use also encompasses the *integrability* of the provided method into the existing workflow of a user. Finally, it includes *simplicity of usage*, which concerns the straightforwardness of the practical applications. For example, users might prefer downloading data rather than running code in a specific programming language, and different data sources might be more or less accessible.

The quest for credibility and the quest for salience may conflict with one another if, for instance, meeting the needs of users generates additional uncertainties within the cascade of uncertainty that already characterizes the model chain of climate impact research. There is, therefore, no universal good and useful way of representing the climate system, in the sense of maximizing both credibility and salience across use cases. In what follows, we will explore the priorities and the resulting trade-offs made by different approaches to bias adjustment regarding these two aspects.

4. Fundamental issues underlying credibility considerations

In this section, we discuss two fundamental assumptions which underlie credibility justifications of the different strategies in technical detail. This discussion then provides the basis for evaluating the different approaches to bias adjustment in terms of their understanding of usability along the multiple dimensions of the concept in the next section.

Given the issues mentioned in Table 1, the evaluation of a bias adjustment method to establish its credibility can be based on different lines of evidence: 1) the evaluation of the performance of the method over a historical validation period, 2) the validity of structural assumptions underlying the method, and 3) evaluation of the preservation or nonpreservation of the future climate trend. Pseudo-reality experiments can also provide additional evidence for or against the usage of a method (Hui et al. 2019; Maraun 2012; Velázquez et al. 2015). Here, we first discuss the ability of a general-purpose evaluation, as defined in section 2 to provide evidence for (1) and (2). Justifications of trend modifications (3) are discussed in section 4b.

a. Performance differences of bias adjustment methods and implications for general-purpose evaluation. Various publications have pointed out that the performance of a bias adjustment method depends on the variable, region, and climate model of interest (Chen et al. 2013a,b; Lafon et al. 2013; Maraun et al. 2017; Spuler et al. 2024; Teutschbein and Seibert 2012; Velázquez et al. 2015; Volosciuk et al. 2017). One reason for the observed differences in performance is that the ability of a bias adjustment method to correct a bias depends on the physical sources of the bias—a large-scale circulation bias might be more difficult to correct using conventional methods than a local-scale bias due to orographic resolution. This means that methods will perform differently across use cases, depending on the source of the bias (Addor et al. 2016; Eden et al. 2012; Maraun et al. 2017, 2021). Furthermore, the assumptions underlying different bias adjustment methods might be adequate or not, depending on the variable and impact studied. For example, a method based on a parametric distribution fit might be robust in some situations but fail in other cases when the distributional assumption—such as the choice of a (censored) gamma distribution to fit daily precipitation data—is not met (see e.g., Lafon et al. 2013). Bias adjustment methods are commonly applied over a running window, the length of which depends on the variable studied and possible rapid changes in this variable through, for example, monsoon onsets. However, if running window lengths are chosen too short, this might impede robust distributional fits. A nonparametric correction might work well for the body of the distribution but might not be appropriate for extremes. Methods based on statistical extreme value theory in return can be more appropriate in the tails but difficult to automate across locations (e.g., Volosciuk et al. 2017; Scarrott and MacDonald 2012). In addition, the assumption that biases are stationary is more problematic for some variables (e.g., precipitation) than for others (e.g., temperature) (Chen et al. 2015; Christensen et al. 2008; Hui et al. 2019, 2020; Maraun 2012; Van de Velde et al. 2022). This means that depending on the variable, impact, and region studied, different bias adjustment methods might perform better over the validation period (1) and be more suitable in terms of structural assumptions made (2).

These findings challenge the ability of a general-purpose evaluation, that is, the evaluation of a bias-adjusted data product by the provider detached from the application, to identify the method that best fulfils evaluation criteria (1) and (2). However, general-purpose evaluation could still be justified if differences between methods are small enough not to impact downstream results or if the overall adjustment made by the method is small.

Comparing different bias adjustment methods globally, Lafferty and Sriver (2023) demonstrate that performance differences between methods lead to large differences in the resulting

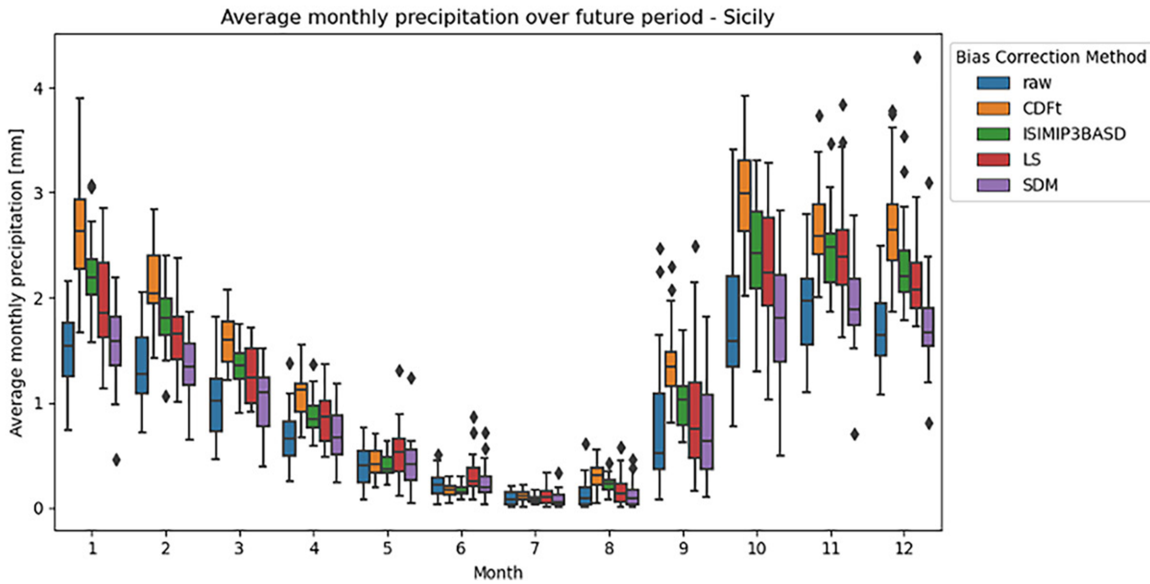

Fig. 1. Ensemble spread of a 25-member CMIP6 ensemble for future monthly average precipitation over Sicily under SSP58.5, under different bias adjustment methods.

bias-adjusted data. Building on the approach proposed by Hawkins and Sutton (2009), they decompose the variance of an ensemble of climate models and bias adjustment/downscaling methods into uncertainty coming from the choice of scenario, model, downscaling/bias adjustment method, and internal variability. In their selected examples, the uncertainty of the choice of bias adjustment and downscaling method combined contributes up to 70% of the total variance, depending on the region and variable studied. Investigating the effect of these differences on downstream results, Laux et al. (2021) find large sensitivities in the results of agricultural impact models to the choice of method (up to twice as large as the variance introduced by the GCM–RCM), while Chen et al. (2013a,b) and Wang et al. (2020) find similar sensitivities for hydrological models. To illustrate this issue, Fig. 1 in the example in the next subsection shows a CMIP6 ensemble spread for future average precipitation over Sicily under different bias adjustment methods. The individual methods significantly alter the CMIP6 ensemble spread, in some cases substantially inflating, reducing, or shifting the CMIP6 ensemble nonuniformly across months.

Finally, Ehret et al. (2012) further argue that if climate model biases are on a scale that makes direct model output unusable for impact modeling, then the modifications made by bias adjustment will be of equal magnitude and, therefore, not small. Even if the model biases at each grid cell are indeed small, or large only in a small number of cases, Ehret et al. (2012) argue that the magnitude of the bias and the impact of bias adjustment require specific evaluation. This means that an evaluation that is indiscriminate across use cases or regions will in general not provide adequate *justificatory capacity* for the application of a given method.

Example 1: Implications of performance differences of bias adjustment methods. Figure 3 shows the monthly average precipitation (2070–2100) projected by a 25-member CMIP6 ensemble (see the appendix) under shared socioeconomic pathway (SSP) 585 over Sicily before and after the application of four different bias adjustment methods. The choice of bias adjustment method significantly alters the spread across climate models—which is often interpreted as uncertainty in future projections. The spread can be increased using some methods, such as the cumulative distribution function transform (CDFt) method in this case, and decreased using other methods, e.g., the ISIMIP3BASD method. This effect is nonuniform across months. For example, ensemble spreads in July are often maintained but shifted and inflated in October.

As average monthly precipitation is a fairly aggregate metric, strong changes to the ensemble spread through bias adjustment raise questions about the interpretation of uncertainty in future projections. Individual analysts consistently employing one bias adjustment method might get very different indications about the scale and the uncertainty in future projections than users of a different method. This highlights the challenges of a general-purpose evaluation of bias adjustment methods.

b. Justificatory capacity and physical interpretability of trend modifications through bias adjustment. The modification of the simulated climate change trend is often an unintended but critical side effect of bias adjustment. Examining plausible future changes in the climate system caused by anthropogenic climate change is the key reason for generating future projections in the first place. Therefore, any modification of this simulated trend and the implications for the credibility of the resulting information need to be carefully investigated.

Simple bias adjustment methods such as linear scaling or quantile mapping modify projected trends, i.e., the forced response to anthropogenic emissions, simulated by a climate model. More advanced, so-called trend-preserving methods exist that often build on quantile mapping and attempt to preserve trends in the mean, individual quantiles or the whole distribution (Maraun 2016a). These include methods such as the CDFt (Michelangeli et al. 2009; Vrac et al. 2016), quantile delta mapping (Cannon et al. 2015; Li et al. 2010), or ISIMIP3BASD (Lange 2019).

