



Max-Planck-Institut
für Biogeochemie



02.01.2016

To

The Editors of *Earth System Dynamics*

Postfach 10 01 64
07701 Jena
Deutschland
Tel. +49-(0)3641-6233
Fax +49-(0)3641-57
E-Mail: ssippel@bgc-jena.mpg.de
Web: www.bgc-jena.mpg.de
Direktorium
Markus Reichstein
Susann Trumbore
Martin Heimann
Ust.-Nr. 143/844/11208
ID-Nr. DE 129517720

Revised manuscript submission “A novel bias correction methodology for climate impact simulations”

Dear Editors:

Please find enclosed the carefully revised version of our latest manuscript entitled „A novel bias correction methodology for climate impact simulations“.

In the paper, we have introduced a novel bias correction methodology that preserves the multivariate structure and physical consistency of regional climate model simulations. The methodology compares favourably to a standard bias correction methodology that has been used widely for climate impact assessments.

Detailed responses to the reviewers’ comments were submitted in the interactive discussion forum of *Earth System Dynamics Discussions*, and the manuscript has been changed accordingly.

Given the importance of an accurate quantification of changing climate extremes and their associated impacts, we believe that our findings are of high relevance both to the climate extremes community as well as to the impacts modelling community. Hence, our study takes an interdisciplinary perspective in linking those two communities and topics and we would appreciate if our study would be considered for publication in the journal *Earth System Dynamics*.

Sincerely yours

Sebastian Sippel

Sebastian Sippel on behalf of all authors

Response to S. Ghosh:

S. Sippel et al., 2015

This is a very interesting work on bias correction of climate simulations preserving the physics. This is worth publishing.

We thank the reviewer for the encouraging feedback.

I have few minor points:

1. Is it possible to show that the physics is preserved, through the calculation of closure term of water and energy cycle. They may follow Trenberth and Fasullo (2013), GRL to compute the closure term. Trenberth, K. E., and J. T. Fasullo (2013), North American water and energy cycles, *Geophysical Research Letters*, 40(2), 365-369, doi: 10.1002/grl.50107.

Yes, this is indeed possible. However, our paper deals with a bias-correction that preserves the physics of the *regional climate model output*. Thus, our approach relies on the assumption that the output of the climate model is physically consistent, i.e. that water and energy balances are closed. We believe that evaluating the water and energy balance in the HadRM3P model is somewhat outside of the scope of our paper.

To this end, Massey et al (2014) showed that the HadRM3P model, including the new parametrization that was introduced in the latter paper, produces a realistic energy balance; and that the closure and realistic representation of the global radiation balance induces an important constraint on the system. In HadRM3P, the water balance is closed. (as much as this is possible in SST driven simulations)

In order to clarify this in the manuscript, we have additionally highlighted that “physical consistent” bias correction requires that water and energy balances are closed and realistic in the Discussion of the manuscript.

2. The authors must mention that bias correction is different from downscaling. There is a recent trend of mixing both of them just by adding a disaggregation to bias correction. This can be a good word of caution.

We agree with the reviewer that this distinction is crucial. In the revised version of the manuscript, this important point is included in the Discussion Section: Here, we emphasize that our bias correction is designed to alleviate biases in a physically consistent way, but is not designed for downscaling or to match any scale mismatches.

3. Is the bias stationary to apply for future climate projections?

An important implicit (and unavoidable) assumption for bias correction of future (or “counterfactual past”) simulations is that the structure of the bias remains constant over time. Maraun (2012) had shown that this might be the case on large scales, but important non-stationarities exist locally and regionally. Therefore, in the Discussion Section we (generally) argue for caution in applying any type of bias correction to future simulations.

However, if one assumes stationarity of the bias structure through time, the method can be readily applied to future (or “counterfactual past”) simulations (as the mapping function for the resampling step is defined between percentiles, not absolute quantities).

Lastly, to address the reviewer’s comment, we have tested that the bias structure in the “bias correction period” in our paper (1985-2010) is stationary, by comparing the bias

structure in the first (1985-1997) and the second period (1998-2010). See the figure below for the results: Although of course this approach is limited by the small sample size (13 years for each period), the bias structure is very similar for both periods and both for the entire temperature signal and detrended temperatures (i.e. the biases (for example, the probability for an event in the hot tail) are not due to single “strange” years but rather occur systematically in every year in the ensemble).

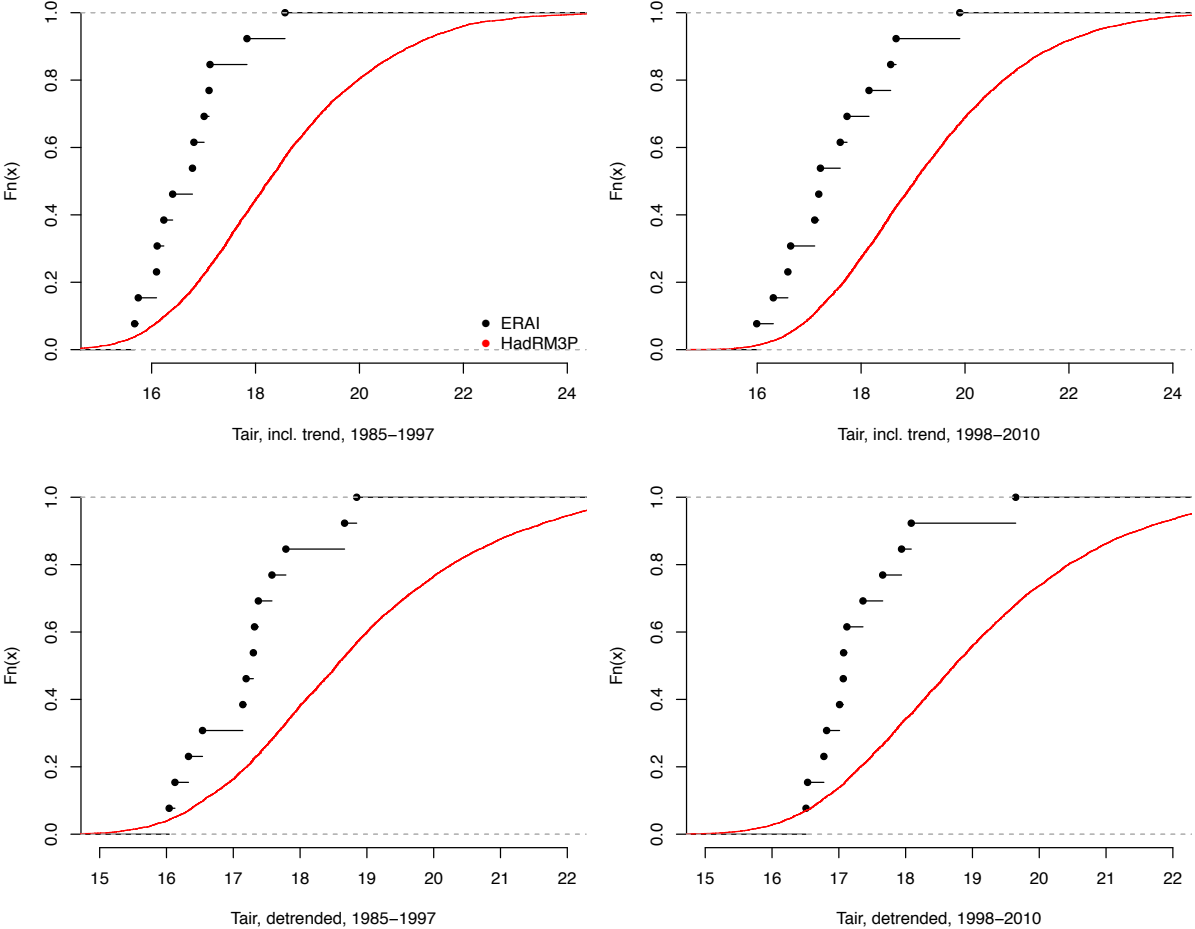


Figure 1: Stationarity of biases in two periods (1985-1997, and 1998-2010), for both the full temperature signal and detrended temperatures.

Response to A. Mondal:

S. Sippel et al., 2015

The paper presents a novel bias correction technique for use in climate change impact assessment studies. The technique is claimed to preserve the physics as well as multivariate dependence structures. Benefits of the proposed technique in comparison to the existing methods are categorically brought out and an end-to-end application is also illustrated using an impact-assessment study. The paper is overall very well written and will be of interest to a wide range of researchers. Therefore, I would favor its publication.

We thank the reviewer for the encouraging feedback.

However, I have a few comments/suggestions as I detail below and would like to see the authors' responses to them. Since the proposed bias correction methods leads to a decrease in effective ensemble size, large ensembles such as the weather@home experiment is necessary for its application. In my opinion, this is a strong limitation of the method as such large experiments are rare, particularly for developing regions. Is the proposed technique also effective on GCM simulations directly?

We agree with the reviewer, this issue is an important consideration for the applicability of our method.

The methodology's applicability depends on a few factors (see also discussion in the paper):

1. (ensemble) sample size and the anticipated application
2. structure of the bias in the model (if the model produces unrealistic simulations in all ensemble members, then our bias correction cannot improve the simulations).

Regarding (1): As demonstrated in the paper, large ensembles are quite useful (but are increasingly available also for many parts of the world, see e.g.

<http://www.climateprediction.net/weatherathome>).

However, also smaller of 5-10 member ensembles could be used for this type of bias correction, if one assumes ergodicity over a certain number of years. E.g. assume an CORDEX-RCM simulation with 10 members over each 50 years. In this case, one could still resample from 500 ensemble members, hence a useful sample size should remain. In this context, the type of application also plays a role: A so-derived ensemble could be well interpreted probabilistically, but of course "strict year-to-year continuity" would be lost.

Regarding (2): In addition to (1), of course the "magnitude" of the biases is important. Consider e.g. Fig. 2d in the revised manuscript: No loss in effective sample size would require that for each decile in the observations, 10% of the original ensemble members are available for resampling. In our case, for the "worst" deciles the effective sample size is reduced roughly by 50% (Fig. 2d). This way, one could estimate whether the effective sample size after resampling for a smaller ensemble is large enough for the anticipated application (which of course is a function of the "severity" of the bias and the number of ensemble members).

We have stressed both points more clearly in the discussion section.

I also have concerns with the quantile mapping based technique for more general applications of the proposed bias correction method. The retention of an ensemble member depends on q_{mod} as given by the transfer function. Therefore, if a model simulated value does not correspond to a quantile of the observed record, that value is rejected, thereby indirectly defining a prescribed range of possible values of the variable based on certain number of years of observations. For bias correction of future values, clearly, there is no way to ensure that the actual values belong to that range.

We agree with the reviewer in that the choice of the resampling constraint is a critical step for the applicability of the methodology. However, we would like to add two comments:

- 1) Present-day bias correction: For transient bias correction, the underlying assumption is that the actual range of variability in the observations is representative, as correctly pointed out by the reviewer (but not fully restricted to that as in quantile-quantile mapping or similar approaches, due to the kernel bandwidth that allows some flexibility). However, for this reason we argue in the paper that an “aggregated” constraint (i.e., aggregated both in time (JJA) and in space (Central Europe)) is more useful than constraints define on short periods or single grid cells, because the resampling would then be very sensitive to the observed variability. We have tried to discuss this point more clearly in the manuscript.
- 2) Bias-correction of future simulations: For future simulations, clearly a direct application of the bias correction based on resampling a range of absolute values is not meaningful (as correctly stated by the reviewer). However, if one assumes that the structure of the biases relative to the observations holds in the future (i.e. not in absolute values, but in the percentiles of model simulations relative to percentiles in the observations, see e.g. Fig. 2b in the revised manuscript), the bias correction is still applicable, as the transfer function is defined as a mapping between the percentiles and calibrated on the present (e.g. Fig. 2b, see also our response to Referee #1). Of course this is a rigorous assumption, and we are not arguing that this should be assumed, but this kind of “stationarity assumption of the bias structure” underlies implicitly all bias correction approaches for future simulations.

Further, selection of Gaussian kernels seem somewhat arbitrary. It is a subjective choice, and so is the choice of Cubic Hermite splines.

This is of course correct.

The choice of the Gaussian kernel is motivated by the choice of the resampling metric and the Central limit theorem: Since the constraint is quite highly aggregated (JJA means over a relatively large region, Central Europe), we believe that the choice for a Gaussian kernel seems somewhat “natural”.

Cubic Hermite splines for interpolation are clearly an arbitrary choice, but the form of the transfer function is almost entirely determined by the two kernels over observations and the model ensemble (because both kernels allow resampling of an arbitrarily large number of random variables).

Additionally, in my opinion, more clarity is solicited in the description of the proposed bias correction methodology.

We thank the reviewer for highlighting the need for a “cleaner” methodological description.

We have redrawn the figures for the methodological illustration (Fig. 2b and 2c). We hope that these now better reflect the procedure how the transfer function is obtained as a mapping between percentiles in the observations and model ensemble?

Additional changes to the methodological description are highlighted below as a response to the reviewer’s comments.

For example, do the authors simply concatenate observed data listed in Table 1? How do they fit the kernel density ‘over the observed meteorological constraint in various observational datasets’ (blue cdf in Figure 2(a)? How are the 800 ensemble members merged to obtain the red cdf of Figure 2(a)? The authors also mention that they derive a bias-corrected sample by ‘randomly resampling n times from f_{obs} ’: what is the length of the sample?

Further, q_{mod_X} and q_{obs_X} represent a given quantile in the model ensemble and observation, respectively. Does this then imply that bias correction is carried out individually for each quantile?

One dataset is used at a time. Concatenation or another form of combination of different observational datasets would be an option, but could result in somewhat “strange” distributions (e.g. if one dataset is simply offset relative to another one would yield a bimodal distribution). Therefore, figures 4 and 5 contain several lines in the return time plots, one for the correction with each observational dataset separately. Since the different observational datasets were very similar to each other for the aggregated temperature constraint (Fig. 4a, b), only the ERA-Interim constraint was used for the LPJmL simulations. This has been clarified in the manuscript.

