Towards process-informed bias correction of climate change simulations

Article

Accepted Version

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutierrez, J. M., Hagemann, S., Richter, I., Soares, P. M. M., Hall, A. and Mearns, L. O. (2017) Towards process-informed bias correction of climate change simulations. Nature Climate Change, 7. pp. 764-773. ISSN 1758-6798 doi: https://doi.org/10.1038/NCLIMATE3418 Available at http://centaur.reading.ac.uk/73495/

To link to this article DOI: http://dx.doi.org/10.1038/NCLIMATE3418

Publisher: Nature Publishing Group

Towards process-informed bias correction of climate change simulations

Douglas Maraun[1], Theodore G. Shepherd[2], Martin Widmann[3], Giuseppe Zappa[2], Daniel Walton[4], Jose M. Gutierrez[5], Stefan Hagemann[6], Ingo Richter[7], Pedro M.M. Soares[8], Alex Hall[4], Linda O. Mearns[9]

[1] University of Graz, Austria
[2] University of Reading, UK
[3] University of Birmingham, UK
[4] University of California, Los Angeles, USA
[5] Institute of Physics of Cantabria (IFCA), Santander, Spain
[6] Max Planck Institute for Meteorology, Hamburg, Germany
[7] Japan-Agency for Marine-Earth Science and Technology (JAMSTEC), Yokohama, Japan
[8] Instituto Dom Luiz, University of Lisbon, Portugal
[9] National Center for Atmospheric Research (NCAR), Boulder, USA

September 15, 2017

Biases in climate model simulations introduce biases in subsequent impact simulations. Therefore, bias correction methods are operationally used to post-process regional climate projections. However many problems have been identified, and some researchers question the very basis of the approach. Here we demonstrate that a typical cross-validation is unable to identify improper use of bias correction. Several examples show the limited ability of bias correction to correct and to downscale variability, and demonstrate that bias correction can cause implausible climate change signals. Bias correction cannot overcome major model errors, and naive application might result in ill-informed adaptation decisions. We conclude with a list of recommendations and suggestions for future research to reduce, post-process, and cope with climate model biases.

Climate scientists are confronted with a growing pressure to support adaptation decisions and face the dilemma of operationalising what is still foundational research[1,2]. The models often used to inform adaptation decisions - global coupled atmosphere ocean general circulation models (GCMs), potentially downscaled with regional climate models (RCMs) - have horizontal resolutions often far coarser than those demanded, and suffer from substantial biases[3,4]. To reduce biases and to overcome the scale gap between the numerical model grid and the desired scale, climate model output is almost routinely post-processed by bias correction (often called bias adjustment) methods. A vast number of bias corrected national and global climate change projections has been published[5,6,7,8,9,10,11,12,13], has served as input for impact studies[14,15,10,16] as well as assessment reports[17,18,19], and has been made available through data portals[20,21,13]. A wide variety of bias correction methods is in use, ranging from simple adjustments of the mean to flexible, potentially multivariate, quantile mapping approaches[22,23,24]. Yet many problems related to bias correction have been identified[25,8,26,27,28,29]. Thus, even though bias correction is often considered a necessary step in climate impact modelling[24], the approach is prone to misuse and best practice still needs to be established[30]. Some authors even question the very basis of bias correction[31].

Current developments on bias correction have largely focused on improving statistical methodology: to better match variability and extremes[24,32,33,34], the dependence between different climatic variables[35,36], the location of features[37], or to retain simulated trends[6,32,11]. This focus has ignored a major issue: a key requirement of climate model projections is credibility[38,1,2]. Here, we argue that current bias correction methods might improve the applicability of climate simulations, but in general cannot improve low model credibility. Indeed, bias correction may hide a lack of credibility or may even reduce credibility. The way bias correction is often applied and evaluated might ultimately lead to ill-informed adaptation decisions.

We start from the basic reason underlying the demand to bias correct: all models are substantial simplifications of a real system. Climate models are based on physical laws such as conservation of energy, mass and momentum, thermodynamic and radiation laws. But models have a limited spatial resolution, their topography is coarse and they will never resolve nor represent all relevant processes from planetary waves down to turbulence. Sub-grid processes are simplified by parameterisations. As a consequence, many relevant atmospheric, oceanic and coupled processes are not realistically represented, with knock-on effects on other processes even far away from where the primary biases occur[39]. Biases in basic quantities such as mean and variance are therefore commonplace, even for something as fundamental as global-mean surface temperature[3]. Often, a realistic behaviour is only achieved by tuning the model[3]. In short, climate model biases are severe enough to in principle justify the use of bias correction techniques to render model output more useful for impact studies.

We therefore argue that bias correction should not be dismissed, but that a solid conceptual and process understanding of climate model biases is required to successfully apply bias correction. The extent to which biases can be mitigated by post-processing depends on their origin. We present several examples, discuss their correctability by state-of-the-art bias correction methods, and propose alternative approaches and future directions of research.

1 Bias correction in a nutshell

We define a bias as the systematic difference between a modelled property of the climate system and the corresponding real property[40,41,25,31,42,43]. Such properties could be mean temperature, variance or a 100-year return value. The term "systematic" refers to all differences that are not due to sampling uncertainty. Biases are typically assumed to be time-independent[44,45,23,11], but in principle may vary in time[41,40,25,42]. Some authors define a bias as the time independent error component of a model[24,46,47]. The problems we discuss below occur irrespective of the specific bias definition.