However, even so-called trend-preserving methods frequently modify the climate change trend for several reasons. For one, all methods are based on some assumptions about which trends to preserve and how to represent them. For example, a method such as linear scaling, which subtracts a bias in the mean from a climate model (or the related delta method), will preserve trends in the mean but not trends in the rest of the distribution. Similarly, a method such as ISIMIP3BASD might aim to preserve trends in certain quantiles but will still modify trends in threshold-based impact metrics such as dry days or heat wave days or in spatiotemporal metrics such as dry spells, or higher moments of the distribution, as discussed by Casanueva et al. (2020) and Spuler et al. (2024), which another method might aim to preserve. Second, any trend-preserving method is necessarily based on assumptions about how to parameterize trends. For example, trends can be assumed additive or multiplicative, they can be assumed constant or varying over an application period/with seasonality, and the bias correction can be based on distributional or other structural assumptions. The success of any trend preservation is contingent on how well these assumptions match the underlying data, which can depend on the variable, impact, and region studied. Third, even if most assumptions are met, empirically trends are not guaranteed to be preserved as assumptions in other parts of the bias adjustment methods might not be entirely adequate and model fits might be imperfect. Thus, trend-preserving methods frequently modify trends in certain attributes of interest (see, e.g., Cannon et al. 2015; Spuler et al. 2024; Padulano et al. 2025; Astagneau et al. 2025).

Trend modifications through different bias adjustment methods can have large impacts on downstream results and impact models, as illustrated in example 1. In addition, the analysis presented in example 2 below shows that bias adjustment not only changes the climate change trend but also impedes the *physical interpretability* of this trend by altering the relationship between local variables such as precipitation and their large-scale dynamical drivers. The resulting climate model spread after bias adjustment is, therefore, not only modified but also made physically less interpretable and coherent which has implications, for example, for the study of compound risks.

Trend modifications can be justified either pragmatically by arguing that certain climate model trends are not relevant for a specific application—for example, trends in the upper tail of precipitation might not be relevant when studying dry spells—or based on physical and statistical reasoning. The former, pragmatic, justification will only be valid for a specific use case. Similarly, any physical or statistical justification of a trend modification through bias adjustment relies on the assumption that gridcell-level biases in historical simulations directly relate to the gridcell-level validity of future projections. However, this assumption is not valid generally. Rather, it can be justified only based on a (physical) understanding of the sources of model bias that are specific to a certain variable and region: For example, Gobiet et al. (2015) argue that climate models have intensity-dependent biases in the temperature trend over Europe relevant for their chosen application that can be improved through quantile mapping. However, in the absence of a physical justification or knowledge of the source of the bias, Maraun et al. (2017) argue that trend modifications through bias adjustment should be avoided.

Example 2: Modification of the physical interpretation of the climate model ensemble spread. As an example of how bias adjustment alters the physical interpretability of climate model trends, we analyze dynamical storylines of future precipitation before and after bias adjustment. Physical storylines represent plausible unfoldings of future trends conditional on the evolution of large-scale drivers in the climate system such as the strength of future Arctic amplification (Shepherd et al. 2018; Zappa and Shepherd 2017), allowing a dynamical interpretation of the spread in an ensemble of climate models.

We bias adjust precipitation over the Mediterranean region in 30 CMIP6 models under SSP585 using seven methods (CDFt, scaled distribution mapping, the ISIMIP3BASD method, linear scaling, quantile mapping, the latter two both with and without running window (RW) implementation—described in detail in Spuler et al. 2024). We then follow the approach as presented in Zappa and Shepherd (2017): Based on a pattern scaling assumption, we regress the ensemble anomaly of the pattern of precipitation change (scaled by the global-mean temperature change) onto large-scale drivers, namely, tropical warming, polar warming, and change in stratospheric polar vortex strength. The results show the influence of each of the drivers on the difference of the model projection from the ensemble mean, or in other words the physical contribution of the large-scale driver to the ensemble spread. More information on the data and method can be found in the appendix.

Figure 2 compares the resulting influence of tropical warming on the ensemble spread between the raw climate model and the model bias adjusted with one of the seven methods (plots for the other drivers can be found in the appendix). We find that the influence of the remote driver on the ensemble spread changes strongly depending on the bias adjustment method used. The influence can be strengthened, weakened, or fully removed in some regions, and this effect appears to be nonuniform across bias adjustment methods.

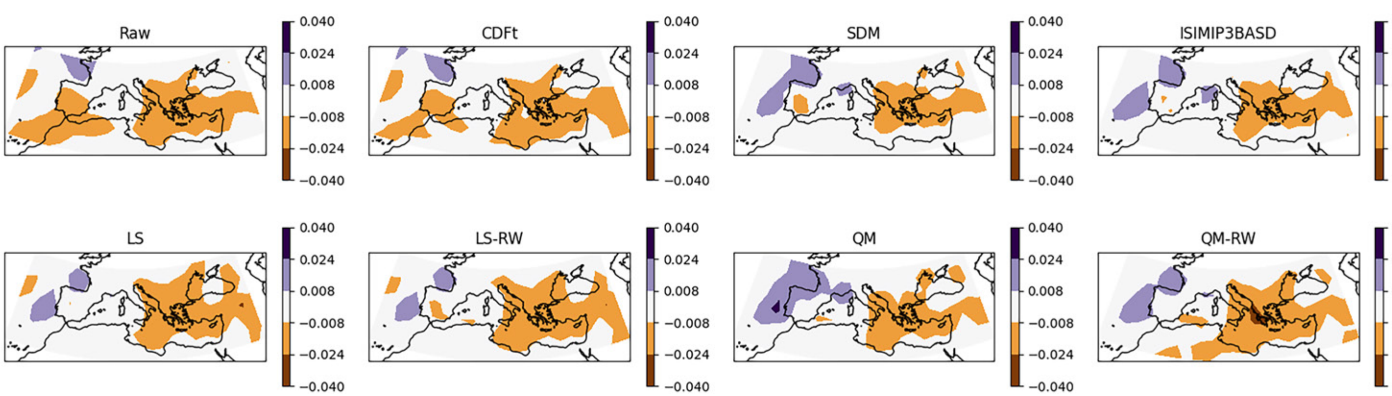

Fig. 2. **Influence of tropical warming on the climate model ensemble spread for precipitation [mm (K day)$^{-1}$] for the raw climate model ensemble and for the bias-adjusted climate models in a storyline approach (see the appendix).**

To some extent, this result is to be expected, as all bias adjustment methods investigated here calibrate the model to observations on a gridcell by gridcell basis. By doing so, they also change the multivariate structure of the precipitation (Spuler et al. 2024) as well as its relationship to dynamical drivers such as the zonal wind and associated storm-track location, which in Zappa and Shepherd (2017) was shown to mediate the response of Mediterranean precipitation to the three large-scale drivers investigated. However, this implication of bias adjustment is often not appreciated in practice. This example illustrates that bias adjustment not only modifies the climate change trend but also impedes the dynamical interpretation and understanding of future ensemble spread and its multivariate structure which is relevant in particular for the study of spatial, temporal, or multivariate compound events (Zscheischler et al. 2020) in future climate.

5. Evaluating approaches to bias adjustment through the lens of usability

The four approaches to bias adjustment introduced in section 2 navigate the tension between fundamental issues of bias adjustment and closing the usability gap of climate information in different ways. Based on the discussion of two key methodological issues in the previous section, we now apply usability as introduced in section 3 as an evaluative framework to examine how different methodological assumptions and value judgments about usable climate information shape the choice of approach to bias adjustment.

Figure 3 illustrates the credibility assumptions made by different approaches to bias adjustment. In particular, it highlights how the different credibility assumptions relate to each other and how the two fundamental issues which were discussed in the previous section—the interpretation of modifications to the climate change trend and performance differences of bias adjustment methods—are addressed in different approaches.

a. Consistency approach. In the consistency approach to bias adjustment, a single bias adjustment method is selected either at institutional level or by a data provider based on a general-purpose evaluation as defined in section 2 and then applied across regions and impacts.

In terms of salience considerations, this strategy prioritizes the *ease of use* of the bias-adjusted information, in particular in terms of *integrability* into existing workflows, *simplicity of usage*, and *opacity* of the method. If a consistent method is decided across use

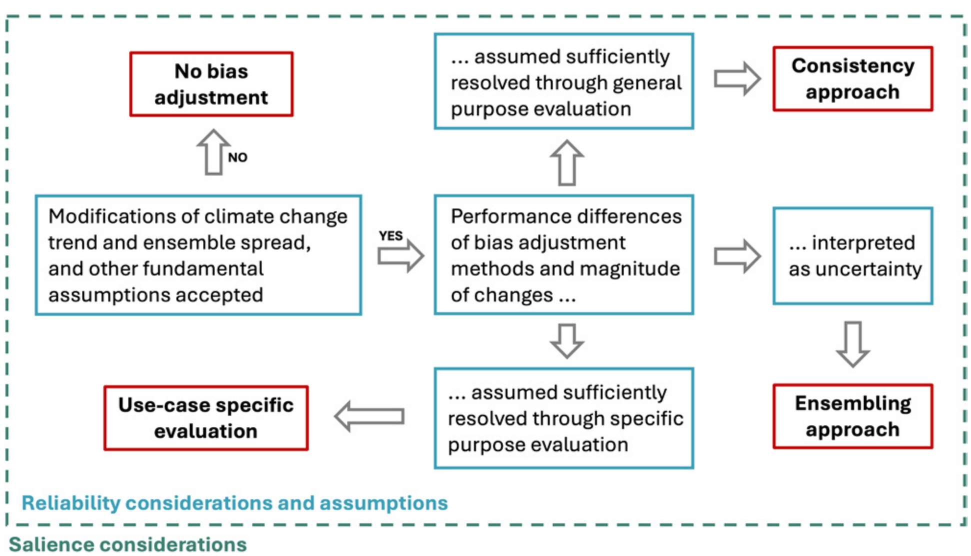

Fig. 3. Assumptions that underly credibility considerations of different approaches to bias adjustment and their relationships.

cases, the burden of engaging with the complexities of bias adjustment is removed from the user which facilitates integration of bias adjustment with subsequent tasks such as downscaling or impact modeling. This choice is not uniquely made in the case of bias adjustment—the intricacies involved in the different steps of the climate modeling chain already necessitate dividing labor and expertise across teams and institutions. For example, most users of climate models have no detailed knowledge of subcomponents of the global or regional climate models used. This has led to a considerable drive toward standardized datasets through coordinated model experiments such as CMIP and Coordinated Regional Climate Downscaling Experiment (CORDEX). Similarly, the consistency approach argues that not every user has the technical knowledge to select the most appropriate bias adjustment method. An already bias-adjusted dataset or preselected method is, therefore, suitable as it assigns the responsibility for the choice of method to experts' external to the use case.