The Gaussian kernel density fit uses a bandwidth estimation procedure following Sheather and Jones (1991). Subsequently, resampling is done by 1) sampling n times with replacement from the respective observations (e.g. x_i), and 2) sampling from a Gaussian distribution with mean x_i and the bandwidth h as the standard deviation (definition of the Gaussian kernel). The Gaussian kernel fit is identically applied to the observations and the model ensemble (i.e. the temperature constraint is obtained from each ensemble member and each year and the Gaussian kernel is fitted over all of them). The length of the sample is $n=800$ for the illustrative application and probabilistic interpretation in the manuscript.

Bias correction is done by resampling percentiles from the observations and retaining the ensemble member that corresponds as given by the transfer function (in that sense, bias correction is applied individually to each resampled (random) percentile).

We have clarified these issues in the Methods section of the manuscript.

For fitting the GEV distribution, though the length of all the observed records listed in Table 1 is greater, the authors mention about a ‘relatively small sample size (1901-2014)’. I did not understand why (why not all 26 years?). Also, statistical extreme value theory requires certain conditions to be held true for application of the GEV distribution to the block maxima. If 10-year samples are ‘randomly concatenated’, the tail behaviour may change, thereby questioning the application of extreme value theory to the concatenated datasets. Another, more fundamental issue concerns the random nature of the model output. The bias corrected variables are after all output of models

that are deterministic in nature; therefore, whether they can be considered as random variables remains a question.

We agree with the reviewer's concerns regarding concatenation of observational datasets. This might change the tail behaviour and could lead to many other problems. Therefore we have clarified in the revised manuscript that we do not concatenate observational datasets, but perform the analysis with each observational dataset separately. Hence, 10-year samples are concatenated only from the same observational dataset (following the procedure outlined in the manuscript), thus assuming ergodicity. Regarding 10-year samples from the model ensemble, we would like to add that each ensemble member (i.e. each year in the ensemble) has been initialized separately, i.e. each year can be considered as a random realization (therefore we believe that this resampling procedure is appropriate).

The reviewer also addresses a more fundamental issue, namely the random nature of the model output. Here, we also agree with the reviewer, but would like to add another comment: For deriving the initial-condition ensemble that is used in our study, initial conditions for each ensemble member are perturbed randomly at the beginning of each year. Therefore, different years in the ensemble (but not different months...) can indeed be regarded as random realizations.

Other points:

Abstract, last line: 'uptake of our methodology. . .for accurately quantify- ing past. . .extremes' – how is bias correction important for quantifying past extremes which have been already observed? Perhaps the authors mean 'quantifying changes in past extremes'?

This is correct. Thanks.

Page 2011, first sentence – this information is repeating for the third time here.

This sentence has been removed.

Page 2021, Para 15: 'Although more sophisticated. . .in this study' – perhaps a 'that' missing?

For readability, sentence has been rewritten.

All references listed contain two years of publication each – please correct this. Also, Coles, 2001 is a single-author book. The reference to Coles, 2001 is incorrect in the list.

Thanks for pointing this out. It's corrected in the revised manuscript version.

Figure 3 (and similar figures) and Section 4.1 – Figure 3 is not self-explanatory. If the x-axis doesn't consist of values/units, then what to the width of each shape represent?

The description was added to the figure caption of Fig. 3 and Fig. 6: "Both sides of each violin are constructed as rotated, equal-area kernel density estimates, and a standard boxplot is drawn inside each violin."

Response to H. Hoffmann:

S. Sippel et al., 2015

Just going through the references:

Talking about 2d or multidimensional bias correction, the paper correctly cites the article "Two dimensional bias correction of temperature and precipitation copulas in climate models" of Piani and Haerter, September 2012. However, it should consequently cite as well the following article of Hoffmann and Rath, February 2012: Meteorologically consistent bias correction of climate time series for agricultural models. *Theor Appl Climatol* (2012) 110:129–141

Both articles deal with multidimensional bias correction and test 2d-examples. To my knowledge, these are the first two doing this.

We agree with the comment of the referee: Hofmann and Rath (2012) also applied a 2-dimensional bias correction methodology. In the revised manuscript, we cite both references that used a bivariate bias correction scheme.

Best regards

Sebastian Sippel on behalf of all authors

A novel bias correction methodology for climate impact simulations

Sebastian Sippel^{1,2}, Friederike E.L. Otto³, Matthias Forkel¹, Myles R. Allen³, Benoit P. Guillo³, Martin Heimann¹, Markus Reichstein¹, Sonia I. Seneviratne², Kirsten Thonicke⁴, and Miguel D. Mahecha^{1,5}

¹Max Planck Institute for Biogeochemistry, Hans-Knöll-Str. 10, 07745 Jena, Germany.

²Institute for Atmospheric and Climate Science, ETH Zürich, Sälimistr. 101, 8075 Zürich, Switzerland.

³Environmental Change Institute, University of Oxford, South Parks Road, Oxford OX1 3QY, UK.

⁴Potsdam Institute for Climate Impact Research, Telegrafenberg, 14473 Potsdam, Germany.

⁵German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Deutscher Platz 5E, 04103 Leipzig, Germany.

Correspondence to: Sebastian Sippel (ssippel@bgc-jena.mpg.de)

Abstract. Understanding, quantifying and attributing the impacts of extreme weather and climate events in the terrestrial biosphere is crucial for societal adaptation in a changing climate. However, climate model simulations generated for this purpose typically exhibit biases in their output that hinders any straightforward assessment of impacts. To overcome this issue, various bias correction strategies are routinely used to alleviate climate model deficiencies most of which have been criticized for physical inconsistency and the non-preservation of the multivariate correlation structure. In this study, we introduce a novel, resampling-based bias correction scheme that fully preserves the physical consistency and multivariate correlation structure of the model output. This procedure strongly improves the representation of climatic extremes and variability in a large regional climate model ensemble (HadRM3P, climateprediction.net/weatherathome), which is illustrated for summer extremes in temperature and rainfall over Central Europe. Moreover, we simulate biosphere-atmosphere fluxes of carbon and water using a terrestrial ecosystem model (LPJmL) driven by the bias corrected climate forcing. The resampling-based bias correction yields strongly improved statistical distributions of carbon and water fluxes, including the extremes. Our results thus highlight the importance to carefully consider statistical moments beyond the mean for climate impact simulations. In conclusion, the present study introduces an approach to alleviate climate model biases in a physically consistent way and demonstrates that this yields strongly improved simulations of climate extremes and associated impacts in the terrestrial biosphere. A wider uptake of our methodology by the climate and impact modelling community therefore seems desirable for accurately quantifying [changes in](#) past, current and future extremes.

1 Introduction

Weather and climate extreme events such as heat waves, droughts or storms cause major impacts upon human societies and ecosystems (IPCC, 2012). In recent years, these climatic events have changed in intensity and frequency in many parts of the world (Barriopedro et al., 2011; Donat et al., 2013; Seneviratne et al., 2014) and changes are likely to continue throughout the 21st century (Sillmann et al., 2013). Therefore, improving the scientific understanding of these events, including the link to impacts, constitutes an important research challenge (IPCC, 2012; Zhang et al., 2014).

The impacts of climate extremes and potential changes therein are strongly felt in the terrestrial biosphere. For example, heat and drought events trigger ecological responses (Reyer et al., 2013; Frank et al., 2015), which in turn induces changes to the cycling of water and carbon through such systems with potential feedback to the atmosphere and climate system (Reichstein et al., 2013; Frank et al., 2015). Indeed, on continental to global scales, it has been shown that large-scale reductions in photosynthetic uptake of carbon by plants are mainly driven by water limitations (Zscheischler et al., 2014a, b). Furthermore, it has been demonstrated that a single large event such as the European heat and drought summer 2003 alone might undo several years of ecosystem carbon sequestration (Ciais et al., 2005), thus potentially jeopardizing the terrestrial carbon sink potential (Lewis et al., 2011).

A widely debated question in this realm is whether the observed changes in the occurrence of climatic extremes and associated impacts can be attributed to specific changes in climate forcing, both anthropogenic or natural (Allen, 2003; Stone and Allen, 2005; Stone et al., 2009). To this end, large climate model ensembles are needed in order to derive robust probabilistic conclusions about changes in the odds of these events (Bindoff et al., 2013; Massey et al., 2014), because direct assessments of rare extremes are often prohibited by the lack of long and good quality observational time series. Hence, climate models are indispensable tools to study present and future climate extremes on various spatial and temporal scales, and the availability of such simulations is often a prerequisite for studying climate impacts.

However, despite considerable progress in recent years, global and regional climate models typically exhibit biases in various statistical moments of their simulated variables (Ehret et al., 2012; Wang et al., 2014), which often impedes direct assessments of climate extremes (Otto et al., 2012; Sippel and Otto, 2014) or simulating impacts (Maraun et al., 2010; Hempel et al., 2013). These biases are often due to an imperfect representation of physical processes in the models, parametrizations of sub-grid scale processes, and an over- or underestimation of feedbacks with the land-atmosphere or ocean-atmosphere feedbacks (Ehret et al., 2012; Mueller and Seneviratne, 2014). Due to the various origins of model biases, these biases are frequently varying depending on weather patterns both spatially and temporally, for instance in the distributed weather@home ensemble-based modelling framework (Massey et al., 2014) or in an ensemble of regional climate models (Vautard et al., 2013).

To alleviate this issue, various bias correction schemes have been ~~developped~~ developed in recent years that generally aim to statistically transform biased model output in order to derive more

realistic simulations (see e.g. Maraun et al., 2010; Teutschbein and Seibert, 2012). To do so, a statistical relationship ('transfer function') is built between the statistical distribution of an observed and simulated variable (Piani et al., 2010). Such methods span a wide range from very simple parametric transformations adjusting simulated means to observations (i.e. also called the 'delta method' (additive) or 'linear scaling' (multiplicative), (Teutschbein and Seibert, 2012)) to sophisticated, non-parametric approaches that aim to correct various statistical moments of the simulated distributions such as quantile mapping approaches (Wood et al., 2004; Gudmundsson et al., 2012).

65 However, the application of bias correction implicitly requires that a range of assumptions are met, which might be questionable in many cases and are discussed in detail in Ehret et al. (2012). Most importantly, the application of bias correction implicitly assumes that the statistical transformation improves the simulated output time series ('effectiveness'), whilst the signal of interest, e.g. the climate change signal or properties of the extremes, remains accurately detectable ('reliability'). Those assumptions are not always fulfilled since statistical bias correction methods are not based on physical principles, but operate rather heuristically on an observed model-data mismatch. To this end, even relatively simple methods that are designed to adjust 'only' simulated long-term monthly means to observations (e.g. Hempel et al., 2013) lack a sound physical rationale to whether these adjustments are to be made additively or multiplicatively. Further, the assumption of time invariant biases that currently underlies state-of-the-art bias correction procedures (Christensen et al., 2008; Ehret et al., 70 2012) might be especially critical for century-long climate simulations spanning several degrees of warming (Christensen et al., 2008; Buser et al., 2009) including changing land-atmosphere feedback processes (Seneviratne et al., 2006). Recent studies have shown that this assumption is questionable for future climate simulations (Maraun, 2012; Bellprat et al., 2013), and have made attempts to address time-dependent biases.

Furthermore, an adjustment of daily variability does not necessarily improve monthly statistics, thus emphasizing the role of time scales at which bias correction is conducted (Haerter et al., 2011). Lastly, if impact simulations are to be conducted with bias-corrected output of numerical climate models, the multivariate correlation structure between climate variables deserves attention: Most bias-correction schemes that are currently in use to simulate impacts have been suggested to correct 85 each variable separately (Hempel et al., 2013) and hence dependencies between variables are often not retained. This is especially critical for assessments of extreme events and 'compound events' (Leonard et al., 2014), where inter-variable interactions, such as soil moisture-temperature feedbacks might play an important role (Seneviratne et al., 2006). Although recent progress has been made to derive bivariate bias correction schemes (Piani and Haerter, 2012; Li et al., 2014) (Hoffmann and Rath, 2012; Piani and Haerter, 2012) to the best of our knowledge currently no bias correction scheme retains a multivariate correlation structure of a larger set of input variables for impact simulations.

In conclusion, accounting for biases in climate model output is crucial in order to produce credible climate model simulations. Nonetheless, statistical transformations are to be applied with caution

95 and the changes induced to the simulated statistical moments, multivariate dependencies and spatio-temporal patterns deserve considerable attention. Since the tails of statistical distributions are especially sensitive to changes in statistical moments such as the mean and variance (Katz and Brown, 1992), the latter holds in particular for assessments of extreme events and highlights the need for physically consistent ways to alleviate climate model biases.

100 In this paper, we demonstrate how a physically consistent bias correction of a regional climate model ensemble might aid to better simulate climatic extreme events and impacts in the terrestrial biosphere (see Fig. 1 for the methodological workflow of the paper).

First, we introduce a novel methodology to alleviate biases in the output of climate model ensembles that successfully circumvents major deficiencies of statistical bias correction (section 3): an ensemble-based probabilistic resampling approach retains the physical consistency of the regional climate model output. This includes the preservation of the multivariate correlation structure, and the procedure is shown to considerably improve the simulation of various statistical moments of the simulated variables. Secondly, we assess contemporary temperature and precipitation extreme events in Central Europe on monthly to seasonal time scales by comparing a widely used ‘standard’ statistical bias correction methodology (Hempel et al., 2013) with the original model simulations and the probabilistic resampling (section 4.1 and 4.2). This evaluation also focuses on the uncertainty induced by different observational datasets used as a basis for any bias correction approach. Thirdly, we explicitly test how differently corrected climatic data propagates into the simulation of impacts on major component fluxes of terrestrial carbon (net ecosystem exchange (NEE), gross primary production (GPP) and ecosystem respiration (Reco)) and water cycling (actual evapotranspiration, AET) in the terrestrial biosphere using a dynamic vegetation model (LPJmL, Section 4.3). To this end, we demonstrate that different ways to deal with biases in climate simulations yield both qualitatively and quantitatively different results regarding simulated impacts, which affect both central moments of the distribution as well as extremes and variability.