As bias correction we consider all methods that calibrate an empirical transfer function between simulated and observed distributional parameters, and apply this transfer function to output simulated by the considered model. Bias correction according to this definition is a mere post-processing.

We focus on two different types of methods which are broadly representative of those commonly used: a simple adjustment of the mean, and quantile mapping. A simple mean bias correction would estimate a bias as the difference (or ratio for, e.g., precipitation) between simulated and observed mean over a reference period, and adjust the simulated time series over a scenario period by the estimated bias (by subtracting it, or rescaling). Quantile mapping individually adjusts each quantile. The transfer functions are then applied to climate change simulations under the assumption that they are time-invariant.

Bias correction relies on observational reference data, which should in many cases be considered a model product themselves. This holds true in particular for gridded data sets. Related issues are an important topic for bias correction, but are outside the scope of this article.

2 The evaluation problem

[Figure 1 about here.]

To begin with, we demonstrate the difficulties to evaluate the performance of bias correction. The evaluation of statistical models, e.g., in weather forecasting, is generally done by cross-validation: the model is calibrated to a subset of the available data only, the evaluation is carried out by assessing the prediction of the remaining (independent) data. Cross-validation is widely used for establishing skill of bias correction, often only for calibrated properties of the marginal distribution[6,47,48,23,49] (some exceptions evaluate temporal or spatial dependence[24,27]). Here we demonstrate that such an evaluation is not suitable to establish bias correction skill.

Consider the rather absurd setting of bias correcting simulated daily temperature from the Southern Ocean against observed daily precipitation over central Europe during boreal winter. The corresponding model grid boxes are simply taken from the exact opposite side of the globe. Whereas the temperature field over the Southern Ocean (mapped onto Europe) is very smooth (Fig. 1a,d), precipitation in Europe has a distinct pattern controlled by distance to sea and orography (b,e). But even though modelled temperature and observed precipitation fields are essentially unrelated and both fields show different long-term changes, the quantile mapping looks reasonable for the validation period, for mean and high values (c,f). The residual bias (g) between corrected model and observation purely stems from the different trends in both regions. The problem is especially severe for non-parametric quantile mapping, as demonstrated for the grid box enclosing Venice (h): even though the temperature and precipitation distributions have completely different shapes, and both distributions change substantially over time (mean precipitation +28%, mean temperature -0.29K in the corresponding Southern Ocean grid box), the QQ plot looks reasonable also for the validation period. In other words: cross-validation of calibrated climatological properties is not able to identify the absurdity of the chosen example, and is thus not sufficient to evaluate the performance of bias correction. The reason for the failure is that, in climate modelling, model and observations are not in synchrony and predictive skill cannot, as in weather forecasting, be established by cross-validation[26]. The evaluation is restricted to long term distributional aspects only, and provided the sampling is adequate, cross-validation will merely reproduce the long-term distribution. But in a non-synchronous setting it is still possible to evaluate non-calibrated aspects, in particular for the temporal and, if required, spatial dependence structure. Such an evaluation would yield essential and indispensable information about the appropriateness of a bias correction.

3 Bias correction under present conditions

[Figure 2 about here.]

Bias correction may introduce artefacts already for present climate conditions which are invisible to an evaluation of marginal distributional properties. As example, consider corrections of the drizzle effect, i.e., the fact that climate models often simulate too high a number of wet days with very low intensities. Quantile mapping adjusts the number of wet days by changing the least wet days into dry days. The adjustment in turn improves the representation of dry spells of typically up to about 20 days[50]. But climate models have considerable deficiencies in representing temporal variability beyond the drizzle effect. Dry spells are often too short, e.g., because the persistence of blocking highs is typically underrepresented[51], or because a dry valley may be represented as an exposed location by a typical climate model with coarse topography. Whereas the drizzle effect may indeed be correctable, an attempt to correct other, more fundamental errors in the spell length distribution may result in unwanted artefacts (Fig. 2). In many cases one may simply miss the long spells (a), in some cases one may by chance even combine short spells into long ones and therefore improve the overall spell length distribution (b). But in a substantial amount of cases, the wet-day adjustment might either produce too many short and medium-length spells (c) or even too long spells (d). This example highlights that bias correction is not a one-size-fits-all approach, but needs to be user-tailored: is the overall wet-day probability relevant or the representation of spell lengths? A careful decision needs to be drawn, and a sensible adjustment carried out. Other examples, where attempts to bias correct temporal structure might cause severely misleading results, are the diurnal cycle of precipitation or the onset of the rainy season[8].

[Figure 3 about here.]

Bias correction may further be infeasible if the climate model variable does not capture the relevant regional processes. Consider a GCM that simulates reasonable ENSO variability, but does not reproduce the clustering of extreme precipitation in Peru during El Niño events (Fig. 3, top and middle left panels). Quantile mapping trivially adjusts the distributions (right panel), but still the result is meaningless as the wrong clustering is not improved (bottom left panel). In this example, already a visual inspection of the resulting time series uncovers the bias correction problem. When evaluating many grid boxes, an evaluation conditional on El Niño events might be required. A similar representativeness problem may be caused by a coarse model topography, which may act as an unrealistically strong meteorological divide[28].