However, the consistency approach to bias adjustment thereby assumes that a general-purpose evaluation by the data provider or institution can sufficiently ensure credibility in terms of both the *justificatory capacity* of modifications to the climate change trend as well as the *representational* and *empirical accuracy* of the resulting information. However, based on the discussion of existing literature and the examples analyzed in section 4, we find that neither of these assumptions holds in practice. Arguments made for justifying modifications to the climate change trend hold only conditional on the context of the application. Similarly, the performance of a bias adjustment method depends on the impact, region, and variable studied.

Prioritizing *standardization* and *comparability of methods* across use cases in the consistency approach can be taken as furthering credibility; however, as discussed in section 3, these two aims are not genuine criteria for scientific credibility. Standardization can, on the other hand, be argued to increase the *relevance* of the data for the user. However, this holds primarily if the use case is global or involves model intercomparison—either across regions in a global assessment or across use cases within a region. If the aim of a given user is to produce the most accurate set of (local) climate projections in a specific application, standardization in line with the consistency approach will not necessarily produce the most relevant information.

The choice to prioritize salience, specifically *ease of use*, therefore, comes at a cost to the credibility of the resulting climate information as neither justificatory capacity of modifications to the climate change trend or *representational* and *empirical accuracy* can be guaranteed based on general-purpose evaluation. However, if the use case is global or involves model intercomparison, standardization of the bias adjustment method improves its *relevance* for the user, another salience consideration.

b. Use-case-specific evaluation. The use-case-specific evaluation approach to bias adjustment encourages the choice of method based on the consideration of context-specific assumptions and subsequent evaluation of bias-adjusted results.

The focus of this approach is to improve the credibility and *relevance* of the resulting climate change information. By situating the evaluation of bias adjustment methods in a specific use case, the context-specific assumptions of individual methods can be justified. In addition, when comparing multiple methods, the most appropriate method in terms of empirical performance, trend preservation, and use-case-specific assumptions can be selected. This enhances both *representational* and *empirical accuracy* as well as *justificatory capacity*. Furthermore, by considering specific user demands, this ensures that the information is relevant for the downstream analysis.

To identify the best-performing bias adjustment method for a given use case, evaluation needs to be based on multiple lines of evidence, as outlined in section 4a. This, however, comes with challenges. Evaluating the performance of a bias adjustment method requires selecting the climate metrics relevant to the impact studied. However, impact models, such as hydrological or fire models, can depend on a range of climate metrics, where the sensitivity of the model output to biases in the different metrics is not always known a priori. Furthermore, uncertainty in the results of historical evaluation is introduced through observational uncertainty, as well as the choice of validation period due to climate variability (Casanueva et al. 2020; Chen et al. 2015; Jain et al. 2023). Structural assumptions underlying the different methods, such as the suitability of a selected parametric distribution or assumptions about the nature of climate change trends, can be hard to verify and require judgments on the side of the researcher applying bias adjustment. This means that multiple methods can be adequate under somewhat equally defensible assumptions (Ho et al. 2012).

Furthermore, the use-case-specific evaluation approach currently comes at the cost of *ease of use* in most applications. To enable the implementation of this evaluation by the user, different bias adjustment methods have to be readily available and use-case-targeted evaluation needs to be straightforward enough to be conducted within the time frame of the research or climate services project. The publication of several bias-adjusted datasets, for example, by CORDEX (2022) and evaluation frameworks such as these developed as part of the VALUE project (Gutiérrez et al. 2019; Maraun et al. 2015, 2019) or sector-specific protocols (e.g., Galmarini et al. 2019) support this. In addition, open software tools such as climate4R framework (Iturbide 2019), the xclim library (Bourgault et al. 2023), or the ibicus software package (Spuler et al. 2024) can help users to easily implement different methods and conduct evaluation. However, even given the current availability of open software, evaluation frameworks, and datasets, the task of evaluating and selecting the bias adjustment method best suited to a use case requires some familiarity and knowledge with the approach that users might not necessarily have.

In addition, the limitations of currently available bias adjustment methods imply that a credibility gap remains, for example, if trends in certain metrics are still modified in an undesirable manner or if relevant metrics are not calibrated by the bias adjustment method. These persistent limitations motivate research into new bias adjustment methods, as well as alternatives to bias adjustment which are discussed in section 5d.

c. Ensembling approach. Performance differences across bias adjustment methods can also be interpreted as uncertainty, which underlies the ensembling approach to bias adjustment.

On first glance, the ensembling approach appears to prioritize credibility. It propagates the lack of knowledge about the most suitable bias adjustment method down to the user, presenting it as uncertainty about future projections, which the user needs to engage with when using the information. It can be argued that it thus maximizes the likelihood of the truth falling within the spread, while potentially compromising the *relevance* of the resulting output.

However, the interpretation of performance differences between bias adjustment methods as a source of epistemic uncertainty (see e.g., Chen et al. 2015; Ho et al. 2012; Hui et al. 2019; Velázquez et al. 2015; Wang et al. 2018) can be disputed when considering on what basis the application of any particular bias adjustment method can be justified. As discussed in section 4, each bias adjustment method is justified only by assumptions which are use-case dependent. These can include assumptions about which quantiles or threshold-based climate impact drivers to correct based on the climate impact studied, which distribution to choose based on the variable and region studied, and what kind of trends to preserve in which statistics. There is, therefore, no general justification for a bias adjustment method, in the way physical laws can be argued to provide for GCMs (Baumberger et al. 2017; Knutti 2008; Oreskes et al. 1994).

Therefore, when different bias adjustment methods are used in an ensemble, it is not an ensemble of equally justified methods that is considered but rather an ensemble of statistical assumptions, some of which can be more or less justified. For example, models with very different assumptions—such as trend-preserving and nontrend-preserving methods—are often bundled together in an ensemble, or methods whose assumptions are clearly disputable are included. In particular, recent applications of the ensembling approach did not conduct any (published) prior evaluation of the methods included in the ensemble (see e.g., Lafferty and Sriver 2023; Laux et al. 2021), which raises the question whether this lack of evaluation on the researchers' side should be interpreted as uncertainty. On the contrary, including methods with opposing assumptions might make the resulting climate model spread less *interpretable*—and thereby less credible—and possibly overly large.

Furthermore, the *ease of use*, in particular *rapidity*, is impeded by the computational cost associated with running an impact model several times using different bias adjustment methods. However, as little to no modification of existing impact modeling pipelines is required by the approach, this poses mostly a resource challenge rather than a challenge to the researcher as both *simplicity of usage* and epistemic *opacity* are relatively unaffected. Particularly in well-resourced application contexts, the *ease of use* is, therefore, arguably only marginally impacted.

d. No bias adjustment approach. Finally, the no bias adjustment approach assumes that the potential gains in salience achieved through bias adjustment in terms of *relevance* to the user do not outweigh the credibility lost in the process of bias adjusting climate model projections. Possible alternative approaches to bias adjustment are outlined in section 2 and include working directly with trends, scenario-neutral, or storyline approaches and local weather generators or constraining an ensemble of climate models based on their performance over the historical period.

However, these approaches also come with several limitations and assumptions: For example, working with trends is difficult if spatiotemporal fields or many climatic variables with adequate dependence structure are required to drive complex impact models (see e.g., Best et al. 2011; Clark et al. 2011 for an example of what is required to drive a state-of-the-art land surface model). Scenario-neutral approaches are often computationally expensive and can be difficult to scale for larger application areas (e.g., Broderick et al. 2019; Bennett et al. 2021), impeding *relevance* and *ease of use*. Finally, constraining model ensembles on the basis of performance over the historical period is an active area of research. Obstacles to operationalizing existing approaches widely include the significant role of observational uncertainty and internal variability (Jain et al. 2023), as well as the fact that detailed research is often required to understand whether a certain climate model is getting the correct distribution for the right reasons. Furthermore, alternative approaches require rethinking the scales and starting points to produce climate information. Weather generators driven by the output of climate models, for example, produce local information and do not lend themselves well to global intercomparison assessments.

Available alternatives to bias adjustment, therefore, in many applications, fail to close the salience gap which arises due to prevailing biases in climate models. As will be discussed in section 6, the approach chosen to address this usability gap depends on the considered starting point for producing climate information.

6. Discussion and conclusions

The most general justification for applying bias adjustment to the output of climate model data is that it makes the resulting information more usable. In the tension between fundamental issues and practical considerations, different approaches to bias adjustment have developed, from consistent application to use-case-specific evaluation, ensembling, and no bias adjustment. In this paper, we evaluate these different approaches through the lens of usability, examining underlying methodological assumptions and value judgments.

The consistency approach prioritizes *standardization* for intercomparison projects and *ease of use* of bias adjustment. This comes at the cost of the credibility of the resulting information, in particular as *empirical accuracy* and *justificatory capacity* cannot be guaranteed based on a general-purpose evaluation. The use-case-specific evaluation approach, on the other hand, focuses on ensuring the credibility and *relevance* of bias-adjusted information for a specific application, but this comes at a cost to *ease of use*, given available tools. The ensembling approach aims to address credibility through "sampling" of different bias adjustment methods in an ensemble. However, the credibility gained by including many bias adjustment methods in an ensemble can be disputed, and the approach also compromises on *relevance* and *ease of use*. Finally, the no bias adjustment strategy argues that the gains in salience through bias adjustment do not outweigh the losses in credibility, but alternatives proposed are largely not operational and come with open issues of their own.