120 **2 Data**

2.1 Climate model simulations

In this study, regional climate model ensemble simulations spanning 26 years (1986-2011) with approx. 800 ensemble members per year from the weather@home distributed computing platform are investigated (Massey et al., 2014). The ‘atmosphere-only’ simulations were conducted over the European region (identical to the EURO-CORDEX region Giorgi et al., 2009) using a regional model (HadRM3P) on a rotated grid nested into the global HadAM3P model. Both models share the same model formulation and are described in Pope et al. (2000). The regional (global) simulations are run with a spatial resolution of $0.44^\circ \times 0.44^\circ$ ($1.875^\circ \times 1.25^\circ$) with 19 vertical levels, and the temporal resolution is set to 5 (15) minutes (Massey et al., 2014). The models are driven by observed sea

130 surface temperatures and sea ice fractions, the observed composition of the atmosphere (greenhouse gases, aerosols) and anomalies in the solar cycle (Massey et al., 2014). To derive different ensemble members, the initial conditions of the driving GCM are perturbed on 1st December of each 1-year simulation (ibid.). For further analysis and bias correction, the ensemble simulations were remapped to 0.5 ° spatial resolution using a conservative remapping scheme (Jones, 1999).

135 Massey et al. (2014) demonstrate that the ensemble setup described above produces a realistic representation and statistics of European weather events, including the extremes for most seasons and regions. However, despite these encouraging results, a relatively large mismatch remains between the statistical distribution of the ensemble simulations and the observations in Northern hemisphere summer, which holds for the means of simulated seasonal temperature and precipitation (Massey et al., 2014) as well as for higher statistical moments, shown in the Supplement against the ERA-Interim reanalysis dataset (Dee et al., 2011). Especially in more continental parts of the European model region, HadRM3P shows a pronounced hot and dry bias in simulated summer weather (Figs. S1-S3 in the Supplement). However, note that the ensemble setup still captures the entire range of the observed distribution (Fig. S1). In HadRM3P, these biases are likely related to an imperfect parametrization of cloud processes in the model, leading to an overestimation of incoming solar radiation, which in turn triggers warm and dry summer conditions (R. Jones, [2015](#), pers. comm.) that are further amplified by strong soil moisture-temperature coupling in the model (Fig. S4). In this context, it is worthwhile to note that these biases are not a peculiarity of the regional climate model employed in this study, but indeed hold for many dynamically downscaled regional climate model simulations over Europe (Buser et al., 2009; Boberg and Christensen, 2012).

2.2 Simulation of atmosphere-biosphere carbon and water fluxes

To assess terrestrial biosphere impacts of bias correcting regional climate simulations (see section 4.3), we simulate ensembles of atmosphere-biosphere fluxes of carbon (NEE, GPP, Reco) and water (AET) using the Lund-Potsdam-Jena managed land scheme (LPJmL, Version 3.5, Sitch et al., 155 2003; Bondeau et al., 2007), a state-of-the-art process-based dynamic global vegetation model that ~~account~~[accounts](#) for human land use. We follow Schulze (2006) and Chapin III et al. (2006) in their definition of major components of carbon cycling in terrestrial ecosystems: Gross primary productivity (GPP) denotes the vegetation's gross photosynthetic uptake of carbon from the atmosphere, whereas ecosystem respiration (Reco) is defined as the respiratory release of carbon by plants and 160 microbes in the ecosystem, i.e. including both (autotrophic) plant respiration and (heterotrophic) soil organic matter decomposition. Net ecosystem exchange (NEE) constitutes the net carbon flux from the ecosystem to the atmosphere, i.e. the difference between Reco and GPP.

LPJmL simulates vegetation dynamics (growth, competition and mortality) and fully coupled cycling of carbon (photosynthesis, autotrophic & heterotrophic respiration) and water (transpiration, 165 evaporation, interception, runoff) in terrestrial ecosystems and managed systems (Sitch et al., 2003;

Gerten et al., 2004; Bondeau et al., 2007). The model is driven with monthly or daily climatic input data (temperature, precipitation, incoming shortwave radiation & net longwave radiation), atmospheric carbon dioxide concentrations and soil texture. Vegetation structure in LPJmL is characterized by the fractional coverage of 11 plant functional types that differ in their bioclimatic limits and ecophysiological parameters. Vegetation dynamics and competition are explicitly represented using
170 a set of allometric and empirical equations and updated annually (Sitch et al., 2003).

GPP in LPJmL follows the process-oriented coupled photosynthesis and water balance scheme of the BIOME3 model (Haxeltine and Prentice, 1996). Subsequently, autotrophic (growth and maintenance) respiration is subtracted from GPP, and the net carbon uptake is allocated to plant compartments based on a set of allometric constraints (Sitch et al., 2003). Ecosystem heterotrophic respiration
175 depends on temperature and moisture in each litter and soil carbon pool; carbon decomposition dynamics are simulated as first-order kinetics with specified decomposition rate in each pool (Sitch et al., 2003). Water cycling in LPJmL has been improved by Gerten et al. (2004) and Schaphoff et al. (2013), where actual evapotranspiration (the sum of evaporation, transpiration and interception)
180 is computed as a function from atmospheric demand and soil moisture supply. Phenology and photosynthesis-related parameters in the LPJmL version used in this paper have been optimized against remote sensing observations for an improved simulation of natural vegetation greenness dynamics (Forkel et al., 2014), including the introduction of a novel phenology scheme.

LPJmL has been applied in a range of studies assessing ecosystem responses to anomalous climatic conditions (Rammig et al., 2014; Van Oijen et al., 2014; Zscheischler et al., 2014b; Rolinski et al., 2015). Rolinski et al. (2015) argued that the model might be able to capture various ecosystem physiological responses to climatic extreme events such as heat or drought through various pathways. These include a water stress response through reduced stomatal conductance, which in turn decreases both photosynthetic carbon uptake and transpiration. Further, the model responds to very
185 high temperatures by a photosynthesis inhibition and increased respiration (Rammig et al., 2014).

In this paper, we use the weather@home climate data to derive ensemble-based simulations of the functioning of terrestrial ecosystems. LPJmL simulations are conducted in natural vegetation mode (i.e. no human land use, fire or permafrost) in 0.5° spatial resolution and monthly time steps over Central Europe. For each bias-corrected ensemble dataset, 2000 years of spinup to equilibrate soil
195 carbon pools were conducted, using randomly chosen years from the first 10 years of the HadRM3P ensemble. Subsequently, atmosphere-biosphere fluxes were simulated at the monthly time scale for 1986-2010 over Central Europe (see Fig. 1 for methodological workflow). This procedure was repeated five times to check that no carry-over effects from the randomized spinup affect simulated biosphere-atmosphere carbon fluxes in the transient period. Since this was not the case, differences
200 in carbon and water fluxes and their extremes can be directly attributed to the bias correction of the climatic forcing in the transient period, and are analyzed in section 4.3.

2.3 Observations

Any statistical assessment or correction method of biases requires reference datasets, and the quality of bias adjustment is thus restricted by the quality of observations or reanalysis data available (Ehret et al., 2012; Hempel et al., 2013). Consequently, the sensitivity of ‘bias corrected’ model output to any given set of observations needs to be tested. In this study, a range of observational datasets is used in order to characterize uncertainty induced by using different observations for bias correction. In total, seven different temperature and precipitation datasets consisting of gridded observations/reanalysis were used ([one at a time](#)) for the univariate bias correction (section 4.2) and are detailed in Table 1. The simultaneous correction of multiple variables for the impact simulations in the terrestrial biosphere presented in section 4.2 are conducted using ERA-Interim as reference dataset (Dee et al., 2011, see table 1).

To conduct the sensitivity analysis of climatic extremes and associated biosphere impacts to the type of bias correction applied, we select one focus region in Central Europe. This region roughly encompasses Germany (47.5 – 55.0°N, 6.0 – 15.0°E, see e.g. Fig. S2a) and consists of temperate mid-latitude climate with maritime influence to the North-West and more continental characteristics to the East. In addition, to sample local (i.e. grid cell scale) variability we test different bias correction scheme on one single grid cell located in Central Germany (‘Jena pixel’, 50.75°N, 11.75°E).

3 Methods

In this section, we describe the different bias correction methods deployed in this study. First, a bias correction methodology for impact [simulation simulations](#) that has been adopted widely is summarized (Hempel et al., 2013). Second, we introduce the novel resampling-based bias correction scheme and lastly the methodologies for evaluation are described.

3.1 Statistical Bias Correction

Hempel et al. (2013) presented a bias-correction that is designed to preserve long-term trends in simulated impacts and that has been used widely in simulating effects of climatic changes in different sectors such as water, agriculture, ecosystems, health, coastal infrastructure, and agro-economy (see Warszawski et al., 2014, for an overview).

The approach builds on earlier, conventional statistical bias correction schemes (Piani et al., 2010; Haerter et al., 2011) and is based on linear transfer functions of the form

$$x_{cor} = a + bx. \tag{1}$$

Here, x and x_{cor} represent the simulated and corrected climatic variable, a and b are coefficients to be calibrated.

In Hempel et al. (2013) the transfer function is applied additively (for temperature, i.e. $b = 1$),
235 such that

$$a = \overline{T_{obs}} - \overline{T_{mod}}; \quad (2)$$

where $\overline{T_{mod}}$ and $\overline{T_{obs}}$ represent the means of simulated and observed monthly temperatures, respec-
tively.

To account for positivity constraints for precipitation and radiation components, Hempel et al.
240 (2013) suggested a multiplicative adjustment of those variables (i.e. $a = 0$), such that

$$b = \frac{\overline{x_{obs}}}{\overline{x_{mod}}}. \quad (3)$$

These parametric transformations are applied on each grid cell and for each month separately to
account for potential temporal and spatial structure in the biases. By applying this transfer function,
long-term monthly means of the simulated distributions are matched with those in observations for
245 each grid cell (Hempel et al., 2013). In addition to adjusting monthly means, Hempel et al. (2013)
also adjust daily variability about the monthly means, but (importantly) the year-to-year variability
at monthly time scales remains unchanged. In our present analysis, we follow this conventional bias
correction scheme for comparison and denote it by ‘ISIMIP’.

Furthermore, to isolate the effects of bias-correcting the full suite of impact variables (temper-
250 ature, precipitation and radiation) vs. correcting simulated precipitation only, we conduct impact
simulations with a ‘precipitation only’ bias-corrected scenario (‘PRECIPCOR’).

3.2 A Novel Resampling-Based Ensemble Bias Correction Scheme

~~Conventional statistical bias correction methods have been criticized due to their physical inconsistency
and the non-preservation of dependency relationships between meteorological variables (Ehret et al., 2012).~~

255 In this subsection, we introduce a novel ‘bias correction scheme’ suitable for ensemble simulations
that retains the physical consistency and multivariate correlation structure of the model output. The
idea is to resample plausible ensemble members from a large ensemble simulation given the statisti-
cal distribution of an observable meteorological metric (‘constraint’). The procedure is illustrated
using the weather@home ensemble described above.

260 The largest biases in the HadRM3P simulation occur in the summer season (JJA) over the Euro-
pean model domain, where the model ensemble produces too frequent and too pronounced hot and
dry conditions (Massey et al., 2014). Importantly however, the ensemble spans the entire distribution
of observed summer conditions in most parts of Europe, i.e. some (but too few) ensemble members
produce relatively wet and cold summers. Therefore, our resampling-based correction approach is
265 designed to alleviate the representation of summer conditions in the model ensemble.

The bias correction procedure consists of the following steps and is illustrated in Fig. 2:

1. Define an observable meteorological metric that is poorly represented (‘biased’) in the model ensemble. In this paper, we use summer mean temperatures over Central Europe, which are relatively well-constrained in observational datasets.
- 270 2. Estimate the probability distribution function of the meteorological constraint from observational datasets using e.g. a kernel density fit ($\hat{f}_{obs}(x)$, see e.g. Fig. 22aa, blue line for an illustration), where x denotes the constraint. Here, we use a Gaussian kernel with reliable data-based bandwidth selection (Sheather and Jones, 1991) fitted over the observed meteorological constraint for the period 1986-2011 ~~in using~~ various observational datasets (one at a
275 time).
3. Estimate the probability distribution of the meteorological constraint in the model ensemble using the same estimation procedure as above over all ensemble members and all years ($\hat{f}_{mod}(x)$, see Fig. 2a, red line). The deviation between the red and blue line in Fig. 2a illustrates the temperature bias in the weather@home ensemble.
- 280 4. Derive a transfer function that maps any given quantile in the observations ($q_{obs,X}$) to the respective quantile in the model ensemble ($q_{mod,X}$, see Fig. 2b), such that $q_{mod,X} = TF(q_{obs,X})$ using the fitted kernels $\hat{f}_{obs}(x)$ and $\hat{f}_{mod}(x)$ to determine empirical quantile functions. For example, a ‘median temperature’ summer over Central Europe (approx. 17.2°C) would correspond to the 50th percentile in the observations-based kernel (by definition). The transfer function would then map the 50th percentile in \hat{f}_{obs} to the corresponding 20.4th percentile in \hat{f}_{mod} (i.e. average summer temperatures of 17.2° would correspond to the 20.4th percentile in the model ensemble, see Fig. 2b). In this study, we use Cubic Hermite splines (Fritsch and Carlson, 1980) to determine the transfer function shown in Fig. 2b.
- 285 5. Derive a new ‘bias-corrected’ ensemble (of sample size n) by randomly resampling n times from ~~\hat{f}_{obs}~~ observed percentiles ($q_{obs,X}$) and retaining the ensemble member that corresponds to $q_{mod,X}$ as given by the transfer function ($n = 800$ per year in our study).
- 290

Hence, the outlined procedure does not adjust any output variable in the model ensemble thus preserving physical consistency, but rather selects plausible ensemble members. This procedure invariably leads to a reduction in the effective ensemble size: For example in the HadRM3P ensemble, 295 roughly the hottest 20% of simulations are effectively not chosen for the resampled ensemble since they are implausibly hot (Fig. 2a). However, an evaluation of the sample size in the bias corrected ensemble shows that at least 4% of the ensemble simulations match any decile of observations (Fig. 2e and 2d, in an unbiased ensemble exactly 10% of ensemble simulations would match each decile of observations), corresponding to an effective sample size of at least approx. 1000 model years (= 300 ensemble members) per decile of observations (Fig. 2ed).