[Figure 4 about here.]

In many cases bias correction is used to downscale to a finer spatial resolution[5,48,49,35,15,12]. Current approaches, however, are unable to generate unexplained subgrid day-to-day variability and may even introduce artefacts, e.g., in the representation of extreme precipitation[27]. But similar effects might also occur for temperature fields in complex terrain. Consider temperature inversions, a common feature in the Central Valley, California (Fig. 4). A bias corrected GCM will trivially reproduce the climatological temperature difference of 2 K between a location in the valley and a nearby location higher up in the Sierra Nevada. But whereas the actual day-to-day temperature difference has a broad distribution - with negative values indicating inversions - the bias corrected difference is essentially constant (it varies slightly because quantile mapping corrects different quantiles individually). Stochastic approaches explicitly modelling unexplained sub-grid variability may thus be required in complex terrain or for highly variable fields.

4 Bias correction under climate change conditions

Some artefacts of bias correction may only appear under changing climatic conditions and may thus be invisible to evaluation against present observations.

One cause of such artefacts are GCMs biases in the large-scale atmospheric circulation[52,53], which themselves result from an insufficient resolution of the atmospheric model[54], a coarse topography[55,56] or from biases in the underlying sea surface temperature[57,58,59]. For instance, over Europe the North Atlantic winter storm track is too zonal in most models and crosses Europe too far south[53]. Such biases exert a strong control on regional climate[26,60]. They are inherited by downscaling and are reflected in regional biases[61].

[Figure 5 about here.]

It has been argued that biases in surface weather resulting from circulation biases cannot be bias corrected[26,30]. For instance, when the frequency of circulation types is misrepresented, bias correction may increase biases for specific circulation types[29]. Here we further show that bias correction in the presence of substantial circulation biases may induce implausible future signals.

Consider precipitation projections based on a GCM with a substantial southward bias of the Atlantic storm track, such that the maximum of present day winter precipitation in Western Europe is shifted southwards by about 20°(Fig. 5 top). The GCM simulates a northward shift of the storm track. A mean bias correction of winter precipitation will perfectly align simulated present-day mean precipitation with observations, by damping precipitation over Southern Europe, and amplifying it over Central and Northern Europe. Applying this correction to the future simulation, however, the northward shift of the uncorrected precipitation peak - indicating a northward shift of the storm track - is transformed into a southward precipitation shift.

In other words: in the presence of major circulation biases, bias correction - even though the local climate change signal is preserved - might create implausible patterns of surface climate change. Such problems can be avoided by a careful climate model selection: for a GCM with a lower circulation bias, the precipitation bias correction preserves the northward precipitation shift consistent with the storm track shift (Fig. 5 right bottom panel).

Two approaches have been suggested to correct atmospheric circulation biases. First, to bias correct GCM fields prior to dynamical downscaling[62]; and second to spatially shift simulated fields[37]. Both approaches, trivially, correct biases in the climatological atmospheric fields. The first approach, however, introduces inconsistencies in the atmospheric dynamics: for instance, individual storms are - in the GCM - still generated at the wrong position of the polar front and then - in the RCM - interact with the corrected climatological polar front. The second approach ignores that the simulated position of circulation features is intricately linked to the model orography, simulated land-sea contrasts and sea surface temperature biases, and thus introduces inconsistencies with these model properties.

Another cause of artefacts is the modification of the climate change signal by variance-adjusting bias correction methods[8,27,63]. A debate has arisen whether these trend modifications might actually improve or deteriorate the raw climate change signal[40,64], and several trend preserving bias correction approaches have been developed[32,11,65,66]. We argue that this issue cannot be resolved based on purely statistical arguments. Again, one needs to refer to process understanding.

Obviously, a credibly simulated trend should not be altered by any postprocessing. In such a case, the assumption of a time invariant correction is fulfilled and a trend preserving bias correction is the method of choice. Often, however, climate model biases depend on the actual state of the climate system[41,25,67], so in a changing climate they are not time-invariant. Two questions arise: first, in what situations are climate model trends implausible? And second, in which situations could bias correction methods like quantile mapping potentially improve such trends?

Many cases have been identified where climate models may simulate implausible changes of large-scale climatic phenomena, because the underlying processes are not realistically represented. Prominent examples are the representation of ENSO feedbacks[68,69], the Indian summer monsoon[70,71,72], the influence of increased diabatic heating on the intensification of extratropical cyclones[73], or European blocking[51]. Current bias correction methods will not succeed in improving these changes, as they result from fundamental climate model errors[30].

At the regional scale, misrepresented land-surface interactions may result in implausible climate change trends. For instance, models simulating unrealistically low summer soil moisture tend to over-represent summer temperature increases[74,75]; similarly the simulated increase of spring temperature is tightly linked to snow-albedo feedback strength[74]. Furthermore trends may be implausible as a result of inadequately parameterised sub-grid processes. For instance, there is evidence that the response of summer convective precipitation extremes to global warming is mis-represented by regional climate models with parameterised convection[76,77].