Overall, none of the currently available approaches to bias adjustment produces climate information that closes the usability gap in the sense of being both credible and salient across use cases. For some applications, it is, therefore, questionable whether the demands of credibility and salience can simultaneously be met. An example of this can be found in the physical risk scenarios published by NGFS (2021), which use the consistency strategy, in particular the ISIMIP3BASD method, and produce scenarios used by central banks across the world to stress test their financial systems to climate risks. A consistent bias adjustment method is necessary as it implies comparability of results when assessing risks across regions for organizations such as the International Monetary Fund. However, as demonstrated by Laux et al. (2021), the choice of bias adjustment method can significantly impact downstream results. Therefore, the indiscriminate application of a method across variables and regions risks the distortion of the information on future impacts and associated risks to financial stability. Here, salience demands stand in opposition to credibility ones, pointing to potential limits of *relevance* imposed by scientific integrity.

The discussion in this paper shows that the choice of bias adjustment approach and the associated best available climate information in terms of usability depends on which aspects of usability are deemed important in the production of climate information and which assumptions are accepted. To some extent, the difference between prioritizing *relevance for global studies or model intercomparison* in the consistency approach, as opposed to *locally relevant* and credible information in the use-case-specific evaluation approach, can be mapped to the distinction between "top-down" climate information which starts from a GCM/RCM in a specific scenario and "bottom-up" climate information which starts from a specific vulnerability and prioritizes the relevance of the information to the impacted community (Kelly and Adger 2000; O'Brien et al. 2007). In this distinction, different approaches to bias adjustment also take different views on the legitimacy of the resulting climate information as the third component of usability, specifically in the sense of "taking into account the values, interests, and concerns of different stakeholders." While bottom–up approaches are, in many cases, preferred when developing local climate information, the time and skill required, given currently available tools, to apply bias adjustment prevent the use-case-specific evaluation approach from being easily applicable in local contexts.

The evaluation conducted in this paper raises important questions on how to address this existing usability gap—through changes in current practice, tool development and deployment, and future research into method development.

In terms of changes to current practice, examples of best practice exist that could already, given available methods and tools, be adopted more widely. So-called process-oriented bias adjustment methods, which take into account physical sources of bias, could be more widely used. For example, Manzanas and Gutiérrez (2019) condition their bias adjustment on phases of ENSO, and Verfaillie et al. (2017) adjust dependent on synoptic weather types. Also, there are good examples of published evaluation; for example, Lehner et al. (2023) assess a number of bias adjustment methods over Austria, and the authors are aware of at least one meteorological service which is currently comprehensively evaluating a number of different bias adjustment methods for data provision for a range of different use cases. These examples primarily fall under the use-case-specific evaluation approach. Available tools can support their implementation, including software tools such as ibicus (Spuler et al. 2024) and the climate4R framework (Iturbide et al. 2019) or evaluation frameworks as developed by the VALUE project (Gutiérrez et al. 2019).

However, implementing this best practice requires some technical expertise on the side of the user, which can be a climate impact research or climate service practitioner, as well as time and associated funding spent on bias adjustment. Moving forward, the shortfall in the ease of use of the use-case-specific evaluation approach can be addressed by continuing the development of available tools for comparing and evaluating existing bias adjustment methods and mainstreaming them into widely used pipelines for processing climate information.

Finally, further method development guided by the usability framework presented in this paper can support closing the usability gap in different application areas. Machine learning approaches are able to improve the multivariate structure which can be valuable especially for studying compound extremes (e.g., Hess et al. 2023). However, they still rely on many of the same assumptions as existing methods and can be much harder to evaluate in the presence of fewer uncalibrated aspects (Maraun 2016a). Process-based bias adjustment methods, as well as some of the approaches listed in the no bias adjustment approach such as constraining an ensemble of future projections based on historical performance, can be improved through further research. On the other hand, bottom-up climate information that is, for example, coproduced with local actors might require different approaches that can be integrated into information pipelines working backward from a local context, or “intermediate technologies” called for by Rodrigues and Shepherd (2022). These local requirements might favor the advancement of methods beyond bias adjustment, such as methods that link local weather generators directly with changes in large-scale drivers.

The aim of this contribution was to situate bias adjustment in the discussion on the usability of climate information, provide theoretical foundations for choices and assumptions often made implicitly in practice, and offer a guide for researchers and practitioners working with bias adjustment. With this contribution, we hope to support the reflection on the different choices when approaching bias adjustment as well as its subsequent evaluation and offer a perspective on possible ways forward considering the interpretation of the usability of climate information.

Acknowledgments. We thank Dr. Uwe Ehret and one anonymous reviewer for their thorough and thoughtful reviews of the manuscript, that helped to improve the final paper significantly. F. R. S. is funded by the Advancing the Frontiers of Earth System Prediction (AFESP) Doctoral Training Program and acknowledges support by the Alan Turing Institute’s Enrichment Scheme. J. B. W. was supported during this work by the Engineering and Physical Sciences Research Council (EPSRC) under Grant Award 2696930. J. J. acknowledges support from the Swiss National Science Foundation (SNF PRIMA project, Grant PR00P1_208469).

Data availability statement. The CMIP6 data used in the case studies can be obtained from ESGF. The ibicus package used for climate model bias adjustment is available from https://github.com/ecmwf-projects/ibicus (Spuler and Wessel 2023).

APPENDIX

Storyline Approach and Physical Interpretation of the Climate Model Spread [Section 4b]

We follow the approach presented in Zappa and Shepherd (2017) to investigate the connection between remote drivers and model spread, before and after bias adjustment. The three remote drivers for each model (cm) analyzed are

- tropical warming $\Delta T^{\mathrm{cm}}_{\mathrm{tropical}}$: temperature change (2070–2100 SSP58.5 compared to 1960–90 hist) at 250 hPa averaged between 30°S and 30°N,
- polar warming $\Delta T^{\mathrm{cm}}_{\mathrm{polar}}$: temperature change (2070–2100 SSP58.5 compared to 1960–90 hist) at 850 hPa averaged between 60° and 90°N, and
- change in stratospheric polar vortex strength $\Delta U^{\mathrm{cm}}_{\mathrm{strat}}$: zonal wind change (2070–2100 SSP58.5 compared to 1960–90 hist) at 20 hPa averaged between 70° and 80°N.

We scale each of the remote drivers by the model-simulated global-mean temperature (GMT) change ΔT^{cm} and compute anomalies $(\cdot)'$ corresponding to the multimodel mean. Under a pattern-scaling assumption, we then regress the multimodel anomaly in precipitation change [change in extended winter, November–April (NDJFMA), gridcell average precipitation 2070–2100 compared to 1950–90], standardized by ΔT^{cm} onto driver anomalies:

$$\left(\frac{\Delta P^{\mathrm{cm}}_{\mathrm{bc}}}{\Delta T^{\mathrm{cm}}}\right)' = \alpha_{\mathrm{bc}} + \beta_{\mathrm{bc}}\left(\frac{\Delta T^{\mathrm{cm}}_{\mathrm{tropical}}}{\Delta T^{\mathrm{cm}}}\right)' + \gamma_{\mathrm{bc}}\left(\frac{\Delta T^{\mathrm{cm}}_{\mathrm{polar}}}{\Delta T^{\mathrm{cm}}}\right)' + \delta_{\mathrm{bc}}\left(\frac{\Delta U^{\mathrm{cm}}_{\mathrm{strat}}}{\Delta T^{\mathrm{cm}}}\right)' + \epsilon_{\mathrm{cm,bc}}.$$

Here, $\epsilon_{\mathrm{cm,bc}} \sim N(0, \sigma)$ is Gaussian noise. This approach enables us to capture (and possibly attribute) the influence of remote drivers on the multimodel spread in precipitation trends. In a second step, the large-scale drivers can then be clustered to define categories of future precipitation responses (or storylines, see Zappa and Shepherd 2017). The CMIP6 models used are listed in Table A1.

Regressions at each grid point are fitted to the raw model as well as all bias-corrected models (bc). To assess the differences between the regression coefficients of raw and bias-corrected precipitation, we also fit regressions to the differences:

$$\left(\frac{\Delta P^{\mathrm{cm}}_{\mathrm{bc}}}{\Delta T^{\mathrm{cm}}}\right)' - \left(\frac{\Delta P^{\mathrm{cm}}_{\mathrm{raw}}}{\Delta T^{\mathrm{cm}}}\right)' = \alpha_{\mathrm{bc}} + \beta_{\mathrm{bc}}\left(\frac{\Delta T^{\mathrm{cm}}_{\mathrm{tropical}}}{\Delta T^{\mathrm{cm}}}\right)' + \gamma_{\mathrm{bc}}\left(\frac{\Delta T^{\mathrm{cm}}_{\mathrm{polar}}}{\Delta T^{\mathrm{cm}}}\right)' + \delta_{\mathrm{bc}}\left(\frac{\Delta U^{\mathrm{cm}}_{\mathrm{strat}}}{\Delta T^{\mathrm{cm}}}\right)' + \epsilon_{\mathrm{cm,bc}}.$$

The sensitivity of the European precipitation response to the uncertainty in each of the three remote driver responses is shown in Fig. A1 for the raw climate model as well as the different bias-adjusted ones. While the response patterns identified in the raw, i.e., not bias adjusted, ensemble shows a similarity to the response patterns identified by Zappa and

Table A1. **CMIP6 models used for the analysis in section 4a and the storyline analysis in the appendix and section 4b.**