In conclusion, the outlined approach to bias correction is conceptually similar to earlier ideas of assigning weights to different regional climate model projections based on each model’s performance in order to derive probabilistic multi-model projections (Piani et al., 2005; Collins, 2007; Knutti, 2010; Christensen et al., 2010). However, instead of a weight assignment specific ensemble members are selected and combined into a new ensemble using the statistical distribution of observed meteorological constraints.

3.3 Analysis Methodology

In section 4.1 the outlined bias correction method is evaluated for the simulation of temperature, rainfall and radiation using standard evaluation metrics such as seasonal mean values and interannual variability. Further, we evaluate soil moisture coupling in the original and bias corrected ensemble against reanalysis data and upscaled observations by computing the correlation between summer mean temperatures and the mean latent heat flux following Seneviratne et al. (2006).

Moreover, we analyze empirical return times of the original and bias-corrected ensembles that are derived by plotting each ensemble value against its rank both for climatological extremes (section 4.2: monthly summer temperatures and cumulative summer rainfall) and simulated ecosystem-atmosphere annual fluxes of water and carbon (section 4.3).

To further understand discrepancies between the bias-corrected ensemble simulations and observed climate extremes (section 4.2), we characterize the tails of simulated and observed variables by extreme value theory (Coles, 2001). Hence, generalized extreme value distributions (GEV) are derived from monthly temperature and precipitation in a procedure similar to Sippel et al. (2015a), i.e. by resampling block-maxima in randomly concatenated 10-year sequences of ensemble data and fitted to a GEV model using generalized maximum likelihood estimation. In observational data, only a relatively small sample size is available (mostly 1901-2014 only) that is additionally plagued by non-stationarity and does not match the period in which ensemble simulations are available (1986-2011). Hence, for monthly temperatures we subtract the trend and seasonal cycle from the original time series using Singular Spectrum Analysis (von Buttlar et al., 2014), and subsequently resample (monthly) summer temperature anomalies (for the whole time series) by adding a trend and seasonal cycle component drawn randomly from the period of available ensemble simulations (1986-2011, [each observational dataset is analysed separately](#)). Approximate stationarity was assumed for seasonal precipitation, and hence no further adjustments were made. Lastly, GEV models were fitted to the observations following the procedure as described above.

4 Results

This section is structured as follows: First, we evaluate the bias correction procedure both for resampling based on an area mean and grid cell based constraint. Second, climate extreme statistics and

335 their sensitivity to bias correction schemes are investigated (section 4.2). More specifically, the prob-
abilistic resampling scheme introduced in sect. 3.2 is evaluated against a conventional bias correction
scheme (Hempel et al., 2013, , section 3.1) and compared against the uncorrected simulations and
different observational datasets. Third, we illustrate how biases and their ‘correction’ propagate into
climatic impacts exemplified by simulations ecosystem water and carbon fluxes in Central European
340 natural vegetation.

4.1 Evaluation of Resampling Bias Correction

An evaluation of the distribution of variables in the resampled ensemble in Central Europe shows that
it not only improves the simulation of seasonal mean temperatures (which it does by construction),
but also yields considerable improvements to the simulation of rainfall and radiation components
345 (Fig. 3). This suggests that these biases are related to specific synoptic situations in summer, justi-
fying to apply the bias correction approach to summer months. Hence, the multivariate covariance
structure between temperature, precipitation and radiation as simulated by HadRM3P appears to be
well represented in the model simulations posterior to the updating procedure given the reanaly-
sis/observational data. Moreover, while this procedure also improves the simulation of summer tem-
350 peratures and precipitation on a monthly time scale, virtually no changes in the ensemble statistics
are induced to non-summer months (Fig. S1), indicating that the time scales of temporal decorrela-
tion are short enough for a successful application of the resampling procedure (Fig. S1). However,
while conventional statistical bias correction following Hempel et al. (2013) adjusts monthly means
of the distributions of precipitation and radiation (by construction), changes are induced by the mul-
355 tiplicative adjustment to the width and shape of the distribution, including its tails (Fig. 3, see also
section 4.2).

An evaluation of the resulting spatial patterns of the resampling bias correction shows that the
representation of the simulated statistical distributions of temperature and precipitation are consider-
ably improved in Central Europe (area mean constraint) and across the entire European model region
360 (single grid cell constraints, Figs. S2-S3). Remarkably, this holds not only for seasonal averages, but
also for higher statistical moments such as the inter-decile range (Figs. S2-S3).

Furthermore, we test the representation of land-atmosphere coupling in the original and resampled
model ensemble by investigating the correlation strength between summer mean temperatures (T)
with latent heat (LE) fluxes following Seneviratne et al. (2006). The original HadRM3P ensemble
365 shows strong water limitation of evapotranspiration in summer (negative correlation between LE
and T) for most temperate and Mediterranean European regions, thus overestimating soil moisture
control compared to reanalysis data and upscaled observations (Fig. S4). In the resampled ensemble,
land-atmosphere coupling remains strongly soil moisture controlled in the Mediterranean regions,
but reduces in temperate European regions, resulting in spatial patterns that resemble those of land-
370 atmosphere coupling in ERA-Interim (Fig. S4). The latter finding indicates that the procedure of

eliminating implausible ensemble members also yields an improved representation of physical processes such as land-atmosphere coupling in the resampled ensemble.

4.2 Sensitivity of Climatic Extremes to Bias Correction

4.2.1 Summertime temperature extremes

375 Summertime monthly extreme temperatures are shown in Fig. 4 as a spatial average for the study region located over Central Europe and for an illustrative and randomly chosen grid cell ('Jena grid cell').

The location, slope and shape of the lines in the return time plots shown in Fig. 4 reveal that the tails of simulated monthly temperature extremes are highly sensitive to the type of bias correction applied, both for a regional average and a single grid cell: Uncorrected simulations overestimate both location and scale (i.e. slope of the line in the return time plot) of positive temperature anomalies in summer, while this is not the case for anomalously cold summer months (Fig. 4). An additive adjustment of monthly means (orange lines in Fig. 4, Hempel et al., 2013) preserves slope and shape of the tail, i.e. preserves the year-to-year variability of simulated monthly temperatures (and biases therein) in the ensemble. Note that this procedure cannot account for the asymmetry between the upper and lower tail of simulated monthly temperatures - i.e. the offset correction leads to an overcorrection of cold months, whereas the statistics of the hot tails improve only marginally. This is confirmed by a statistical extreme value analysis (Figs. S5-S6): The temperature offset approach adjusts only the location of the GEV yielding spurious artefacts in the (originally well simulated) cold tail, whilst not accounting fully for the upper tail due to the aforementioned asymmetries. This is a fundamental drawback of using linear parametric transfer functions, i.e. even if the variability of the simulated distributions would have been adjusted along with the means (see e.g. Sippel and Otto, 2014), the outlined 'asymmetry' issue would not necessarily improve. On the other hand, the probabilistic resampling procedure alters both the location and slope of the lines in the return time plot, where resampling based on a spatial average as well as on a grid cell constraint yield relatively similar representations of the tails. An evaluation of the extreme value statistics shows that the probabilistic procedure indeed considerably improves the statistical characteristics of the simulated tails in the ensemble compared to (long-term) observations (Figs. S5-S6). To this end, resampling the original ensemble changes location and scale of the extreme value distributions, but the shape parameter of the tails remain effectively unchanged. Some caution is required due to the relatively scarce availability of observed monthly mean temperatures (i.e. 1901-2014), which induces considerable uncertainties to the parameters of the fitted GEV distributions (Figs. S5-S6). Moreover, the different time periods of observations and ensemble simulations (1986-2011) impede a direct 'evaluation' of the bias correction. Nonetheless, this indicative comparison yields very promising results of bias-correcting without invasive changes to the simulated statistical distribution.

Lastly, our analysis shows that any bias correction based on a single grid-cell level induces some uncertainty due to the choice of observational dataset. This is an important issue to consider if impact model simulations on a grid cell scale are to be conducted, whereas regional averages are not as strongly affected. Fig. 4 shows that resampling the ensemble based on a spatial average constraint
410 reduces this uncertainty as compared to adjusting monthly means or resampling on a grid cell scale.

4.2.2 Summertime rainfall extremes

We extend the analysis of the previous paragraph to investigate how resampling based on a temperature constraint alters the representation of summer precipitation in a large ensemble simulation. The original HadRM3P simulated summer seasons are too dry in average over Central Europe (Fig. S2), which is largely due to a much too dry lower tail (Fig. 5), whilst simulated heavy monthly
415 precipitation matches relatively well the available observational data (Fig. 5).

The tails of simulated (cumulative) seasonal precipitation are sensitive to bias correction. As above, the plots in Fig. 5 illustrate that a statistical adjustment of the means can be detrimental to statistics of extremes and variability. For instance, scaling monthly means to match observations
420 (Hempel et al., 2013) leads to an inflation of very wet seasons that are physically implausible given the observations (Fig. 5, orange lines). Likewise, the (biased) dry tail in HadRM3P improves only to a very limited extent if the scaling approach is used. The extreme value analysis (Fig. S6) shows that the multiplicative adjustment changes both location and scale of the tail distribution - and that both parameters are not necessarily improving (indeed often deteriorating, see e.g. scale parameters in
425 Fig. S6) by applying a simple statistical bias correction. However, resampling based on a *temperature* constraint yields a new ensemble, in which the simulation of both tails has improved (Fig. 5, Fig. S6). Only minor changes have been induced to the (well-simulated) wet tail, whilst the previously strongly biased dry tail has considerably improved (Fig. 5, Fig. S6), indicating that temperature-based resampling as deployed here successfully separates ‘plausible’ ensemble members from the
430 (unrealistic) hot and dry members. The extreme value analysis shows that resampling largely alters the location of the simulated distribution of seasonal rainfall extremes, whilst the scale and shape of the tails remain largely unchanged.

To conclude, it was shown that resampling based on a univariate observations-based temperature constraint improves the simulation of rainfall variability and extremes by teasing out ensemble
435 members that are implausibly hot and dry in our case study region.

4.3 The impact of bias correction on simulated ecosystem water and carbon fluxes

In this subsection, we present HadRM3P-LPJmL ensembles of simulated fluxes of carbon and water and discuss bias correction methods with a focus on the extreme tails of the simulated distributions. Further, we investigate the sensitivity of the simulated carbon fluxes to an accurate representation of
440 rainfall in the climatic input data.

Annual mean fluxes across the large ensemble of NEE, GPP, Reco, and AET are shown in Table 2 for the 1986-2010 period for each bias correction and the control simulation. Conventional statistical bias correction that matches monthly means of the HadRM3P ensemble *exactly* to those of the ERA-Interim control climate simulation yields differences in fluxes of -6.6% , -7.5% and -4.7% for
445 GPP, Reco and AET, respectively. Note that differences in the resampled HadRM3P ensemble are less pronounced (-4.2% , -4.5% , and -2.0% , respectively), although no attempt has been made to adjust the statistical properties of the model output. Those differences in simulated annual mean fluxes are related to higher statistical moments of the statistical distributions and shown in Fig. 6.

To this end, simulated GPP, NEE, and AET show strong asymmetry in their simulated distributions (Fig. 6): Negative anomalies in GPP and AET are much more pronounced than positive ones;
450 this holds also for NEE but with an inverted sign (ecosystem carbon release corresponds to positive fluxes). However, the simulation of these extremes is highly sensitive to bias correction, where the lower tails of GPP and AET in the original and statistically bias corrected ensemble strongly overestimate reductions in carbon and water flux. In contrast, negative GPP and AET anomalies in the
455 resampled ensemble (corresponding to positive ones in NEE) exhibit a much less pronounced lower tail and asymmetry and agree well with the control simulations.

For example, a positive anomaly in NEE corresponding to a 30-year return period exceeds $+200$ $\text{g C m}^{-2} \text{ year}^{-1}$ in the conventionally bias corrected simulations and the original ensemble, whereas such an anomaly in the resampled ensemble hardly reaches $+150$ $\text{g C m}^{-2} \text{ year}^{-1}$ (Fig. 6b) roughly
460 corresponding to an empirical 30-year return event in the ERA-Interim control simulations. Similar arguments can be made for negative anomalies in annual GPP and annual AET (Fig. 6). The different tails of the simulations occur because the original meteorological ensemble implies large hot and dry biases in summer, inducing negative anomalies in ecosystem-atmosphere carbon and water cycling. These biases are not accounted for by conventional statistical bias correction but they are alleviated
465 if an ensemble resampling scheme is used (see previous subsection). However, this is remarkable because monthly means of precipitation in PRECIPCOR and ISIMIP are identical to the control climate simulation, which highlights the importance to consider statistical moments beyond the mean for impact simulations.

However, note that the positive tails of GPP and AET are not as strongly affected. Furthermore,
470 ecosystem respiratory fluxes show a relatively lower sensitivity to bias correction (i.e. to hot and dry summer conditions).