In such situations, it has been argued that quantile mapping may improve implausible trends[40,64], because its correction is value-dependent: a simulated value of, say, 25°C will be adjusted with a specific correction irrespective of the actual state of the climate system, i.e., in present and future climate. The distributions typically adjusted by quantile mapping are mostly spanned by day-to-day variability, which is mainly caused by the passage of different types of airmasses. Under climate change, the properties of airmasses themselves will change. If a temperature of 25°C corresponds to a rare, sunshiny day in present climate, such a temperature might correspond to an overcast rainy day in a warmer climate. It is thus conceivable that the value dependence of biases found for present day climate[40] might be different in the future. The same reasoning can be made from a time-scale point of view: as bias correction is calibrated on daily time scales, also the modification of the climate change signal stems from the rescaling of modelled day-to-day variability[27,63]. Therefore, a trend modification by quantile mapping can only be sensible if - in a given context - the transfer function calibrated on short time scales can sensibly be applied to correct biases on long time scales.

[Figure 6 about here.]

We illustrate this issue with spring temperature trends in mountaineous terrain. Consider again the example from California (Fig. 6). A GCM misses the complex topography of the region and thus simulates a rather smooth temperature field for present climate (a). Quantile mapping trivially produces the correct present temperature fields (b). Similarly, a high resolution RCM simulates a realistic temperature field (c). The RCM also simulates a plausible climate change signal which varies systematically across topography (f): at high elevations, the warming is amplified by the snow-albedo feedback. The climate change signal of the GCM, however, is again unrealistically smooth (d); no elevation dependent warming is produced. A trend preserving bias correction would fully inherit this implausible climate change signal. Standard quantile mapping modifies the large-scale changes, but in an unsystematic way (e). We do not know whether the RCM simulation is correct, but the preserved and bias corrected GCM signals are highly implausible.

Thus, bias correction is trapped in a fundamental dilemma: in situations where the driving model simulates a credible change a trend preserving bias correction[32,11] is a sensible choice. In many cases, however, we may have strong evidence that the simulated regional climate change is implausible - we would like to improve the change. Standard quantile mapping modifies simulated trends. But as discussed above and demonstrated for the snow albedo feedback, we know that these modifications may not be physically justified. Here, one would have to assess the raw and modified changes on a case-by-case basis, referring to the relevant climatic processes and their model representation.

5 Ways Ahead

We presented examples of artefacts that may occur when bias correction is applied without considering the underlying processes. These examples illustrate that bias correction is only recommended if, in a given context, the following assumptions hold: first, relevant processes are reasonably well captured by the chosen climate models, including the temporal structure (Figure 2) and location (Figure 5) of the large-scale circulation, as well as the regional response to large-scale processes (Figure 3) and local feedbacks (Figure 6). Second, the climate models resolve the local spatial-temporal variability (Figure 4) and climate change (Figure 6). Over areas where some of these assumptions are not valid, the bias corrected output should be handled with great care. To avoid the related artefacts, we advocate research along four major strands. Process understanding should inform bias correction already during the climate model selection, as part of the actual bias correction procedure, when evaluating the correction and when shifting to alternative approaches.

5.1 Understanding Model Biases

Any regional climate projection that is intended to serve for decision making relies on a realistic simulation of all relevant processes controlling climate change. It has thus to be recognised that the appropriateness of a bias correction is only partly a statistical issue, but importantly an issue of the credibility of the driving model. Thus it is important to understand the origins of model biases, from the large-scale circulation to regional-scale forcings and feedbacks.

Emergent constraints[78] are a promising approach to understand the influence of model biases in present climate on the climate change signal. The essence of this approach is to identify strong statistical relationships between (1) an observable feature of the simulated present climate and (2) a future climate change signal in a large ensemble of climate models. If the statistical relationship is associated with robust physics, then the most realistic models in the present climate can be declared to have the most credible future climate change signal. Basically, emergent constraints allow one to determine which present climate biases are most consequential for future climate change signals. Emergent constraints have already been applied extensively to global-scale processes and feedbacks. However, there is no reason they cannot be applied to regional-scale processes, either in ensembles of global models or associated downscaled data products. Examples are the influence of location biases in the large-scale atmospheric circulation on regional precipitation changes[79], or the influence of biases in snow-albedo feedbacks on the regional warming signal[80]. We advocate searching for emergent constraints along these lines at the regional scale. This technique would exploit regional biases to improve the credibility of future climate change signals, instead of trying to get rid of them in some unphysical way.

As discussed above, a key issue is also to understand the relationship of biases across time-scales: how do biases in day-to-day or interannual variability translate into biases in the climate change signal? Identifying such linkages may help to judge the feasibility of trend modifications.

Given that fundamental model errors cannot be corrected by bias correction[30], we advocate for a region-targeted selection of the driving GCMs prior to any downscaling exercise. The aim of such a procedure would neither be to identify the overall best performing GCM, nor to discard models simulating biased surface variables. Rather, it would be to discard those GCMs that unrealistically simulate the processes controlling the regional climate of interest, and those that have strong location biases in the large-scale atmospheric circulation (see Figure 5). The selection of course has to account in some manner for the range of uncertainty in global climate sensitivity.

There is realistic hope that further model improvements and increased model resolution may improve the representation of both local and large-scale processes[81,54,82,58,83]. The resulting reduction in location biases and the increase in credibility of future projections will render subsequent bias correction a more defensible approach.

5.2 New Bias Correction Approaches

We identified two major limitations of current bias correction methods: their difficulties in downscaling to finer spatial scales, and their inability to improve the local climate change signal. To address both these issues, we advocate the development of new methods, combining advanced statistical modelling with physical understanding.