	Model	Analysis section 4a	Analysis section 4b
1	AS-RCEC__TaiESM1	*	*
2	BCC__BCC-CSM2-MR	*	
3	CAS__FGOALS-g3	*	*
4	CCCR-IITM__IITM-ESM	*	*
5	CCCma__CanESM5	*	*
6	CMCC__CMCC-CM2-SR5	*	*
7	CMCC__CMCC-ESM2	*	*
8	CNRM-CERFACS__CNRM-CM6-1	*	*
9	CNRM-CERFACS__CNRM-CM6-1-HR	*	*
10	CNRM-CERFACS__CNRM-ESM2-1	*	*
11	CSIRO-ARCCSS__ACCESS-CM2	*	
12	EC-Earth-Consortium__EC-Earth3-CC	*	*
13	EC-Earth-Consortium__EC-Earth3-Veg		*
14	EC-Earth-Consortium__EC-Earth3-Veg-LR	*	*
15	INM__INM-CM4-8	*	*
16	INM__INM-CM5-0	*	*
17	IPSL__IPSL-CM6A-LR	*	*
18	KIOST__KIOST-ESM		*
19	MIROC__MIROC-ES2L	*	*
20	MIROC__MIROC6	*	
21	MOHC__HadGEM3-GC31-LL		*
22	MOHC__HadGEM3-GC31-MM		*
23	MOHC__UKESM1-0-LL		*
24	MPI-M__MPI-ESM1-2-LR	*	*
25	MRI__MRI-ESM2-0	*	*
26	NCAR__CESM2	*	*
27	NCAR__CESM2-WACCM	*	*
28	NCC__NorESM2-LM		*
29	NCC__NorESM2-MM	*	*
30	NIMS-KMA__KACE-1-0-G		*
31	NOAA-GFDL__GFDL-CM4		*
32	NOAA-GFDL__GFDL-ESM4	*	*
33	NUIST__NESM3	*	*

Shepherd (2017), some differences emerge, the most prominent of which is the response to tropical warming in the western Mediterranean. These differences are likely due to the different models and model generation (CMIP5 vs CMIP6) investigated.

In addition, Fig. A2 showcases the difference between the bias-adjusted response patterns and the response pattern in the raw ensemble. Depending on the bias adjustment method, the influence of the remote drivers onto model spread can be quite substantially different. This appears to be nonhomogeneous across bias adjustment methods.

Table A1 provides an overview of climate models participating in CMIP6, which were used in the analysis in this paper. We did not apply any other criteria for the inclusion of a model other than their participation in CMIP and the availability of the required data through the Centre for Environmental Data Analysis (CEDA) archive.

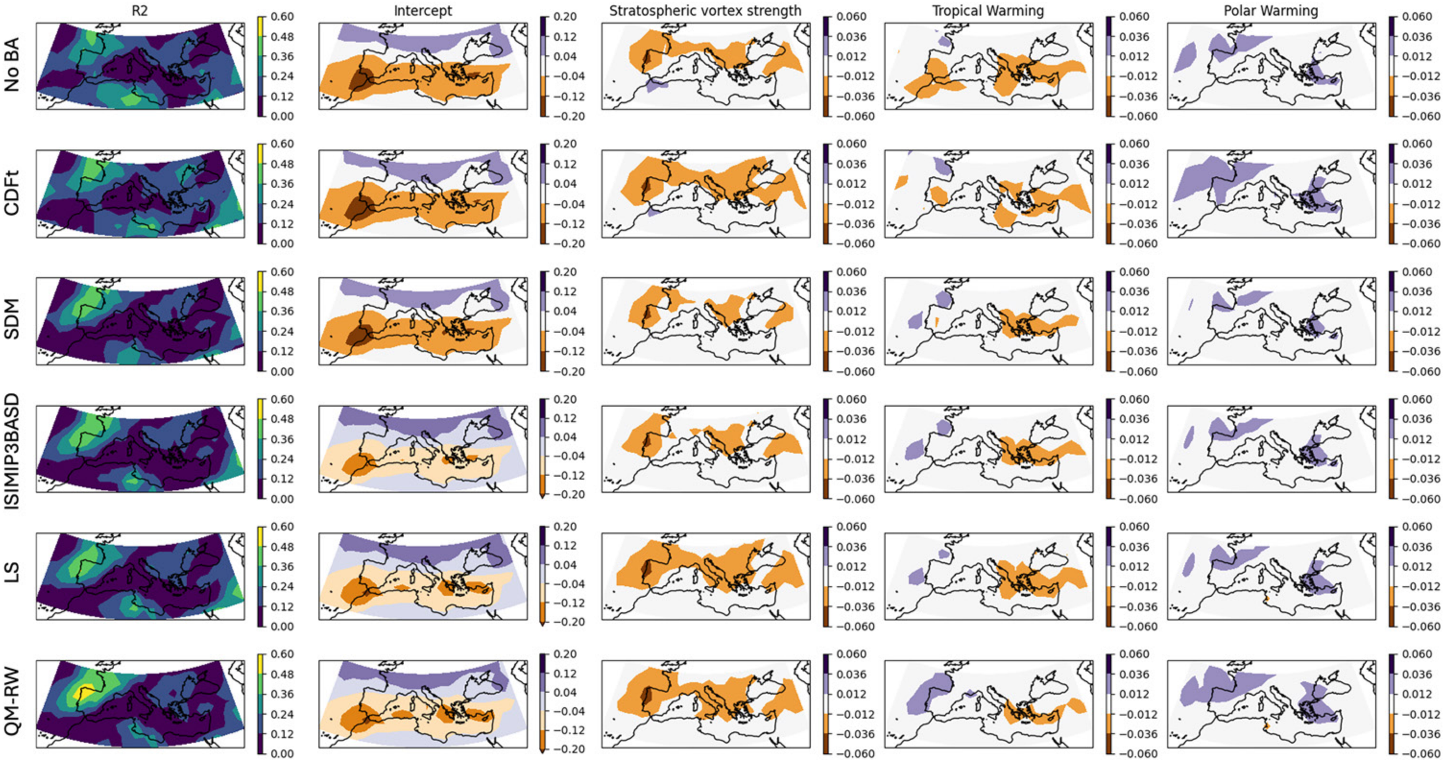

Fig. A1. Regression R2 as well as the influence of stratospheric vortex strength, tropical warming, and polar warming on the climate model ensemble spread for precipitation [mm (day K)$^{-1}$] for the raw climate model ensemble and for the bias-adjusted climate models.

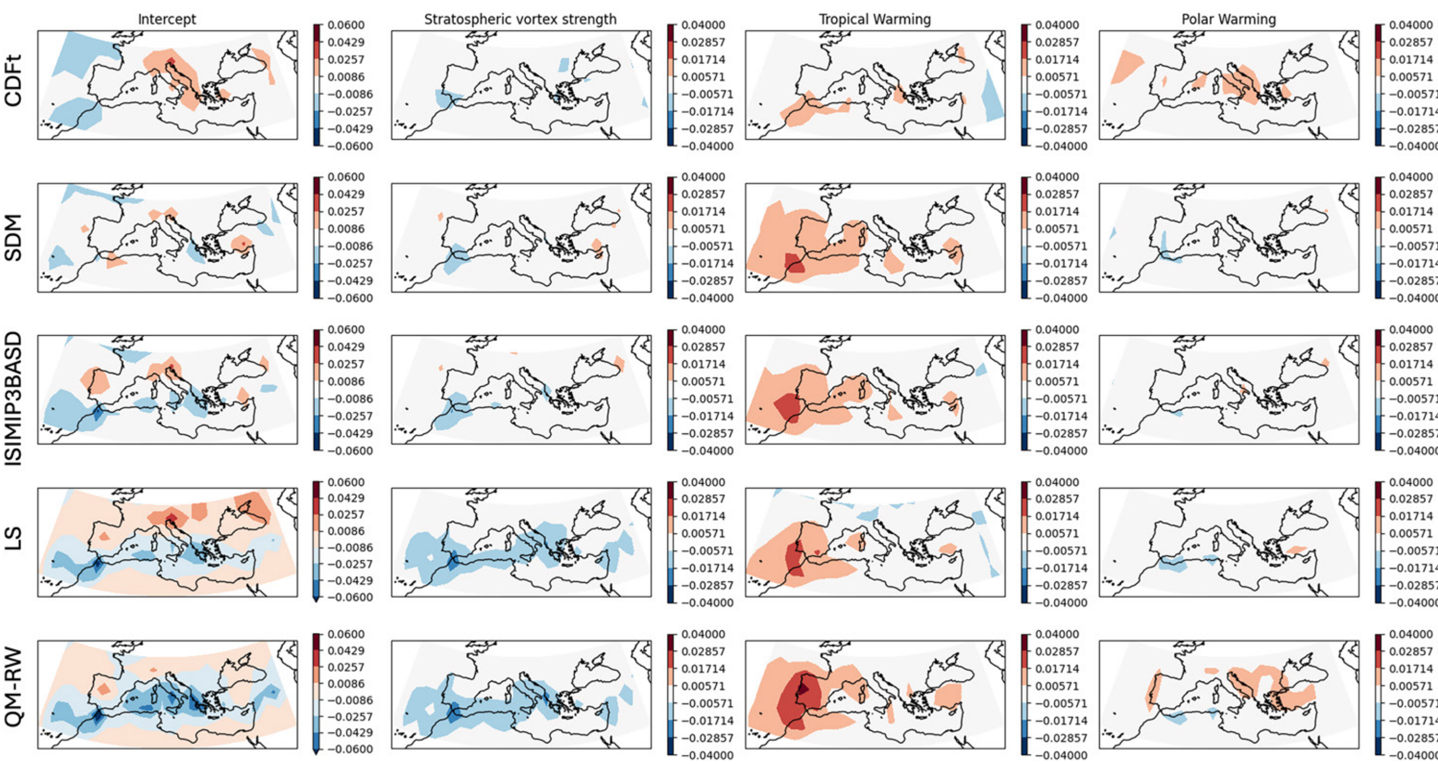

Fig. A2. Difference in response pattern between bias-adjusted and raw climate model for the influence of stratospheric vortex strength, tropical warming, and polar warming on the climate model ensemble spread for precipitation [mm (day K)$^{-1}$].