Further, we investigate whether different bias correction schemes induce different sensitivities of LPJmL simulated carbon fluxes to rainfall. Here, the relationship between a growing season rainfall proxy (April-September rainfall sums) and annual NEE is characterized using piecewise linear
475 regression (Fig. 7a-d). Fig. 7e shows that LPJmL simulated annual NEE responds to rainfall in a roughly similar way across different bias correction schemes, which highlights the need of an accurate representation of precipitation in climate impact simulations in the terrestrial biosphere. How-

ever, characterizing the annual NEE response for each quantile of the rainfall distribution shows that the resampled rainfall distribution (PROBCOR) leads to a less negative NEE response to rainfall (larger slopes in Fig. 7f), whereas a dry summer tail (in the ORIG, ISIMIP, and PRECIPCOR simulations) yields a generally stronger NEE response (more negative sloped in Fig. 7f).
480

In conclusion, different bias correction methods induce different statistical properties of simulated ecosystem-atmosphere fluxes of carbon and water. This affects the variability and skewness of NEE, GPP and AET simulations (as shown in Fig. 6), where hot and dry biases in summer imply a
485 disproportional reduction in carbon and water fluxes in climatically ‘unfavourable’ years. Conventional statistical bias correction cannot account for this issue, whereas the novel probabilistic bias correction schemes alleviates those biases to a very large extent.

5 Discussion

In this paper, we have introduced a novel ensemble-based resampling bias correction approach that
490 retains the physical consistency and multivariate correlation structure of regional climate model output. The [approach thus relies on a physically consistent set of climate model simulations \(i.e. closure of water and energy balances\)](#). The methodology is conceptually similar to earlier approaches designed to constrain future probabilistic climate predictions based on observational constraints (Piani et al., 2005; Collins, 2007). Its application has been shown in this paper to yield considerably improved simulations of weather and climate extremes. Remarkably, the improvement holds for variables that have not been constrained upon (i.e. constraining on seasonal mean temperatures improves the representation of mean and extreme precipitation), which indeed emphasizes the importance to
495 bias correct in a physically meaningful way.

Furthermore, simple but widely used statistical bias correction methodologies (e.g. Hempel et al.,
500 2013) have been evaluated with respect to the effect on the representation of weather and climate extremes on monthly to seasonal time scales. These methods cannot account for biases associated with e.g. specific synoptic situations that result in biases in higher statistical moments of the simulated distributions, which indeed emphasizes the importance to bias correct in a physically meaningful way. We demonstrated that this shortcoming of conventional methodologies can be detrimental to
505 statistics of weather and climate extremes and their variability. ~~Although more~~ [More](#) sophisticated statistical bias-correction schemes (see Gudmundsson et al., 2012, for an overview) [that](#) might have an improved skill in rectifying biases in higher statistical moments (such as e.g. asymmetries in simulated distributions) have not been explicitly tested in this study. [However](#), the fundamental question of how physical consistency can be preserved after bias correction (Ehret et al., 2012), including
510 multivariate dependencies between variables, remains elusive. Therefore non-linear and nonparametric bias correction techniques (Gudmundsson et al., 2012) might potentially improve statistics of extreme events if a large enough sample of observations is available, but cannot retain physi-

cal consistency (Sippel and Otto, 2014) and may ultimately fall short for correcting a set of input variables.

515 To this end, we have explicitly simulated an ensemble of ecosystem-atmosphere fluxes of carbon and water using a state-of-the-art biosphere model (LPJmL) in order to test the sensitivity to bias correction. Similarly to above, we find that bias correction induces strong effects on the representation of extremes and variability in carbon and water fluxes (section 4.3). Mechanistically, the stark contrast between the bias correction schemes can be traced back to the sensitivity of the LPJmL
520 model to dry conditions (see e.g. Rammig et al., 2014; Rolinski et al., 2015): NEE, GPP and AET in Central Europe are to a large extent driven by the availability of rainfall in the growing season, except for wet conditions, under which the relationship levels off (Fig. 7). Bias correction strongly affects the variability and extremes of rainfall (as shown above), thus inducing pronounced asymmetries in simulated water and carbon fluxes (Fig. 7f, Fig. 6). Therefore, our results highlight the importance
525 to account not only for biases in the mean but also for higher moments in the climatic input in order to generate robust insights into the past, present and future climate impacts. Our results demonstrate that physically consistent bias correction schemes might be preferable for this task. Moreover, it has been shown recently that climatic drivers exert multivariate controls on ecosystem responses such as phenology and vegetation greenness dynamics (Forkel et al., 2015), therefore accurate ecosystem impact simulations requires bias correction schemes that preserve the correlation structure of
530 climatic data.

However, several limitations of the present methodology should be discussed: First, probabilistic resampling based on a regional observational constraint cannot account for biases on very large regional or continental scales if the biases show a spatially or temporally heterogeneous structure or
535 gradients. In the latter case, resampling-based bias correction could lead to spurious artefacts in the spatio-temporal structure of the bias-corrected model domain. ~~Secondly~~Hence, a careful evaluation of the ensemble resampling approach has to be made - particularly with a focus on the spatial and temporal extent of the constraint and the resampled ensemble: A trade-off exists between resampling on small domains (e.g. grid-cell based) that is sensitive to the choice of observational dataset, and
540 very large domains that might be prone to a spatio-temporal bias structure. ~~Thirdly, the applicability of bias correction methods for future projections is currently unclear, since previous studies have shown that biases in climate projections (e.g. for the 21st century) are unlikely to be stationary (Ehret et al., 2012; Maraun, 2012). However, an application of the resampling approach to future projections similarly to the current practice of statistical bias correction (Hempel et al., 2013, e.g.) would~~
545 ~~be straightforward, i.e. based on a calibration using present or past conditions. Lastly, the resampling approach~~Secondly, the resampling approach requires relatively large ensemble sizes to be effective: in order to plausibly cover the climate space in any particular location, the simulated ensemble should cover the entire observed distribution. However, this condition does not necessarily restrict resampling-based bias correction methods to large ~~ensembles~~ensemble simulations: For example,

550 under the assumption of ergodicity for a given time period, resampling shorter time periods (e.g. single years) from smaller ensembles such as CORDEX regional simulations (Giorgi et al., 2009) would provide a promising topic for further study. In this context, the applicability of the resampling methodology would depend on the remaining effective sample size after the resampling step. The latter is a function of the biases in the model and the number of ensemble members available, and
555 could be tested in an evaluation step similarly to Fig. 2d. Thirdly, the applicability of bias correction methods for future projections is currently unclear, since previous studies have shown that biases in climate projections (e.g. for the 21st century) might not be stationary (Ehret et al., 2012; Maraun, 2012) . However, an application of the resampling approach to future projections similarly to the current practice of statistical bias correction (Hempel et al., 2013, e.g.) would be straightforward, i.e. based
560 on a calibration using present or past conditions. Lastly, a clear distinction between bias correction and statistical downscaling is crucial (Maraun, 2013) : While the resampling bias correction is designed to account for the former, no attempt of statistical downscaling or bridging any scale mismatches is made (see, e.g. Maraun, 2013, for a detailed discussion) .

Notwithstanding these limitations however, we show the usefulness of the novel bias correction
565 scheme that might be a useful and physically consistent alternative to conventional statistical bias correction as long as global and regional dynamical climate models suffer from pertinent biases.

6 Conclusions

In this paper, we introduced a novel bias correction method that retains physical consistency and the multivariate correlation structure of the climate model output based on an ensemble resampling
570 approach. We showed that such an approach strongly improves

- a) statistics of weather and climate extreme events, and
- b) the simulation of climate impacts such as ecosystem-atmosphere fluxes of carbon and water, including extremes and variability therein.

The methodology could be readily taken up in probabilistic event attribution studies that deploy
575 large ensembles simulations (see Stott et al., 2013, for an overview) in order to more realistically describe the statistics of (changing) extreme events.

Furthermore, detecting and attributing the impacts of climatic variability and extremes on hydro-
logical and socio-ecological systems has emerged as a highly topical research area (Stone et al.,
2009, 2013), including demonstrated interest by stakeholders across various sectors (Schiermeier,
580 2011; Stott and Walton, 2013; Sippel et al., 2015b). To this end, our study showed that it is crucial to account for higher statistical moments in biased climatic input data, and to correct climatic biases in a physically consistent way. Therefore, our methodology could be taken up by the climate impact modelling community to reduce climate forcing biases to a very large extent without requiring any modifications to the climate model output.

585 *Acknowledgements.* We would like to thank our colleagues at the Oxford eResearch Centre for their technical
expertise, the Met Office Hadley Centre PRECIS team for their technical and scientific support for the devel-
opment and application of weather@home and all of the volunteers who have donated their computing time to
climateprediction.net. We thank Richard Jones, Antje Weisheimer, Lukas Gudmundsson, Jose Manuel Gutier-
rez Llorente, Nuno Carvalhais, Mirco Migliavacca and Soenke Zaehle for comments and productive discussions
590 regarding the bias correction of regional climate model output.

We appreciate the creators and providers of observational and reanalysis datasets, including in particular the
Berkeley Earth Observations (Rohde et al., 2013), University of East Anglia Climatic Research Unit (CRU,
TS3.2, Phil Jones & others; Harris et al. (2014)), CRUNCEP dataset ([http://dods.extra.cea.fr/data/p529viov/
cruncep/readme.htm](http://dods.extra.cea.fr/data/p529viov/cruncep/readme.htm)), Global Precipitation Climatology Centre (GPCC) Full Data Reanalysis Version 6.0 at
595 Deutscher Wetterdienst (Schneider et al., 2011, 2014), the E-OBS dataset from the EU-FP6 project ENSEM-
BLES (<http://ensembles-eu.metoffice.com>) and the data providers in the ECA&D project (<http://www.ecad.eu>),
ERA-Interim data provided courtesy ECMWF, and the WATCH-Forcing-Data-ERA-Interim (~~WFDEI~~) ([WFDEI, a product of the EU-EMBRACE pr](#)

S.S. is grateful to the German National Academic Foundation (Studienstiftung des Deutschen Volkes) for
support and the International Max Planck Research School for Global Biogeochemical Cycles (IMPRS-gBGC)
600 for training.

References

- Allen, M.: Liability for climate change, *Nature*, 421, 891–892, 2003.
- Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., and García-Herrera, R.: The hot summer of 2010: redrawing the temperature record map of Europe, *Science*, 332, 220–224, 2011.
- 605 Beer, C., Weber, U., Tomelleri, E., Carvalhais, N., Mahecha, M., and Reichstein, M.: Harmonized European Long-Term Climate Data for Assessing the Effect of Changing Temporal Variability on Land–Atmosphere CO₂ Fluxes*, *J. Clim.*, 27, 4815–4834, 2014.
- Bellprat, O., Kotlarski, S., Lüthi, D., and Schär, C.: Physical constraints for temperature biases in climate models, *Geophys. Res. Lett.*, 40, 4042–4047, 2013.
- 610 Bindoff, N. L., Stott, P. A., AchutaRao, M., Allen, M. R., Gillett, N., Gutzler, D., Hansingo, K., Hegerl, G., Hu, Y., Jain, S., et al.: Detection and attribution of climate change: from global to regional, in: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.
- 615 Boberg, F. and Christensen, J. H.: Overestimation of Mediterranean summer temperature projections due to model deficiencies, *Nat. Clim. Change*, 2, 433–436, 2012.
- Bondeau, A., Smith, P. C., Zaehle, S., Schaphoff, S., Lucht, W., Cramer, W., Gerten, D., LOTZE-CAMPEN, H., Müller, C., Reichstein, M., et al.: Modelling the role of agriculture for the 20th century global terrestrial carbon balance, *Glob. Chang. Biol.*, 13, 679–706, 2007.
- 620 Buser, C. M., Künsch, H., Lüthi, D., Wild, M., and Schär, C.: Bayesian multi-model projection of climate: bias assumptions and interannual variability, *Clim. Dyn.*, 33, 849–868, 2009.
- Chapin III, F. S., Woodwell, G. M., Randerson, J. T., Rastetter, E. B., Lovett, G. M., Baldocchi, D. D., Clark, D. A., Harmon, M. E., Schimel, D. S., Valentini, R., et al.: Reconciling carbon-cycle concepts, terminology, and methods, *Ecosystems*, 9, 1041–1050, 2006.
- 625 Christensen, J., Boberg, F., Christensen, O., and Lucas-Picher, P.: On the need for bias correction of regional climate change projections of temperature and precipitation, *Geophys. Res. Lett.*, 35, 2008.
- Christensen, J., Kjellström, E., Giorgi, F., Lenderink, G., and Rummukainen, M.: Weight assignment in regional climate models, *Clim. Res.*, 44, 179–194, 2010.
- Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogee, J., Allard, V., Aubinet, M., Buchmann, N., Bernhofer, C., Carrara, A., et al.: Europe-wide reduction in primary productivity caused by the heat and drought in 2003, *Nature*, 437, 529–533, 2005.
- 630 Coles, S.: *An introduction to statistical modeling of extreme values*, vol. 208, Springer, London, UK, 2001.
- Collins, M.: Ensembles and probabilities: a new era in the prediction of climate change, *Philos. Trans. R. Soc., A*, 365, 1957–1970, 2007.
- 635 Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Quart. J. Roy. Meteor. Soc.*, 137, 553–597, 2011.
- Donat, M., Alexander, L., Yang, H., Durre, I., Vose, R., Dunn, R., Willett, K., Aguilar, E., Brunet, M., Caesar, J., et al.: Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset, *J. Geophys. Res.: Atmos.*, 118, 2098–2118, 2013.
- 640

- Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions "Should we apply bias correction to global and regional climate model data?", *Hydrol. Earth Syst. Sci.*, 16, 3391–3404, 2012.
- Forkel, M., Carvalhais, N., Schaphoff, S., Migliavacca, M., Thurner, M., Thonicke, K., et al.: Identifying environmental controls on vegetation greenness phenology through model–data integration, *Biogeosciences*, 11, 7025–7050, 2014.
- 645 Forkel, M., Migliavacca, M., Thonicke, K., Reichstein, M., Schaphoff, S., Weber, U., and Carvalhais, N.: Co-dominant water control on global inter-annual variability and trends in land surface phenology and greenness, *Glob. Chang. Biol.*, 21, 3414–3435, 2015.
- Frank, D., Reichstein, M., Bahn, M., Thonicke, K., Frank, D., Mahecha, M. D., Smith, P., Velde, M., Vicca, S., Babst, F., et al.: Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts, *Glob. Chang. Biol.*, 21, 2861–2880, 2015.
- 650 Fritsch, F. N. and Carlson, R. E.: Monotone piecewise cubic interpolation, *SIAM J. Numer. Anal.*, 17, 238–246, 1980.
- Gerten, D., Schaphoff, S., Haberlandt, U., Lucht, W., and Sitch, S.: Terrestrial vegetation and water balance-hydrological evaluation of a dynamic global vegetation model, *J. Hydrol.*, 286, 249–270, 2004.
- 655 Giorgi, F., Jones, C., Asrar, G. R., et al.: Addressing climate information needs at the regional level: the CORDEX framework, *World Meteorological Organization (WMO) Bulletin*, 58, 175, 2009.
- Gudmundsson, L., Bremnes, J., Haugen, J., and Skaugen, T. E.: Technical Note: Downscaling RCM precipitation to the station scale using quantile mapping—a comparison of methods, *Hydrol. Earth Syst. Sci.*, 9, 6185–6201, 2012.
- 660 Haerter, J., Hagemann, S., Moseley, C., and Piani, C.: Climate model bias correction and the role of timescales, *Hydrol. Earth Syst. Sci.*, 15, 1065–1079, 2011.
- Harris, I., Jones, P., Osborn, T., and Lister, D.: Updated high-resolution grids of monthly climatic observations—the CRU TS3. 10 Dataset, *Int. J. Climatol.*, 34, 623–642, 2014.
- 665 Haxeltine, A. and Prentice, I. C.: BIOME3: An equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types, *Global Biogeochem. Cycles*, 10, 693–709, 1996.
- Haylock, M., Hofstra, N., Klein Tank, A., Klok, E., Jones, P., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, *J. Geophys. Res.: Atmos.*, 113, 2008.
- 670 Hempel, S., Frieler, K., Warszawski, L., Schewe, J., and Piontek, F.: A trend-preserving bias correction—the ISI-MIP approach, *Earth Syst. Dynam.*, 4, 219–236, 2013.
- Hoffmann, H. and Rath, T.: Meteorologically consistent bias correction of climate time series for agricultural models, *Theor. Appl. Climatol.*, 110, 129–141, 2012.
- 675 IPCC: Summary for Policymakers, in: *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change*, edited by Field, C., Barros, V., Stocker, T., Dahe, Q., Dokken, D., Ebi, K., Mastrandrea, M., Mach, K., Plattner, G., Allen, S., Tignor, M., and Midgley, P., Cambridge University Press, 2012.
- Jones, P. W.: First- and second-order conservative remapping schemes for grids in spherical coordinates, *Mon. Weather Rev.*, 127, 2204–2210, 1999.
- 680

- Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., et al.: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, *J. Geophys. Res.*, G: Biogeosci., 116, 2011.
- 685 Katz, R. W. and Brown, B. G.: Extreme events in a changing climate: variability is more important than averages, *Clim. Chang.*, 21, 289–302, 1992.
- Knutti, R.: The end of model democracy?, *Clim. Chang.*, 102, 395–404, 2010.
- Leonard, M., Westra, S., Phatak, A., Lambert, M., van den Hurk, B., McInnes, K., Risbey, J., Schuster, S., Jakob, D., and Stafford-Smith, M.: A compound event framework for understanding extreme impacts, *Wiley*
- 690 *Interdiscip. Rev. Clim. Change*, 5, 113–128, 2014.
- Lewis, S. L., Brando, P. M., Phillips, O. L., van der Heijden, G. M., and Nepstad, D.: The 2010 amazon drought, *Science*, 331, 554–554, 2011.
- Li, C., Sinha, E., Horton, D. E., Diffenbaugh, N. S., and Michalak, A. M.: Joint bias correction of temperature and precipitation in climate model simulations, *J. Geophys. Res.: Atmos.*, 119, 13–153, 2014.
- 695 Maraun, D.: Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums, *Geophys. Res. Lett.*, 39, 2012.
- Maraun, D.: Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue, *J. Clim.*, 26, 2137–2143, 2013.
- Maraun, D., Wetterhall, F., Ireson, A., Chandler, R., Kendon, E., Widmann, M., Brienen, S., Rust, H., Sauter,
- 700 T., Themeßl, M., et al.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, *Rev. Geophys.*, 48, 2010.
- Massey, N., Jones, R., Otto, F., Aina, T., Wilson, S., Murphy, J., Hassell, D., Yamazaki, Y., and Allen, M.: weather@home-development and validation of a very large ensemble modelling system for probabilistic event attribution, *Quart. J. Roy. Meteor. Soc.*, 2014.
- 705 Mueller, B. and Seneviratne, S.: Systematic land climate and evapotranspiration biases in CMIP5 simulations, *Geophys. Res. Lett.*, 41, 128–134, 2014.
- Otto, F., Massey, N., Oldenborgh, G., Jones, R., and Allen, M.: Reconciling two approaches to attribution of the 2010 Russian heat wave, *Geophys. Res. Lett.*, 39, 2012.
- Piani, C. and Haerter, J.: Two dimensional bias correction of temperature and precipitation copulas in climate
- 710 models, *Geophys. Res. Lett.*, 39, 2012.
- Piani, C., Frame, D., Stainforth, D., and Allen, M.: Constraints on climate change from a multi-thousand member ensemble of simulations, *Geophys. Res. Lett.*, 32, 2005.
- Piani, C., Weedon, G., Best, M., Gomes, S., Viterbo, P., Hagemann, S., and Haerter, J.: Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models, *J. Hydrol.*,
- 715 395, 199–215, 2010.
- Pope, V., Gallani, M., Rowntree, P., and Stratton, R.: The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3, *Clim. Dyn.*, 16, 123–146, 2000.
- Rammig, A., Wiedermann, M., Donges, J., Babst, F., von Bloh, W., Frank, D., Thonicke, K., and Mahecha, M.: Coincidences of climate extremes and anomalous vegetation responses: comparing tree ring patterns to
- 720 simulated productivity, *Biogeosciences*, 12, 373–385, 2014.

- Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M., Seneviratne, S., Zscheischler, J., Beer, C., Buchmann, N., and Frank, D.: Climate extremes and the carbon cycle, *Nature*, 500, 287–295, 2013.
- Reyer, C. P., Leuzinger, S., Rammig, A., Wolf, A., Bartholomeus, R. P., Bonfante, A., de Lorenzi, F., Dury, M., Gloning, P., Abou Jaoudé, R., et al.: A plant’s perspective of extremes: terrestrial plant responses to changing climatic variability, *Glob. Chang. Biol.*, 19, 75–89, 2013.
- Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., Rosenfeld, A., Wurtele, J., Curry, J., Wickham, C., and Mosher, S.: Berkeley earth temperature averaging process, *Geoinfor. Geostat.: An Overview*, 1, 1–13, 2013.
- Rolinski, S., Rammig, A., Walz, A., von Bloh, W., Van Oijen, M., and Thonicke, K.: A probabilistic risk assessment for the vulnerability of the European carbon cycle to weather extremes: the ecosystem perspective, *Biogeosciences*, 12, 1813–1831, 2015.
- Schaphoff, S., Heyder, U., Ostberg, S., Gerten, D., Heinke, J., and Lucht, W.: Contribution of permafrost soils to the global carbon budget, *Environ. Res. Lett.*, 8, 014026, 2013.
- Schiermeier, Q.: Extreme measures, *Nature*, 477, 131–132, 2011.
- Schneider, U., Becker, A., Meyer-Christoffer, A., Ziese, M., and Rudolf, B.: Global Precipitation Analysis Products of the GPCC., Tech. rep., Global Precipitation Climatology Centre (GPCC), Deutscher Wetterdienst, Offenbach a. M., Germany, 2011.
- Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., and Rudolf, B.: GPCC’s new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle, *Theor. Appl. Climatol.*, 115, 15–40, 2014.
- Schulze, E.-D.: Biological control of the terrestrial carbon sink, *Biogeosciences*, 3, 147–166, 2006.
- Seneviratne, S. I., Lüthi, D., Litschi, M., and Schär, C.: Land–atmosphere coupling and climate change in Europe, *Nature*, 443, 205–209, 2006.
- Seneviratne, S. I., Donat, M. G., Mueller, B., and Alexander, L. V.: No pause in the increase of hot temperature extremes, *Nat. Clim. Change*, 4, 161–163, 2014.
- Sheather, S. J. and Jones, M. C.: A reliable data-based bandwidth selection method for kernel density estimation, *J. R. Stat. Soc. Series B Stat. Methodol.*, 53, 683–690, 1991.
- Sillmann, J., Kharin, V., Zwiers, F., Zhang, X., and Bronaugh, D.: Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections, *J. Geophys. Res.: Atmos.*, 118, 2473–2493, 2013.
- Sippel, S. and Otto, F.: Beyond climatological extremes—assessing how the odds of hydrometeorological extreme events in South-East Europe change in a warming climate, *Clim. Chang.*, 125, 381–398, 2014.
- Sippel, S., Mitchell, D., Black, M. T., Dittus, A. J., Harrington, L., Schaller, N., and Otto, F. E.: Combining large model ensembles with extreme value statistics to improve attribution statements of rare events, *Weather Clim. Extrem.*, 9, 25–35, 2015a.
- Sippel, S., Walton, P., and Otto, F. E.: Stakeholder Perspectives on the Attribution of Extreme Weather Events: An Explorative Enquiry, *Weather Clim. Soc.*, 7, 224–237, 2015b.
- Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J., Levis, S., Lucht, W., Sykes, M. T., et al.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, *Glob. Chang. Biol.*, 9, 161–185, 2003.
- Stone, D. and Allen, M.: The end-to-end attribution problem: from emissions to impacts, *Clim. Chang.*, 71, 303–318, 2005.

- Stone, D., Auffhammer, M., Carey, M., Hansen, G., Huggel, C., Cramer, W., Lobell, D., Molau, U., Solow, A., Tibig, L., et al.: The challenge to detect and attribute effects of climate change on human and natural systems, *Clim. Chang.*, 121, 381–395, 2013.
- 765 Stone, D. A., Allen, M. R., Stott, P. A., Pall, P., Min, S.-K., Nozawa, T., and Yukimoto, S.: The detection and attribution of human influence on climate, *Ann. Rev. Environ. Resour.*, 34, 1–16, 2009.
- Stott, P. A. and Walton, P.: Attribution of climate-related events: understanding stakeholder needs, *Weather*, 68, 274–279, 2013.
- Stott, P. A., Allen, M., Christidis, N., Dole, R. M., Hoerling, M., Huntingford, C., Pall, P., Perlwitz, J., and Stone, D.: Attribution of weather and climate-related events, in: *Climate Science for Serving Society*, pp. 770 307–337, Springer, Dordrecht Heidelberg New York London, 2013.
- Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, *J. Hydrol.*, 456, 12–29, 2012.
- Van Oijen, M., Balkovic, J., Beer, C., Cameron, D., Ciais, P., Cramer, W., Kato, T., Kuhnert, M., Martin, R., Myneni, R., et al.: Impact of droughts on the carbon cycle in European vegetation: a probabilistic risk analysis using six vegetation models, *Biogeosciences*, 11, 6357–6375, 2014.
- 775 Vautard, R., Gobiet, A., Jacob, D., Belda, M., Colette, A., Déqué, M., Fernández, J., García-Díez, M., Goergen, K., Güttler, I., et al.: The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project, *Clim. Dyn.*, 41, 2555–2575, 2013.
- von Buttler, J., Zscheischler, J., and Mahecha, M. D.: An extended approach for spatiotemporal gapfilling: 780 dealing with large and systematic gaps in geoscientific datasets, *Nonlinear Proc. Geoph.*, 21, 203–215, 2014.
- Wang, C., Zhang, L., Lee, S.-K., Wu, L., and Mechoso, C. R.: A global perspective on CMIP5 climate model biases, *Nat. Clim. Change*, 4, 201–205, 2014.
- Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., and Schewe, J.: The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): Project framework, *Proc. Natl. Acad. Sci. U.S.A.*, 111, 3228–785 3232, 2014.
- Weedon, G., Gomes, S., Viterbo, P., Shuttleworth, W., Blyth, E., Österle, H., Adam, J., Bellouin, N., Boucher, O., and Best, M.: Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century, *J. Hydrometeorol.*, 12, 823–848, 2011.
- Weedon, G., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological 790 forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, *Water Resour. Res.*, 50, 7505–7514, 2014.
- Wood, A. W., Leung, L. R., Sridhar, V., and Lettenmaier, D.: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, *Clim. Chang.*, 62, 189–216, 2004.
- Zhang, X., Hegerl, G., Seneviratne, S., Stewart, R., Zwiers, F., and Alexander, L.: WCRP Grand Challenge: 795 Understanding and Predicting Weather and Climate Extremes, Tech. rep., World Climate Research Program, http://www.wcrp-climate.org/images/documents/grand_challenges/GC_Extremes_v2.pdf, White Paper, 2014.
- Zscheischler, J., Mahecha, M. D., von Buttler, J., Harmeling, S., Jung, M., Rammig, A., Randerson, J. T., Schölkopf, B., Seneviratne, S. I., Tomelleri, E., et al.: A few extreme events dominate global interannual 800 variability in gross primary production, *Environ. Res. Lett.*, 9, 035 001, 2014a.

Zscheischler, J., Reichstein, M., Harmeling, S., Rammig, A., Tomelleri, E., and Mahecha, M. D.: Extreme events in gross primary production: a characterization across continents, *Biogeosciences*, 11, 2909–2924, 2014b.