The downscaling problem requires stochastic approaches which generate sub-grid spatial variability: to simulate fine-scale precipitation fields, or to simulate sub-grid temperature variations such as inversions. Recently it has been proposed to carry out the bias correction at the grid-box scale, and then to stochastically downscale to finer scales[84]. More realistic fields can be obtained by including process information, e.g., by conditioning the downscaling on the atmospheric circulation[29].

As laid out above, a misrepresentation of regional feedbacks may result in an implausible regional climate change signal, and quantile mapping will likely not be able to improve it. Avenues should be explored to explicitly account for regional-scale processes and feedbacks for improving the climate change signal in the statistical postprocessing. One such avenue is, again, process-based bias correction. For instance, summer temperature biases may depend on temperature because of soil moisture feedbacks. Here it has been suggested to condition the correction on simulated soil moisture[67]. Another avenue are emulators of high-resolution RCMs, which simulate a credible climate change signal. For instance, local variations in the warming signal could be statistically expressed by covariates such as elevation, continentality or large-scale warming patterns. These expressions can be calibrated across a range of dynamically downscaled GCMs, and then applied to statistically downscale the climate change signal of other GCMs[85]. Such emulators could also be developed for other regional processes such as convection: measures of stability and moisture convergence could serve as input to emulate high-resolution convection permitting models. Thereby the representation of extreme events could be improved, a weak point of essentially all statistical post-processing methods so far.

5.3 Evaluating Bias Correction

None of the artefacts we presented would have been identified by a standard cross-validation of marginal aspects. Rigorous standards for evaluating bias correction methods need thus be developed. These should encompass temporal as well as process-oriented aspects[86]. For instance, an investigation of the spell length distribution (Figure 2), or an evaluation conditional on the state of the relevant climatic phenomenon (Figure 3) may help to reveal bias correction problems. In any case, the resulting bias corrected time series should be - at least for some selected grid boxes - visually inspected and compared with observational data. A useful indicator for an unphysical bias correction is the dis-similarity between modelled and observed distribution (Figure 1): major differences point to a misrepresentation of key processes, and a bias correction is unlikely to be sensible. In any case one should investigate the projected signals for implausible change (Figures 5 and 6). The use of pseudo realities for evaluating simulated trends[86] should further be explored.

5.4 Alternative approaches

Finally, we advocate to explore alternative approaches in any given context. In some cases, perfect prognosis statistical downscaling and change factor weather generators[22] may be more appropriate than bias correction. In other cases, response surfaces[87] with qualitative input of possible climate changes might suffice to obtain decision relevant information, or expert knowledge combined with raw climate model simulations might provide useful information. Location biases of the atmospheric circulation may be reduced by surrogate climate warming studies[88]. Finally, storyline simulations of how single but relevant past events might look in a warmer future may substantially improve the representation of local feedbacks: they reduce computational costs and thereby enable much higher model resolutions[89].

6 Final Remarks

Bias correction is not a Swiss Army knife, many issues remain unresolved, and research is needed to understand its limitations and to develop new concepts for mitigating the effects of climate model biases. Bias correction is not a purely statistical problem and cannot overcome fundamental deficiencies in climate models.

We recommend carrying out any bias correction or downscaling based on solid knowledge about the relevant climatic phenomena and the ability of the employed climate models to simulate them. To identify implausible results, a successful bias correction thus requires a close collaboration with global and regional climate modellers as well as experts both in the relevant large scale climatic phenomena and the local weather and climate of the target region. We recommend a concerted action among all involved disciplines to build up the necessary knowledge and to develop best practice guidelines to make bias correction a rigorous science.

In any case, it is essential to disclose relevant expert decisions affecting the results and to transparently discuss the usefulness and limitations of the output with users, in particular as the use of climate model data by non-experts is more and more operationalised by climate service providers[2].

References

[1] B.C. Hewitson, J. Daron, R.G. Crane, M.F. Zermoglio, and C. Jack. Interrogating empirical-statistical downscaling. *Clim. Change*, 122:539–554, 2014.

[2] P. Adams, E. Eitland, B. Hewitson, C. Vaughan, R. Wilby, and S. Zebiak. Toward an ethical framework for climate services. A white paper of the climate services partnership working group on climate services ethics, 2015.

[3] G. Flato, J. Marotzke, B. Abiodun, P. Braconnot, S.C. Chou, W. Collins, P. Cox, F. Driouech, S. Emori, V. Eyring, C. Forest, P. Gleckler, E. Guilyardi, C. Jakob, V. Kattsov, C. Reason, and M. Rummukainen. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, chapter Evaluation of Climate Models. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

[4] S. Kotlarski, K. Keuler, O.B. Christensen, A. Colette, M. Déqué, A. Gobiet, K. Goergen, D. Jacob, D. Lüthi, E. van Meijgaard, G. Nikulin, C. Schär, C. Teichmann, R. Vautard, K. Warrach-Sagi, and V. Wulfmeyer. Regional climate modelling on European scales: A joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geosci. Model. Dev. Discuss.*, 7:217–293, 2014.

[5] E.P. Maurer. Fine-resolution climate projections enhance regional climate change impact studies. *EOS*, 88(47):504, 2007.