References

Addor, N., M. Rohrer, R. Furrer, and J. Seibert, 2016: Propagation of biases in climate models from the synoptic to the regional scale: Implications for bias adjustment. *J. Geophys. Res. Atmos.*, **121**, 2075–2089, https://doi.org/10.1002/2015JD024040.

Astagneau, P. C., R. R. Wood, M. Vrac, S. Kotlarski, P. Vaittinada Ayar, B. François, and M. I. Brunner, 2025: Impact of bias adjustment strategy on ensemble projections of hydrological extremes. EGUsphere, https://doi.org/10.5194/egusphere-2024-3966.

Baumberger, C., R. Knutti, and G. Hirsch Hadorn, 2017: Building confidence in climate model projections: An analysis of inferences from fit. *Wiley Interdiscip. Rev.: Climate Change*, **8**, e454, https://doi.org/10.1002/wcc.454.

Bennett, B., A. Devanand, S. Culley, S. Westra, D. Guo, and H. R. Maier, 2021: A modelling framework and R-package for evaluating system performance under hydroclimate variability and change. *Environ. Modell. Software*, **139**, 104999, https://doi.org/10.1016/j.envsoft.2021.104999.

Best, M. J., and Coauthors, 2011: The Joint UK Land Environment Simulator (JULES), model description—Part 1: Energy and water fluxes. *Geosci. Model Dev.*, **4**, 677–699, https://doi.org/10.5194/gmd-4-677-2011.

Bourgault, P., and Coauthors, 2023: xclim: xarray-based climate data analytics. *J. Open Source Software*, **8**, 5415, https://doi.org/10.21105/joss.05415.

Bremer, S., A. Wardekker, S. Dessai, S. Sobolowski, R. Slaattelid, and J. van der Sluijs, 2019: Toward a multi-faceted conception of co-production of climate services. *Climate Serv.*, **13**, 42–50, https://doi.org/10.1016/j.cliser.2019.01.003.

Broderick, C., C. Murphy, R. L. Wilby, T. Matthews, C. Prudhomme, and M. Adamson, 2019: Using a scenario-neutral framework to avoid potential maladaptation to future flood risk. *Water Resour. Res.*, **55**, 1079–1104, https://doi.org/10.1029/2018WR023623.

Cannon, A. J., S. R. Sobie, and T. Q. Murdock, 2015: Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *J. Climate*, **28**, 6938–6959, https://doi.org/10.1175/JCLI-D-14-00754.1.

Casanueva, A., S. Herrera, M. Iturbide, S. Lange, M. Jury, A. Dosio, D. Maraun, and J. M. Gutiérrez, 2020: Testing bias adjustment methods for regional climate change applications under observational uncertainty and resolution mismatch. *Atmos. Sci. Lett.*, **21**, e978, https://doi.org/10.1002/asl.978.

Cash, D., W. C. Clark, F. Alcock, N. M. Dickson, N. Eckley, and J. Jäger, 2002: Salience, credibility, legitimacy and boundaries: Linking research, assessment and decision making. John F. Kennedy School of Government, Harvard University, Faculty Research Working Papers Series, 25 pp., https://doi.org/10.2139/ssrn.372280.

——, ——, ——, ——, ——, D. H. Guston, J. Jäger, and R. B. Mitchell, 2003: Knowledge systems for sustainable development. *Proc. Natl. Acad. Sci. USA*, **100**, 8086–8091, https://doi.org/10.1073/pnas.1231332100.

Chae, S. T., and E.-S. Chung, 2024: Significant contribution of bias correction methods to uncertainty in future runoff projections under CMIP6 climate change. *J. Hydrol.*, **56**, 101973, https://doi.org/10.1016/j.ejrh.2024.101973.

Chandel, V. S., U. Bhatia, A. R. Ganguly, and S. Ghosh, 2024: State-of-the-art bias correction of climate models misrepresent climate science and misinform adaptation. *Environ. Res. Lett.*, **19**, 094052, https://doi.org/10.1088/1748-9326/ad6d82.

Chandler, R. E., 2013: Exploiting strength, discounting weakness: Combining information from multiple climate simulators. *Philos. Trans. A Math. Phys. Eng. Sci.*, **371**, 20120388, https://doi.org/10.1098/rsta.2012.0388.

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

——, ——, ——, and ——, 2013b: Performance and uncertainty evaluation of empirical downscaling methods in quantifying the climate change impacts on hydrology over two North American river basins. *J. Hydrol.*, **479**, 200–214, https://doi.org/10.1016/j.jhydrol.2012.11.062.

——, ——, and P. Lucas-Picher, 2015: Assessing the limits of bias-correcting climate model outputs for climate change impact studies. *J. Geophys. Res. Atmos.*, **120**, 1123–1136, https://doi.org/10.1002/2014JD022635.

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

Clark, D. B., and Coauthors, 2011: The Joint UK Land Environment Simulator (JULES), model description—Part 2: Carbon fluxes and vegetation dynamics. *Geosci. Model Dev.*, **4**, 701–722, https://doi.org/10.5194/gmd-4-701-2011.

CORDEX, 2022: Summary of bias adjustment methods applied to CORDEX simulations. Accessed 4 March 2024, https://is-enes-data.github.io/CORDEX_adjust_summary.html.

Dike, V. N., and Coauthors, 2018: Obstacles facing Africa's young climate scientists. *Nat. Climate Change*, **8**, 447–449, https://doi.org/10.1038/s41558-018-0178-x.

Dosio, A., 2016: Projections of climate change indices of temperature and precipitation from an ensemble of bias-adjusted high-resolution EURO-CORDEX regional climate models. *J. Geophys. Res. Atmos.*, **121**, 5488–5511, https://doi.org/10.1002/2015JD024411.

Dowling, D., 1999: Experimenting on theories. *Sci. Context*, **12**, 261–273, https://doi.org/10.1017/S0269889700003410.

Dumitrescu, A., V.-A. Amihaesei, and S. Cheval, 2020: RoCliB–bias-corrected CORDEX RCMdataset over Romania. *Geosci. Data J.*, **10**, 262–275, https://doi.org/10.1002/gdj3.161.

Eden, J. M., M. Widmann, D. Grawe, and S. Rast, 2012: Skill, correction, and downscaling of GCM-simulated precipitation. *J. Climate*, **25**, 3970–3984, https://doi.org/10.1175/JCLI-D-11-00254.1.

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

François, B., M. Vrac, A. J. Cannon, Y. Robin, and D. Allard, 2020: Multivariate bias corrections of climate simulations: Which benefits for which losses? *Earth Syst. Dyn.*, **11**, 537–562, https://doi.org/10.5194/esd-11-537-2020.

Frieler, K., and Coauthors, 2017: Assessing the impacts of 1.5°C global warming—Simulation protocol of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2b). *Geosci. Model Dev.*, **10**, 4321–4345, https://doi.org/10.5194/gmd-10-4321-2017.

Fung, F., 2018: UKCP18 guidance: Bias correction. Met Office, 7 pp., https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/ukcp/ukcp18-guidance-how-to-bias-correct.pdf.

Galmarini, S., and Coauthors, 2019: Adjusting climate model bias for agricultural impact assessment: How to cut the mustard. *Climate Serv.*, **13**, 65–69, https://doi.org/10.1016/j.cliser.2019.01.004.

Gergel, D. R., S. B. Malevich, K. E. McCusker, E. Tenezakis, M. T. Delgado, M. A. Fish, and R. E. Kopp, 2024: Global Downscaled Projections for Climate Impacts Research (GDPCIR): Preserving quantile trends for modeling future climate impacts. *Geosci. Model Dev.*, **17**, 191–227, https://doi.org/10.5194/gmd-17-191-2024.

Gobiet, A., M. Suklitsch, and G. Heinrich, 2015: The effect of empirical-statistical correction of intensity-dependent model errors on the temperature climate change signal. *Hydrol. Earth Syst. Sci.*, **19**, 4055–4066, https://doi.org/10.5194/hess-19-4055-2015.

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

Guo, D., S. Westra, and H. R. Maier, 2017: Use of a scenario-neutral approach to identify the key hydro-meteorological attributes that impact runoff from a natural catchment. *J. Hydrol.*, **554**, 317–330, https://doi.org/10.1016/j.jhydrol.2017.09.021.

Gutiérrez, J. M., and Coauthors, 2019: An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment. *Int. J. Climatol.*, **39**, 3750–3785, https://doi.org/10.1002/joc.5462.

Gutmann, E. D., J. J. Hamman, M. P. Clark, T. Eidhammer, A. W. Wood, and J. R. Arnold, 2022: En-GARD: A statistical downscaling framework to produce and test large ensembles of climate projections. *J. Hydrometeor.*, **23**, 1545–1561, https://doi.org/10.1175/JHM-D-21-0142.1.

Hall, A., P. Cox, C. Huntingford, and S. Klein, 2019: Progressing emergent constraints on future climate change. *Nat. Climate Change*, **9**, 269–278, https://doi.org/10.1038/s41558-019-0436-6.

Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.*, **90**, 1095–1108, https://doi.org/10.1175/2009BAMS2607.1.

Hess, P., S. Lange, C. Schötz, and N. Boers, 2023: Deep learning for bias-correcting CMIP6-class Earth system models. *Earth's Future*, **11**, e2023EF004002, https://doi.org/10.1029/2023EF004002.

Ho, C. K., D. B. Stephenson, M. Collins, C. A. T. Ferro, and S. J. Brown, 2012: Calibration strategies: A source of additional uncertainty in climate change projections. *Bull. Amer. Meteor. Soc.*, **93**, 21–26, https://doi.org/10.1175/2011BAMS3110.1.

Hui, Y., J. Chen, C.-Y. Xu, L. Xiong, and H. Chen, 2019: Bias nonstationarity of global climate model outputs: The role of internal climate variability and climate model sensitivity. *Int. J. Climatol.*, **39**, 2278–2294, https://doi.org/10.1002/joc.5950.