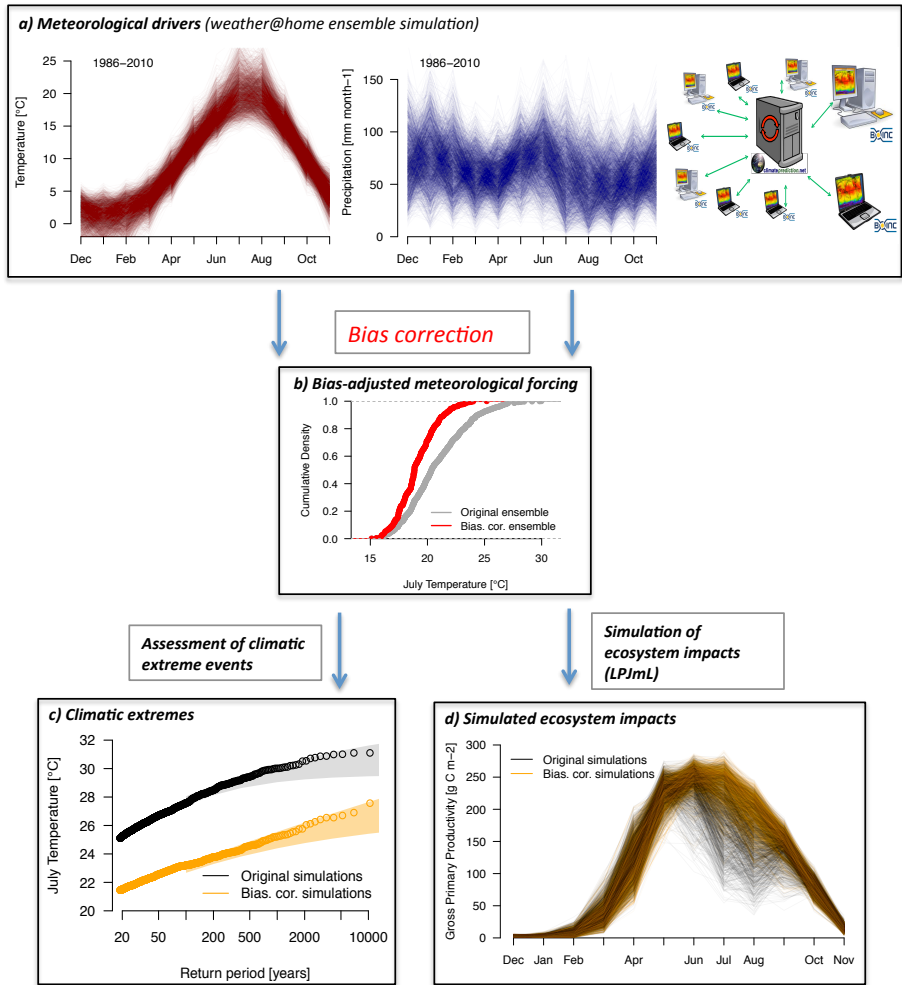


Figure 1. Methodological workflow of the study. (a) Generation of regional climate model simulations using a large ensemble modelling framework (*climateprediction.net/weatherathome*). (b) Adjustment of biases in the regional climate model’s output. (c) Assessment of weather and climate extreme events. (d) Ensemble simulation of ecosystem-atmosphere fluxes of carbon and water using the LPJmL model.

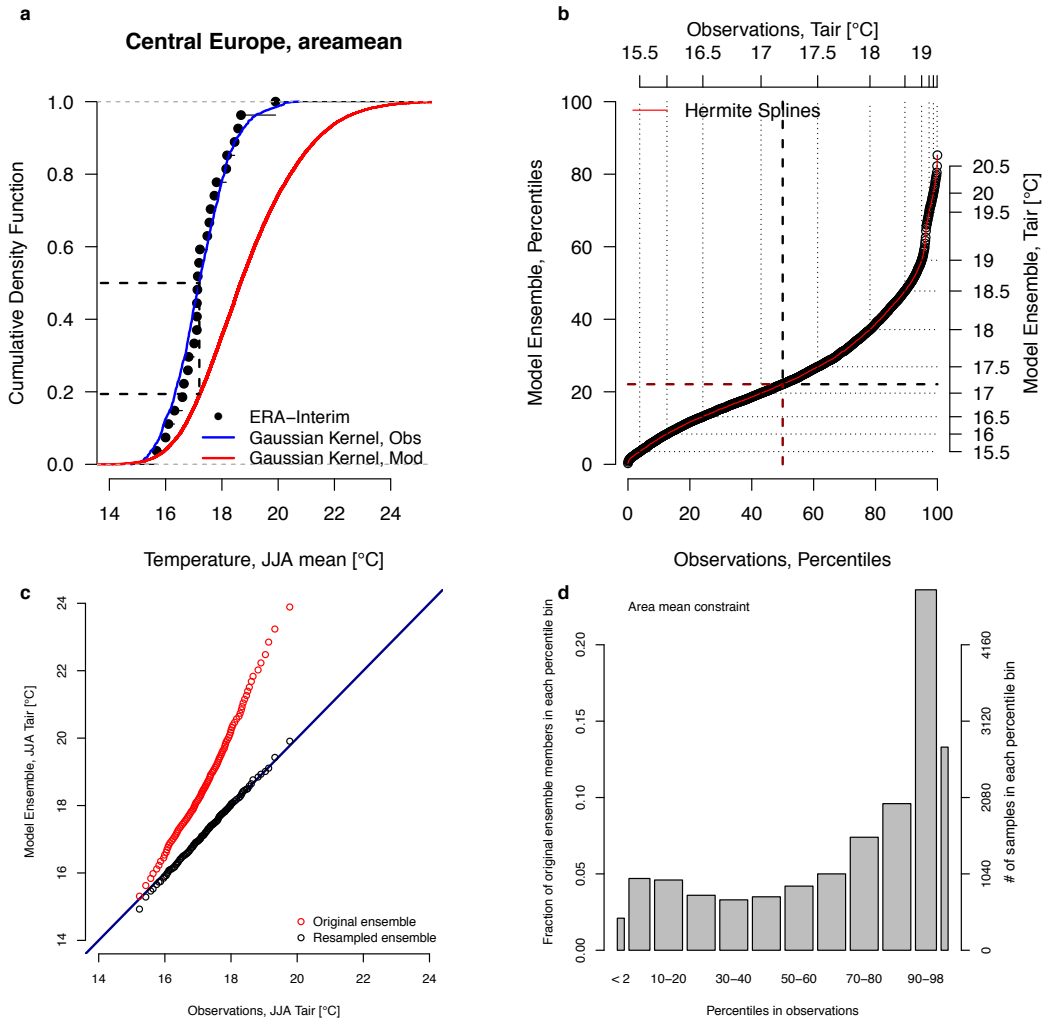


Figure 2. Illustration of ensemble-based resampling methodology. (a) Empirical cumulative density function of JJA mean temperatures over Central Europe in ERA-Interim. The non-parametric fit to the cumulative density using a Gaussian kernel for observations and the model ensemble are shown by the blue and red lines, respectively. (b) A transfer function between the observed and modelled distribution is derived using Cubic Hermite splines. (c) [Quantile-quantile plot for the original and resampled ensemble for the JJA temperature constraint.](#) (d) Fraction of original ensemble members in percentile bins of the observed distribution (blue line in (a)), i.e. ‘effective ensemble size’ after resampling.

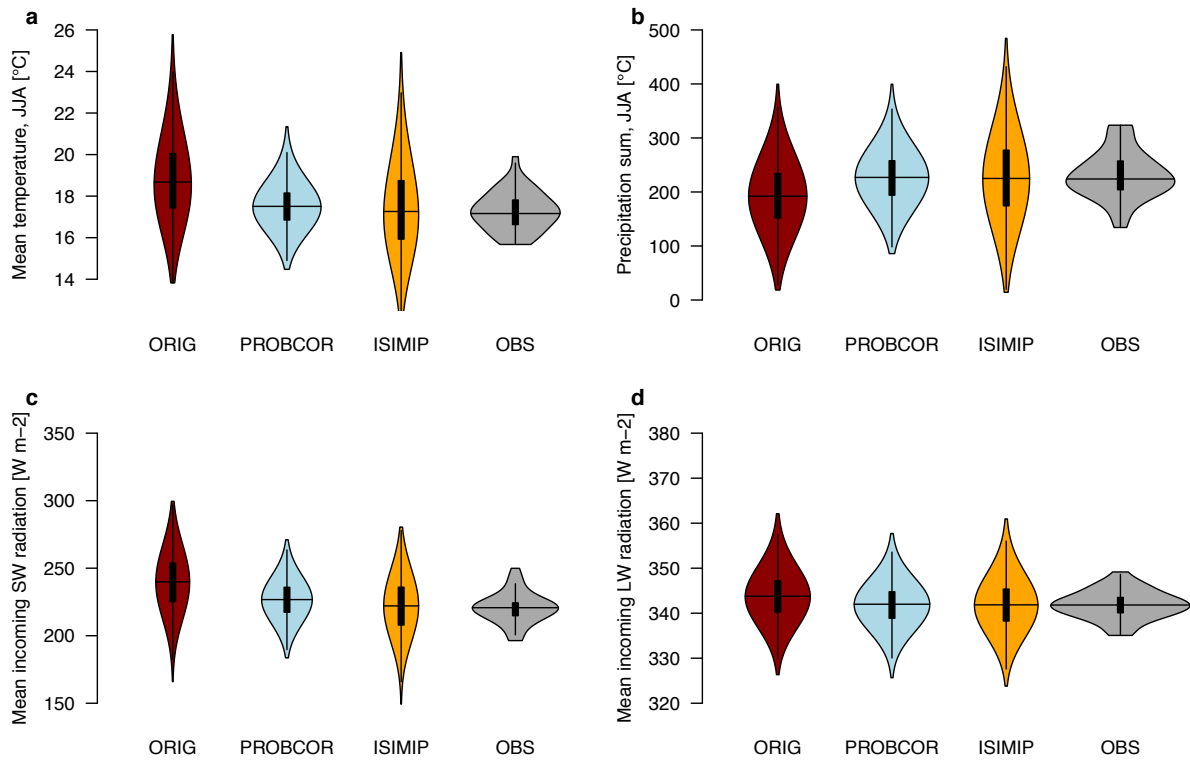


Figure 3. Evaluation of the resampling bias correction methodology for the study area in Central Europe for (a) temperature, (b) precipitation, (c) incoming short-wave radiation, and (d) incoming long-wave radiation. Both sides of each violin are constructed as rotated, equal-area kernel density estimates, and a standard boxplot is drawn inside each violin.

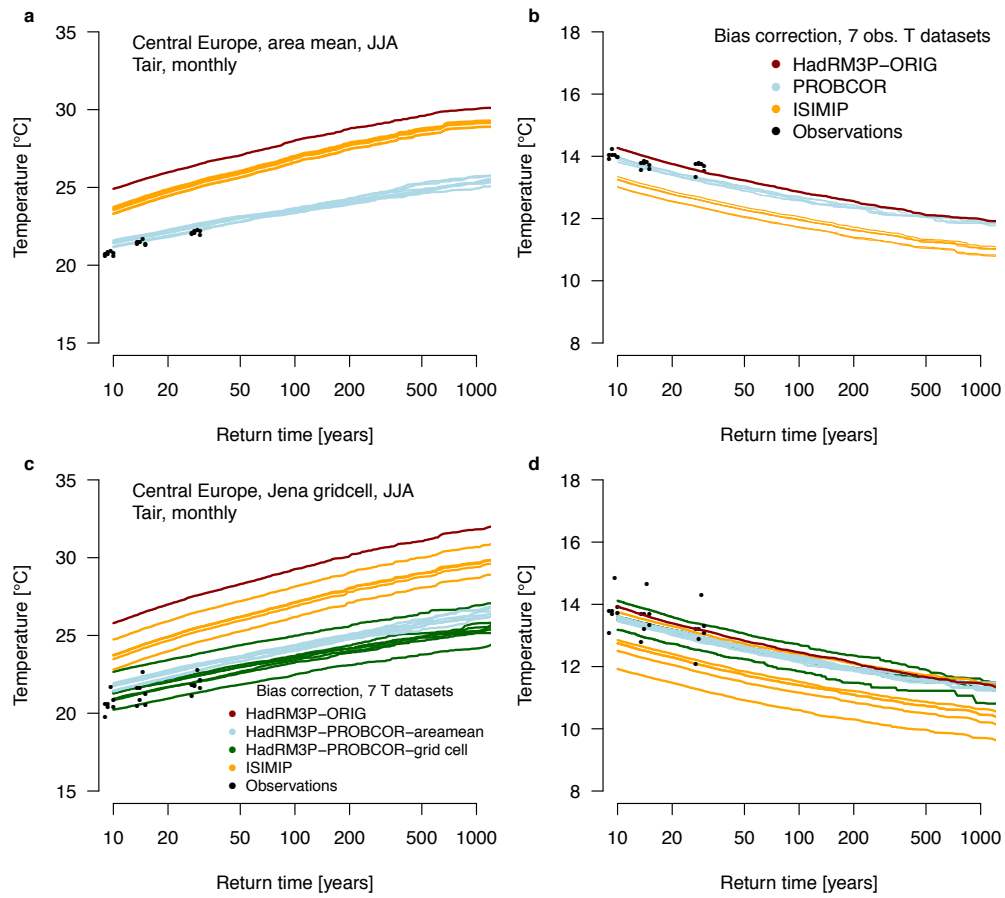


Figure 4. Return times of hot (a,c) and cold (b,d) temperature extremes in summer (JJA) in the original regional model simulations ('ORIG'), in the resampled ensemble ('PROBCOR') and the mean-adjusted ensemble ('ISIMIP'). Plots are shown as spatial averages over Central Europe (top panels) and for an illustrative grid cell (Jena, bottom panels). Black dots in each plot indicate empirical return times estimated from observations taken from 7 different datasets that were used for bias correction.