[6] H. Li, J. Sheffield, and E.F. Wood. Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *J. Geophys. Res.*, 115:D10101, 2010.

[7] A. Dosio, P. Paruolo, and R. Rojas. Bias correction of the ENSEMBLES high resolution climate change projections for use by impact models: Analysis of the climate change signal. *J. Geophys. Res. Atmos.*, 117(D17), 2012.

[8] S. Hagemann, C. Chen, J.O. Haerter, J. Heinke, D. Gerten, and C. Piani. Impact of a statistical bias correction on the projected hydrological changes obtained from three GCMs and two hydrology models. *J. Hydrometeorol.*, 12(4):556–578, 2011.

[9] A.M.K Stoner, K. Hayhoe, X. Yang, and D.J. Wuebbles. An asynchronous regional regression model for statistical downscaling of daily climate variables. *Int. J. Climatol.*, 33:2473–2494, 2013.

[10] E.H. Girvetz, E.P. Maurer, P.B. Duffy, A. Ruesch, B. Thrasher, and C. Zganjar. Making climate data relevant to decision making: the important details of spatial and temporal downscaling. The World Bank, 2013.

[11] S. Hempel, K. Frieler, L. Warszawski, J. Schewe, and F. Piontek. A trend-preserving bias correction - the ISI-MIP approach. *Earth Syst. Dynam.*, 4:219–236, 2013.

[12] E.P. Maurer, L. Brekke, T. Pruitt, B. Thrasher, J. Long, P. Duffy, M. Dettinger, D. Cayan, and J. Arnold. An enhanced archive facilitating climate impact and adaptation analysis. *Bull. Amer. Meteorol. Soc.*, 95(7):1011–1019, 2014.

[13] CORDEX. Bias-adjusted RCM data. http://www.cordex.org/index.php?option=com_content&view=a 2016.

[14] S. Gangopadhyay, T. Pruitt, L. Brekke, and D. Raff. Hydrologic projections for the Western United States. *EOS*, 92(48):441–442, 2011.

[15] S. Hagemann, C. Chen, D. Clark, S. Folwell, S.N. Gosling, I. Haddeland, N. Hannasaki, J. Heinke, F. Ludwig, F. Voss, and A. Wiltshire. Climate change impact on available water resources obtained using multiple global climate and hydrology models. *Earth Syst. Dynam.*, 4:129–144, 2013.

[16] L. Warszawski, K. Frieler, V. Huber, F. Piontek, O. Serdeczny, and J. Schewe. The Inter-Sectoral Impact Model Intercomparison Project (ISIMIP): Project framework. *Proc. Nat. Acad. Sci.*, 111(9):3228–3232, 2014.

[17] D. Cayan, K. Kunkel, C. Castro, A. Gershunov, J. Barsugli, A. Ray, J. Overpeck, M. Anderson, J. Russell, B. Rajagopalan, I. Rangwala, and P. Duffy. *Assessment of Climate Change in the Southwest United States: A Report Prepared for the National Climate Assessment*, chapter Future climate: Projected average., pages 101–125. Island Press, 2013.

[18] World Bank. Turn Down the Heat: Climate Extremes, Regional Impacts, and the Case for Resilience. A report for the World Bank by the Potsdam Institute for Climate Impact Research and Climate Analytics, 2013.

[19] A. Georgakakos, P. Fleming, M. Dettinger, C. Peters-Lidard, T.C. Richmond, K. Reckhow, K. White, and D. Yates. *Climate Change Impacts in the United States: The Third National Climate Assessment*, chapter Water Resources, pages 69–112. U.S. Global Change Research Program, 2014.

[20] Worldbank. Climate change knowledge portal. http://sdwebx.worldbank.org/climateportal/.

[21] prepdata. Partnership for Resilience and Preparedness (PREP). www.prepdata.org.

[22] D. Maraun, F. Wetterhall, A.M. Ireson, R.E. Chandler, E.J. Kendon, M. Widmann, S. Brienen, H.W. Rust, T.Sauter, M.Themeßl, V.K.C. Venema, K.P. Chun, C.M. Goodess, R.G. Jones, C. Onof, M. Vrac, and I. Thiele-Eich. Precipitation downscaling under

climate change. Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.*, 48:RG3003, 2010.

[23] C. Teutschbein and J. Seibert. 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.

[24] C. Piani, J.O. Haerter, and E. Coppola. Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.*, 99(1-2):187–192, 2010.

[25] S. Vannitsem. Bias correction and post-processing under climate change. *Nonlin. Proc. Geophys.*, 18:911–924, 2011.

[26] J. Eden, M. Widmann, D. Grawe, and S. Rast. Skill, correction, and downscaling of GCM-simulated precipitation. *J. Climate*, 25:3970–3984, 2012.

[27] D. Maraun. Bias correction, quantile mapping and downscaling: Revisiting the inflation issue. *J. Climate*, 26:2137–2143, 2013.

[28] D. Maraun and M. Widmann. The representation of location by a regional climate model in complex terrain. *Hydrol. Earth Syst. Sci.*, 19:3449–3456, 2015.

[29] N. Addor, M. Rohrer, R. Furrer, and J. Seibert. Propagation of biases in climate models from the synoptic to the regional scale: Implications for bias adjustment. *J. Geophys. Res.*, 121(5):2075–2089, 2016.