——, Y. Xu, J. Chen, C.-Y. Xu, and H. Chen, 2020: Impacts of bias nonstationarity of climate model outputs on hydrological simulations. *Hydrol. Res.*, **51**, 925–941, https://doi.org/10.2166/nh.2020.254.

Humphreys, P., 2009: The philosophical novelty of computer simulation methods. *Synthese*, **169**, 615–626, https://doi.org/10.1007/s11229-008-9435-2.

Iizumi, T., H. Takikawa, Y. Hirabayashi, N. Hanasaki, and M. Nishimori, 2017: Contributions of different bias-correction methods and reference meteorological forcing data sets to uncertainty in projected temperature and precipitation extremes. *J. Geophys. Res. Atmos.*, **122**, 7800–7819, https://doi.org/10.1002/2017JD026613.

Iturbide, M., and Coauthors, 2019: The R-based climate4R openframework for reproducible climate data access and post-processing. *Environ. Modell. Software*, **111**, 42–54, https://doi.org/10.1016/j.envsoft.2018.09.009.

Jägermeyr, J., and Coauthors, 2021: Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. *Nat. Food*, **2**, 873–885, https://doi.org/10.1038/s43016-021-00400-y.

Jain, S., A. A. Scaife, T. G. Shepherd, C. Deser, N. Dunstone, G. A. Schmidt, K. E. Trenberth, and T. Turkington, 2023: Importance of internal variability for climate model assessment. *npj Climate Atmos. Sci.*, **6**, 68, https://doi.org/10.1038/s41612-023-00389-0.

Jebeile, J., 2024: From regional climate models to usable information. *Climate Change*, **177**, 53, https://doi.org/10.1007/s10584-024-03693-7.

——, and J. Roussos, 2023: Usability of climate information: Toward a new scientific framework. *Wiley Interdiscip. Rev.: Climate Change*, **14**, e833, https://doi.org/10.1002/wcc.833.

Kahlenborn, W., and Coauthors, 2021: *Klimawirkungs- und Risikoanalyse für Deutschland 2021 (Kurzfassung).* Umweltbundesamt, 117 pp., https://www.umweltbundesamt.de/en/publikationen/KWRA-Zusammenfassung.

Kelly, P. M., and W. N. Adger, 2000: Theory and practice in assessing vulnerability to climate change and facilitating adaptation. *Climatic Change*, **47**, 325–352, https://doi.org/10.1023/A:1005627828199.

Kirchhoff, C. J., M. C. Lemos, and S. Dessai, 2013: Actionable knowledge for environmental decision making: Broadening the usability of climate science. *Annu. Rev. Environ. Resour.*, **38**, 393–414, https://doi.org/10.1146/annurev-environ-022112-112828.

Knutti, R., 2008: Should we believe model predictions of future climate change? *Philos. Trans. A Math. Phys. Eng. Sci.*, **366**, 4647–4664, https://doi.org/10.1098/rsta.2008.0169.

——, J. Sedláček, B. M. Sanderson, R. Lorenz, E. M. Fischer, and V. Eyring, 2017: A climate model projection weighting scheme accounting for performance and interdependence. *Geophys. Res. Lett.*, **44**, 1909–1918, https://doi.org/10.1002/2016GL072012.

Lafferty, D. C., and R. L. Sriver, 2023: Downscaling and bias-correction contribute considerable uncertainty to local climate projections in CMIP6. *npj Climate Atmos. Sci.*, **6**, 158, https://doi.org/10.1038/s41612-023-00486-0.

Lafon, T., S. Dadson, G. Buys, and C. Prudhomme, 2013: Bias correction of daily precipitation simulated by a regional climate model: A comparison of methods. *Int. J. Climatol.*, **33**, 1367–1381, https://doi.org/10.1002/joc.3518.

Lange, S., 2019: Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0). *Geosci. Model Dev.*, **12**, 3055–3070, https://doi.org/10.5194/gmd-12-3055-2019.

——, 2021: ISIMIP3b bias adjustment fact sheet. Inter-Sectoral Impact Model Intercomparison Project, 40 pp.

Laux, P., and Coauthors, 2021: To bias correct or not to bias correct? An agricultural impact modelers' perspective on regional climate model data. *Agric. For. Meteor.*, **304–305**, 108406, https://doi.org/10.1016/j.agrformet.2021.108406.

Lavoie, J., P. Bourgault, T. J. Smith, T. Logan, M. Leduc, L.-P. Caron, S. Gammon, and M. Braun, 2024: An ensemble of bias-adjusted CMIP6 climate simulations based on a high-resolution North American reanalysis. *Sci. Data*, **11**, 64, https://doi.org/10.1038/s41597-023-02855-z.

Lehner, F., I. Nadeem, and H. Formayer, 2023: Evaluating skills and issues of quantile-based bias adjustment for climate change scenarios. *Adv. Stat. Climate Meteor. Oceanogr.*, **9**, 29–44, https://doi.org/10.5194/ascmo-9-29-2023.

Lemos, M. C., C. J. Kirchhoff, and V. Ramprasad, 2012: Narrowing the climate information usability gap. *Nat. Climate Change*, **2**, 789–794, https://doi.org/10.1038/nclimate1614.

Li, H., J. Sheffield, and E. F. Wood, 2010: Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *J. Geophys. Res.*, **115**, D10101, https://doi.org/10.1029/2009JD012882.

Liess, S., T. E. Twine, P. K. Snyder, W. D. Hutchison, G. Konar-Steenberg, B. L. Keeler, and K. A. Brauman, 2022: High-resolution climate projections over Minnesota for the 21st century. *Earth Space Sci.*, **9**, e2021EA001893, https://doi.org/10.1029/2021EA001893.

Manzanas, R., and J. M. Gutiérrez, 2019: Process-conditioned bias correction for seasonal forecasting: A case-study with ENSO in Peru. *Climate Dyn.*, **52**, 1673–1683, https://doi.org/10.1007/s00382-018-4226-z.

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

——, 2013: Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *J. Climate*, **26**, 2137–2143, https://doi.org/10.1175/JCLI-D-12-00821.1.

——, 2016a: Bias correcting climate change simulations—A critical review. *Curr. Climate Change Rep.*, **2**, 211–220, https://doi.org/10.1007/s40641-016-0050-x.

——, 2016b: Reply to "Comment on 'Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue'." *J. Climate*, **29**, 8669–8671, https://doi.org/10.1175/JCLI-D-16-0592.1.

——, and M. Widmann, 2018a: Cross-validation of bias-corrected climate simulations is misleading. *Hydrol. Earth Syst. Sci.*, **22**, 4867–4873, https://doi.org/10.5194/hess-22-4867-2018.

——, and ——, 2018b: Weather generators. *Statistical Downscaling and Bias Correction for Climate Research.* Cambridge University Press, 201–219, https://doi.org/10.1017/9781107588783.

——, and ——, Eds., 2018c: User needs. *Statistical Downscaling and Bias Correction for Climate Research*, Cambridge University Press, 33–40, https://doi.org/10.1017/9781107588783.006.

——, and Coauthors, 2010: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.*, **48**, RG3003, https://doi.org/10.1029/2009RG000314.

——, and Coauthors, 2015: VALUE: A framework to validate downscaling approaches for climate change studies. *Earth's Future*, **3** (1), 1–14, https://doi.org/10.1002/2014EF000259.

——, and Coauthors, 2017: Towards process-informed bias correction of climate change simulations. *Nat. Climate Change*, **7**, 764–773, https://doi.org/10.1038/nclimate3418.

——, M. Widmann, and J. M. Gutiérrez, 2019: Statistical downscaling skill under present climate conditions: A synthesis of the VALUE perfect predictor experiment. *Int. J. Climatol.*, **39**, 3692–3703, https://doi.org/10.1002/joc.5877.

——, H. Truhetz, and A. Schaffer, 2021: Regional climate model biases, their dependence on synoptic circulation biases and the potential for bias adjustment: A process-oriented evaluation of the Austrian regional climate projections. *J. Geophys. Res. Atmos.*, **126**, e2020JD032824, https://doi.org/10.1029/2020JD032824.

Michelangeli, P.-A., M. Vrac, and H. Loukos, 2009: Probabilistic downscaling approaches: Application to wind cumulative distribution functions. *Geophys. Res. Lett.*, **36**, L11708, https://doi.org/10.1029/2009GL038401.

Mishra, V., U. Bhatia, and A. D. Tiwari, 2020: Bias-corrected climate projections for South Asia from Coupled Model Intercomparison Project-6. *Sci. Data*, **7**, 338, https://doi.org/10.1038/s41597-020-00681-1.

Nahar, J., F. Johnson, and A. Sharma, 2017: Assessing the extent of non-stationary biases in GCMs. *J. Hydrol.*, **549**, 148–162, https://doi.org/10.1016/j.jhydrol.2017.03.045.

Navarro-Racines, C., J. Tarapues, P. Thornton, A. Jarvis, and J. Ramirez-Villegas, 2020: High-resolution and bias-corrected CMIP5 projections for climate change impact assessments. *Sci. Data*, **7**, 7. https://doi.org/10.1038/s41597-019-0343-8.

NGFS, 2021: NGFS Climate Scenarios for central banks and supervisors. Network for Greening the Financial System, 48 pp., https://www.ngfs.net/en/ngfs-climate-scenarios-central-banks-and-supervisors-september-2022.

O'Brien, K., S. Eriksen, L. P. Nygaard, and A. Schjolden, 2007: Why different interpretations of vulnerability matter in climate change discourses. *Climate Policy*, **7**, 73–88, https://doi.org/10.1080/14693062.2007.9685639.

Olschewski, P., P. Laux, J. Wei, B. Böker, Z. Tian, L. Sun, and H. Kunstmann, 2023: An ensemble-based assessment of bias adjustment performance, changes in hydrometeorological predictors and compound extreme events in EAS-CORDEX. *Wea. Climate Extremes*, **39**, 100531, https://doi.org/10.1016/j.wace.2022.100531.