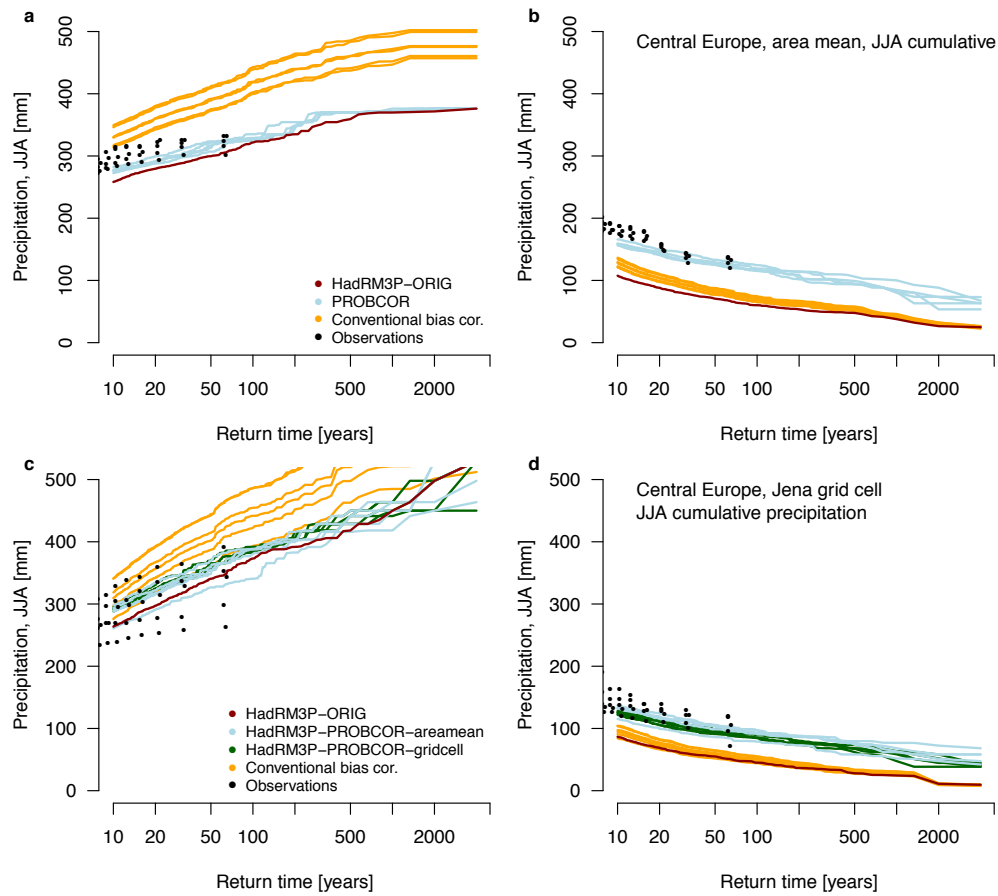


Figure 5. Return times of wet (a,c) and dry (c,d) rainfall extremes in summer (JJA) in the original regional model simulations ('ORIG'), in the resampled ensemble ('PROBCOR') and the mean-adjusted ensemble ('ISIMIP'). Plots are shown as spatial averages over Central Europe (top panels) and for an illustrative grid cell ('Jena pixel', bottom panels). Black dots in each plot indicate empirical return times estimated from observations taken from 7 different datasets that were used for bias correction.

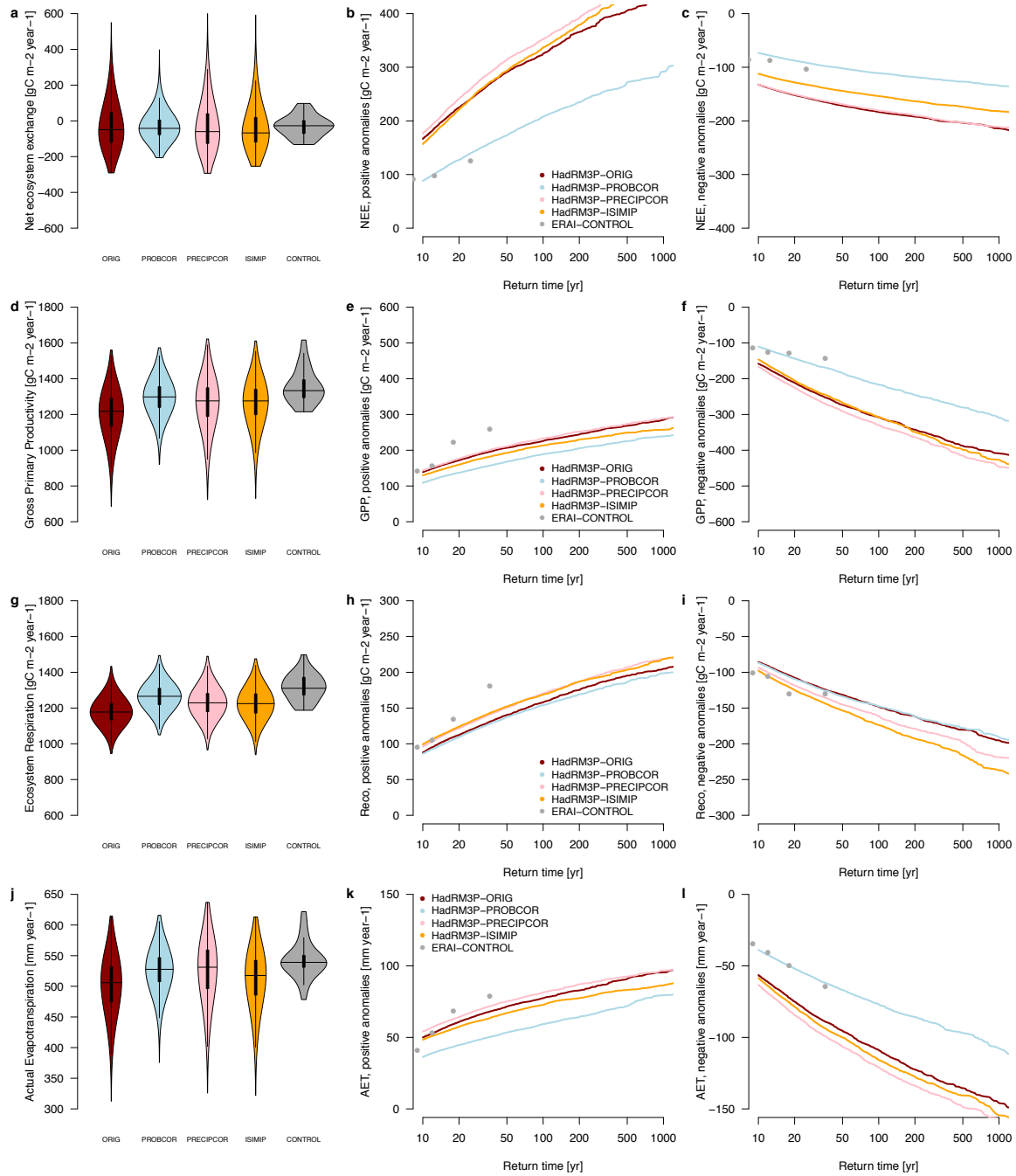


Figure 6. LPJmL simulated distributions of ecosystem-atmosphere carbon and water fluxes for Central European natural vegetation for each bias correction scheme. Each row shows the simulated distribution and the upper and lower tail of NEE (a,b,c), GPP (d,e,f), Reco (g,h,i) and AET (j,k,l), respectively. [\(a,d,g,j\) Both sides of each violin are constructed as rotated, equal-area kernel density estimates, and a standard boxplot is drawn inside each violin.](#)

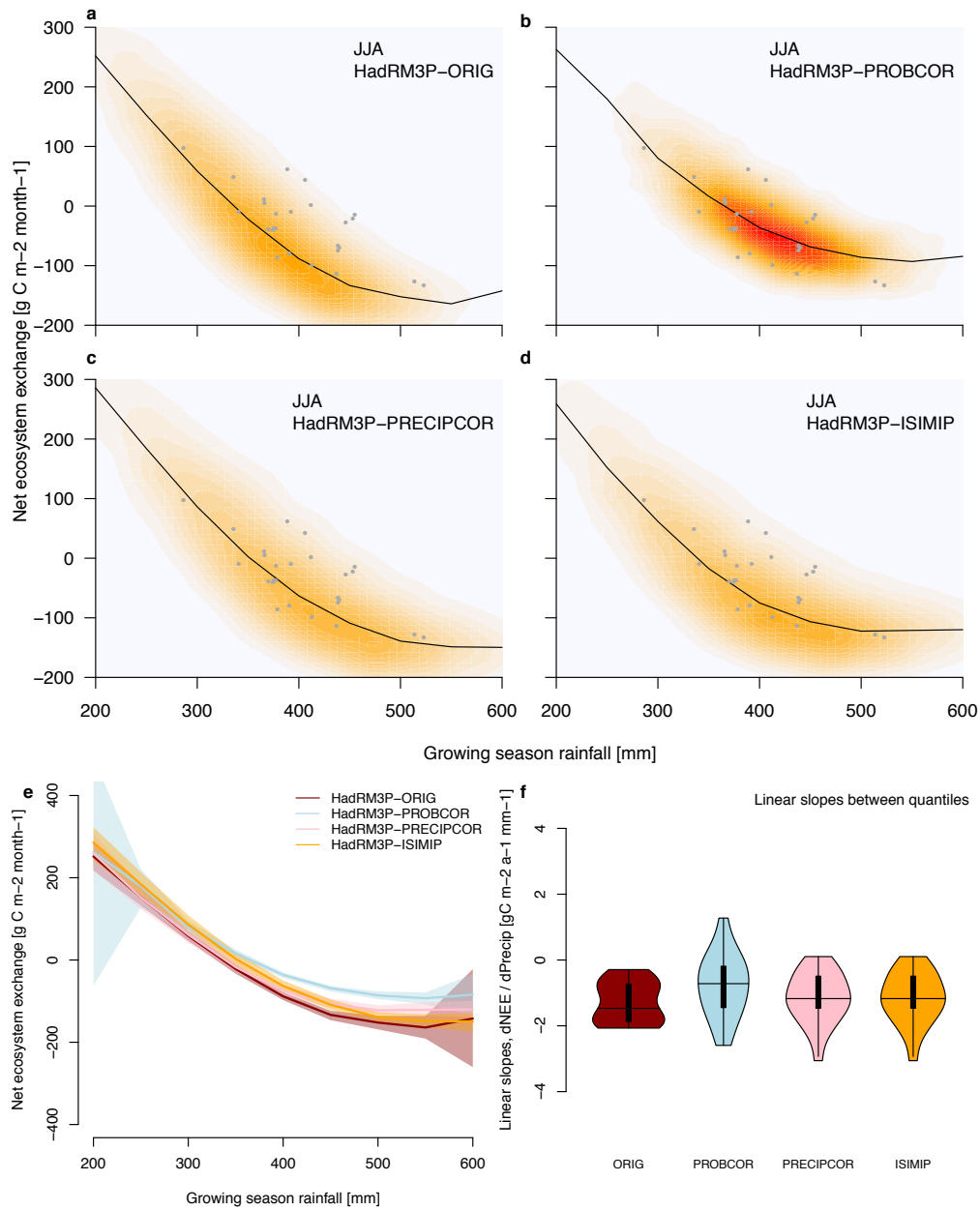


Figure 7. (a-d) Kernel density plots of the sensitivity of simulated annual NEE to growing season rainfall in LPJmL under four different bias correction schemes. Grey dots denote ERA-Interim control simulations in each plot, black lines indicate piecewise linear regressions. (e) Piecewise linear regression relations for each bias correction scheme. Shaded colours indicate confidence intervals (5-95th percentile of piecewise linear regression derived by bootstrapping). (f) Distribution of linear regression slopes ($\text{dNEE} / \text{dRainfall}$) between regularly spaced quantiles of the rainfall distribution for each bias correction scheme, [shown as violin plots](#).

Table 1. Datasets used for bias correction and evaluation.

Name of dataset	Climate variables	Resolution	Domain & Orig.	Provider & Reference
Berkeley Earth Observations (gridded experimental)	Tair	Europe, monthly, 1850-2012	0.25°, 1850-2012	http://www.berkeleyearth.org , Rohde et al. (2013)
Climate Research Unit (CRU), High-resolution datasets	Tair, Precip.	Global, monthly, 1901-2012	0.5°, 1901-2012	Climate Research Unit, http://www.cru.uea.ac.uk/cru/data/hrg/ , Harris et al. (2014)
CRUNCEP	Tair, Precip., SWdown, LWdown	Global, 1948-2012	0.5°, daily, 1948-2012	http://dods.extra.cea.fr/data/p529viov/cruncep/readme.htm
Global Precipitation Climatology Centre monthly precipitation (GPCC) E-OBS gridded dataset	Precip.	Global, monthly, 1901-2012	0.5°, 1901-2012	Global Precipitation Climatology Center (GPCC), http://gpcc.dwd.de/ , Schneider et al. (2014)
ERA-Interim, Version 2 (ERA-I)	Tair, Precip., SWdown, LWdown, LE	Global, 1979-2014	0.5°, 6-hourly, 1979-2014	European Centre for Medium Range Weather Forecasts (ECMWF), http://apps.ecmwf.int/datasets/data/interim-full-daily/ , Dee et al. (2011)
Model Tree Ensembles	LE	Global, monthly, 1982-2011	0.5°, 1982-2011	MPI Biogeochemistry Jena, Jung et al. (2011)
WATCH-harmonized (WFD-harmonized)	Tair, Precip., SWdown, LWdown	Europe, 1901-2012	0.5°, daily, 1901-2012	MPI Biogeochemistry Jena, Weedon et al. (2011); Beer et al. (2014)
WATCH (WFDEI)	Tair, Precip., SWdown, LWdown	Global, 1979-2012	0.5°, daily, 1979-2012	EU-WATCH, Weedon et al. (2011) ; Weedon et al. (2014)

Table 2. Annual mean ecosystem-atmosphere water and carbon fluxes simulated by LPJml.

Bias Correction Method	NEE (g C m ⁻² a ⁻¹)	GPP (g C m ⁻² a ⁻¹)	Reco (g C m ⁻² a ⁻¹)	ET (mm a ⁻¹)
HadRM3P-ORIG	-26.5	1206.4	1179.9	501.5
HadRM3P-PROBCOR	-30.3	1295.8	1265.5	525.9
HadRM3P-ISIMIP	-31.6	1262.3	1230.7	525.2
HadRM3P-PRECIPCOR	-38.2	1263.2	1225.0	511.2
ERA-INTERIM-CONTROL	-28.4	1353.3	1324.8	536.7