[30] T.F. Stocker, Q. Dahe, G.-K. Plattner, and M. Tignor. IPCC Workshop on Regional Climate Projections and their Use in Impacts and Risk Analysis Studies. https://www.ipcc.ch/pdf/supporting-material/RPW_WorkshopReport.pdf, 2015.

[31] U. Ehret, E. Zehe, V. Wulfmeyer, K. Warrach-Sagi, and J. Liebert. Should we apply bias correction to global and regional climate model data? *Hydrol. Earth Syst. Sci.*, 16, 2012.

[32] J.O. Haerter, S. Hagemann, C. Moseley, and C. Piani. Climate model bias correction and the role of timescales. *Hydrol. Earth Syst. Sci.*, 15(3):1065–1079, 2011.

[33] F. Johnson and A. Sharma. A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations. *Wat. Resour. Res.*, 48:W01504, 2012.

[34] P.-A. Michelangeli, M. Vrac, and H. Loukos. Probabilistic downscaling approaches: Application to wind cumulative distribution functions. *Geophys. Res. Lett.*, 36(11), 2009.

[35] C. Piani and J.O. Haerter. Two dimensional bias correction of temperature and precipitation copulas in climate models. *Geophys. Res. Lett.*, 39(20):L20401, 2012.

[36] M. Vrac and P. Friederichs. Multivariate-intervariable, spatial, and temporal-bias correction. *J. Climate*, 28(1):218–237, 2015.

[37] A.A.L. Levy, W. Ingram, M. Jenkinson, C. Huntingford, F.H. Lambert, and M. Allen. Can correcting feature location in simulated mean climate improve agreement on projected changes? *Geophys. Res. Lett.*, 40:354–358, 2013.

[38] J.J. Barsugli, G. Guentchev, R. M. Horton, A. Wood, L.O. Mearns, X.-Z. Liang, J.A. Winkler, K. Dixon, K. Hayhoe, R.B. Rood, L. Goddard, A. Ray, L. Buja, and C. Ammann. The practitioner's dilemma: How to assess the credibility of downscaled climate projections. *EOS*, 94(46):424–425, 2013.

[39] C. Wang, L. Zhang, S.-K. Lee, L. Wu, and C.R. Mechoso. A global perspective on CMIP5 climate model biases. *Nat. Clim. Change*, 4:201–205, 2014.

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

[41] C.M. Buser, H.R. Künsch, D. Lüthi, M. Wild, and C. Schär. Bayesian multi-model projection of climate: bias assumptions and interannual variability. *Clim. Dynam.*, 33:849–868, 2009.

[42] D. Maraun. Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. *Geophys. Res. Lett.*, 39:L06706, 2012.

[43] C. Teutschbein and J. Seibert. Is bias correction of regional climate model (rcm) simulations possible for non-stationary conditions? *Hydrol. Earth Syst. Sci.*, 17(12):5061–5077, 2013.

[44] F. Johnson and A. Sharma. Accounting for interannual variability: A comparison of options for water resources climate change impact assessments. *Wat. Resour. Res.*, 47:W04508, 2011.

[45] P. Berg, H. Feldmann, and H.-J. Panitz. Bias correction of high resolution regional climate model data. *J. Hydrol.*, 448-449:80–92, 2012.

[46] C. Chen, J.O. Haerter, S. Hagemann, and C. Piani. On the contribution of statistical bias correction to the uncertainty in the projected hydrological cycle. *Geophys. Res. Lett.*, 38:L20403, 2011.

[47] A. Dosio and P. Paruolo. Bias correction of the ENSEMBLES high resolution climate change projections for use by impact models: Evaluation on the present climate. *J. Geophys. Res. Atmos.*, 116(D16), 2011.

[48] M. J. Themeßl, A. Gobiet, and A. Leuprecht. Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *Int. J. Climatol.*, 31:1530–1544, 2011.

[49] L. Gudmundson, J.B. Bremnes, J.E. Haugen, and T. Engen-Skaugen. Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations a comparison of methods. *Hydrol. Earth Syst. Sci.*, 16:3383–3390, 2012.

[50] J. Rajczak, S. Kotlarski, and C. Schär. Does quantile mapping of simulated precipitation correct for biases in transition probabilities and spell lengths? *J. Climate*, 29:1605–1615, 2016.

[51] G. Masato, B.J. Hoskins, and T. Woollings. Winter and summer northern hemisphere blocking in CMIP5 models. *J. Climate*, 26:7044–7059, 2013.

[52] T. Woollings. Dynamical influences on European climate: An uncertain future. *Phil. Trans. R. Soc. A*, 368:3733–3756, 2010.

[53] G. Zappa, L.C. Shaffrey, and K.I. Hodges. The ability of CMIP5 models to simulate North Atlantic extratropical cyclones. *J. Climate*, 26:5379–5396, 2013.

[54] P. Davini, J. von Hardenberg, S. Corti, H.M. Christensen, S. Juricke, A. Subramanian, P.A.G. Watson, A. Weisheimer, and T.N. Palmer. Climate SPHINX: evaluating the impact of resolution and stochastic physics parameterisations in the EC-Earth global climate model. *Geosci. Model Dev.*, 10(3):1383–1402, 2017.