Oreskes, N., K. Shrader-Frechette, and K. Belitz, 1994: Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, **263**, 641–646, https://doi.org/10.1126/science.263.5147.641.

Padulano, R., L. A. Gomez-Mogollon, L. Napolitano, and G. Rianna, 2025: Quantile-based bias-correction of extreme rainfall: Pros & cons of popular methods for climate signal preservation. *J. Hydrol.*, **653**, 132814, https://doi.org/10.1016/j.jhydrol.2025.132814.

Philip, S., and Coauthors, 2020: A protocol for probabilistic extreme event attribution analyses. *Adv. Stat. Climate Meteor. Oceanogr.*, **6**, 177–203, https://doi.org/10.5194/ascmo-6-177-2020.

Priestley, M. D. K., D. Ackerley, J. L. Catto, and K. I. Hodges, 2023: Drivers of biases in the CMIP6 extratropical storm tracks. Part I: Northern Hemisphere. *J. Climate*, **36**, 1451–1467, https://doi.org/10.1175/JCLI-D-20-0976.1.

Prudhomme, C., R. L. Wilby, S. Crooks, A. L. Kay, and N. S. Reynard, 2010: Scenario-neutral approach to climate change impact studies: Application to flood risk. *J. Hydrol.*, **390**, 198–209, https://doi.org/10.1016/j.jhydrol.2010.06.043.

Pulkkinen, K., and Coauthors, 2022: The value of values in climate science. *Nat. Climate Change*, **12**, 4–6, https://doi.org/10.1038/s41558-021-01238-9.

Rabin, S. S., and Coauthors, 2017: The Fire Modeling Intercomparison Project (FireMIP), phase 1: Experimental and analytical protocols with detailed model descriptions. *Geosci. Model Dev.*, **10**, 1175–1197, https://doi.org/10.5194/gmd-10-1175-2017.

Rodrigues, R. R., 2021: Downside up: Science matters equally to the Global South. *Commun. Earth Environ.*, **2**, 100, https://doi.org/10.1038/s43247-021-00171-1.

——, and T. G. Shepherd, 2022: Small is beautiful: Climate-change science as if people mattered. *PNAS Nexus*, **1**, pgac009, https://doi.org/10.1093/pnasnexus/pgac009.

Rosenzweig, C., and Coauthors, 2013: The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. *Agric. For. Meteor.*, **170**, 166–182, https://doi.org/10.1016/j.agrformet.2012.09.011.

Sansom, P. G., D. B. Stephenson, and T. J. Bracegirdle, 2021: On constraining projections of future climate using observations and simulations from multiple climate models. *J. Amer. Stat. Assoc.*, **116**, 546–557, https://doi.org/10.1080/01621459.2020.1851696.

Scarrott, C., and A. MacDonald, 2012: Review of extreme value threshold estimation and uncertainty quantification. *REVSTAT Stat. J.*, 33–60, https://doi.org/10.57805/REVSTAT.V10I1.110.

Schipper, E. L. F., and Coauthors, 2021: Equity in climate scholarship: A manifesto for action. *Climate Dev.*, **13**, 853–856, https://doi.org/10.1080/17565529.2021.1923308.

Senatore, A., D. Fuoco, M. Maiolo, G. Mendicino, G. Smiatek, and H. Kunstmann, 2022: Evaluating the uncertainty of climate model structure and bias correction on the hydrological impact of projected climate change in a Mediterranean catchment. *J. Hydrol.*, **42**, 101120, https://doi.org/10.1016/j.ejrh.2022.101120.

Shepherd, T. G., and Coauthors, 2018: Storylines: An alternative approach to representing uncertainty in physical aspects of climate change. *Climate Change*, **151**, 555–571, https://doi.org/10.1007/s10584-018-2317-9.

Sippel, S., and Coauthors, 2016: A novel bias correction methodology for climate impact simulations. *Earth Syst. Dyn.*, **7**, 71–88, https://doi.org/10.5194/esd-7-71-2016.

Skelton, M., J. J. Porter, S. Dessai, D. N. Bresch, and R. Knutti, 2017: The social and scientific values that shape national climate scenarios: A comparison of the Netherlands, Switzerland and the UK. *Reg. Environ. Change*, **17**, 2325–2338, https://doi.org/10.1007/s10113-017-1155-z.

Spuler, F., and J. Wessel, 2023: Ibicus v1.0.1. Zenodo, https://doi.org/10.5281/zenodo.8101898.

Spuler, F. R., J. B. Wessel, E. Comyn-Platt, J. Varndell, and C. Cagnazzo, 2024: ibicus: A new open-source Python package and comprehensive interface for statistical bias adjustment and evaluation in climate modelling (v1.0.1). *Geosci. Model Dev.*, **17**, 1249–1269, https://doi.org/10.5194/gmd-17-1249-2024.

Tandon, A., 2021: Analysis: The lack of diversity in climate-science research. Carbon Brief, https://www.carbonbrief.org/analysis-the-lack-of-diversity-in-climate-science-research/.

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

Thrasher, B., W. Wang, A. Michaelis, F. Melton, T. Lee, and R. Nemani, 2022: NASA global daily downscaled projections, CMIP6. *Sci. Data*, **9**, 262, https://doi.org/10.1038/s41597-022-01393-4.

Tootoonchi, F., A. Todorović, T. Grabs, and C. Teutschbein, 2023: Uni- and multivariate bias adjustment of climate model simulations in Nordic catchments: Effects on hydrological signatures relevant for water resources management in a changing climate. *J. Hydrol.*, **623**, 129807, https://doi.org/10.1016/j.jhydrol.2023.129807.

Tradowsky, J. S., and Coauthors, 2023: Attribution of the heavy rainfall events leading to severe flooding in Western Europe during July 2021. *Climatic Change*, **176**, 90, https://doi.org/10.1007/s10584-023-03502-7.

Vandal, T., E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, 2017: DeepSD: Generating high resolution climate change projections through single image super-resolution. arXiv, 1703.03126v1, https://doi.org/10.48550/arXiv.1703.03126.

Van de Velde, J., M. Demuzere, B. De Baets, and N. E. C. Verhoest, 2022: Impact of bias nonstationarity on the performance of uni- and multivariate bias-adjusting methods: A case study on data from Uccle, Belgium. *Hydrol. Earth Syst. Sci.*, **26**, 2319–2344, https://doi.org/10.5194/hess-26-2319-2022.

Velázquez, J. A., M. Troin, D. Caya, and F. Brissette, 2015: Evaluating the time-invariance hypothesis of climate model bias correction: Implications for hydrological impact studies. *J. Hydrometeor.*, **16**, 2013–2026, https://doi.org/10.1175/JHM-D-14-0159.1.

Verfaillie, D., M. Déqué, S. Morin, and M. Lafaysse, 2017: The method ADAMONT v1.0 for statistical adjustment of climate projections applicable to energy balance land surface models. *Geosci. Model Dev.*, **10**, 4257–4283, https://doi.org/10.5194/gmd-10-4257-2017.

Volosciuk, C., D. Maraun, M. Vrac, and M. Widmann, 2017: A combined statistical bias correction and stochastic downscaling method for precipitation. *Hydrol. Earth Syst. Sci.*, **21**, 1693–1719, https://doi.org/10.5194/hess-21-1693-2017.

Vrac, M., T. Noël, and R. Vautard, 2016: Bias correction of precipitation through singularity stochastic removal: Because occurrences matter. *J. Geophys. Res. Atmos.*, **121**, 5237–5258, https://doi.org/10.1002/2015JD024511.

Wang, H.-M., J. Chen, C.-Y. Xu, J. Zhang, and H. Chen, 2020: A framework to quantify the uncertainty contribution of GCMs over multiple sources in hydrological impacts of climate change. *Earth's Future*, **8**, e2020EF001602, https://doi.org/10.1029/2020EF001602.

Wang, Y., G. Sivandran, and J. M. Bielicki, 2018: The stationarity of two statistical downscaling methods for precipitation under different choices of cross-validation periods. *Int. J. Climatol.*, **38**, e330–e348, https://doi.org/10.1002/joc.5375.

Wilby, R. L., and S. Dessai, 2010: Robust adaptation to climate change. *Weather*, **65**, 180–185, https://doi.org/10.1002/wea.543.

Williamson, D. B., and P. G. Sansom, 2019: How are emergent constraints quantifying uncertainty and what do they leave behind? *Bull. Amer. Meteor. Soc.*, **100**, 2571–2588, https://doi.org/10.1175/BAMS-D-19-0131.1.

Wootten, A., A. Terando, B. J. Reich, R. P. Boyles, and F. Semazzi, 2017: Characterizing sources of uncertainty from global climate models and downscaling techniques. *J. Appl. Meteor. Climatol.*, **56**, 3245–3262, https://doi.org/10.1175/JAMC-D-17-0087.1.

Xu, Z., Y. Han, C.-Y. Tam, Z.-L. Yang, and C. Fu, 2021: Bias-corrected CMIP6 global dataset for dynamical downscaling of the historical and future climate (1979–2100). *Sci. Data*, **8**, 293, https://doi.org/10.1038/s41597-021-01079-3.

Zappa, G., and T. G. Shepherd, 2017: Storylines of atmospheric circulation change for European regional climate impact assessment. *J. Climate*, **30**, 6561–6577, https://doi.org/10.1175/JCLI-D-16-0807.1.

Zscheischler, J., E. M. Fischer, and S. Lange, 2019: The effect of univariate bias adjustment on multivariate hazard estimates. *Earth Syst. Dyn.*, **10**, 31–43, https://doi.org/10.5194/esd-10-31-2019.

——, and Coauthors, 2020: A typology of compound weather and climate events. *Nat. Rev. Earth Environ.*, **1**, 333–347, https://doi.org/10.1038/s43017-020-0060-z.