[55] F. Pithan, T.G. Shepherd, G. Zappa, and I. Sandu. Climate model biases in jet streams, blocking and storm tracks resulting from missing orographic drag. *Geophys. Res. Lett.*, 43:7231–7240, 2016. doi:10.1002/2016GL069551.

[56] A. van Niekerk, J.F. Scinocca, and T.G. Shepherd. The modulation of stationary waves, and their response to climate change, by parameterized orographic drag. *J. Atmos. Sci.*, 74:2557–2574, 2017. doi:10.1175/JAS-D-17-0085.1.

[57] M. Ashfaq, C.B. Skinner, and N.S. Diffenbaugh. Influence of SST biases on future climate change projections. *Clim. Dynam.*, 36:1303–1319, 2011.

[58] A.A. Scaife, D. Copsey, C. Gordon, C. Harris, T. Hinton, S. Keeley, A. O'Neill, M. Roberts, and K. Williams. Improved Atlantic winter blocking in a climate model. *Geophys. Res. Lett.*, 38:L23703, 2011.

[59] S.P.E. Keeley, R.T. Sutton, and L.C. Shaffrey. The impact of North Atlantic sea surface temperature errors on the simulation of North Atlantic European region climate. *Q. J. R. Meteorol. Soc.*, 138:1774–1783, 2012.

[60] T.G. Shepherd. Atmospheric circulation as a source of uncertainty in climate change projections. *Nat. Geosci.*, 7:703–708, 2014.

[61] A. Hall. Projecting regional change. *Science*, 346(6216):1461–1462, 2014.

[62] A. Colette, R. Vautard, and M. Vrac. Regional climate downscaling with prior statistical correction of the global climate forcing. *Geophys. Res. Lett.*, 39(13):L13707, 2012.

[63] E.P. Maurer and D.W. Pierce. Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean. *Hydrol. Earth Syst. Sci.*, 18(3):915–925, 2014.

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

[65] D.W. Pierce, D.R. Cayan, E.P. Maurer, J.T. Abatzoglou, and K.C. Hegewisch. Improved Bias Correction Techniques for Hydrological Simulations of Climate Change. *J. Hydrometeorol.*, 16(6):2421–2442, 2015.

[66] M.B. Switanek, P.A. Troch, C.L. Castro, A. Leuprecht, H.-I. Chang, R. Mukherjee, and E.M.C. Demaria. Scaled distribution mapping: a bias correction method that preserves raw climate model projected changes. *Hydrol. Earth Syst. Sci.*, 21:2649–2666, 2017.

[67] O. Bellprat, S. Kotlarski, D. Lüthi, and C. Schär. Physical constraints for temperature biases in climate models. *Geophys. Res. Lett.*, 40:4042–4047, 2013.

[68] H. Bellenger, E. Guilyardi, J. Leloup, M. Lengaigne, and J. Vialard. ENSO representation in climate models: from CMIP3 to CMIP5. *Clim. Dynam.*, 42:1999–2018, 2014.

[69] L. Chen, T. Li, and Y. Yu. Causes of strengthening and weakening of ENSO amplitude under global warming in four CMIP5 models. *J. Climate*, 28:3250–3274, 2015.

[70] M. Collins, R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, W.J. Gutowski, T. Johns, G. Krinner, M. Shongwe, C. Tebaldi, A.J. Weaver, and M. Wehner. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, chapter Long-term Climate Change: Projections, Committments and Irreversibility. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

[71] H. Annamalai, J. Hafner, K.P. Sooraj, and P. Pillai. Global warming shifts the monsoon circulation, drying South Asia. *J. Climate*, 26:2701–2718, 2013.

[72] M.K. Roxy, K. Ritika, P. Terray, R. Murtugudde, K. Ashok, and B.N. Goswami. Drying of indian subcontinent by rapid indian ocean warming and a weakening land-sea thermal gradient. *Nat. Commun.*, 6:7423, 2015.

[73] J. Willison, W.A. Robinson, and G.M. Lackmann. North atlantic storm-track sensitivity to warming increases with model resolution. *J. Climate*, 28:4513–4524, 2015.

[74] A. Hall, X. Qu, and J.D. Neelin. Improving predictions of summer climate change in the united states. *Geophys. Res. Lett.*, 35:L01702, 2008.

[75] J.H. Christensen and F. Boberg. Temperature dependent climate projection deficiencies in CMIP5 models. *Geophys. Res. Lett.*, 39:L24705, 2012.

[76] E.J. Kendon, N.M. Roberts, H.J. Fowler, M.J. Roberts, S.C. Chan, and C.A. Senior. Heavier summer downpours with climate change revealed by weather forecast resolution model. *Nat. Clim. Change*, 4:570–576, 2014.

[77] E.P. Meredith, D. Maraun, V.A. Semenov, and W. Park. Evidence for added value of convection permitting models for studying changes in extreme precipitation. *J. Geophys. Res. Atmos.*, 120:12,500–12,513, 2015.

[78] M. Collins, R.E. Chandler, P.M. Cox, J.M. Huthnance, J. Rougier, and D.B. Stephenson. Quantifying future climate change. *Nat. Clim. Change*, 2(6):403–409, 2012.

[79] I.R. Simpson, R. Seager, M. Ting, and T.A. Shaw. Causes of change in northern hemisphere winter meridional winds and regional hydroclimate. *Nat. Clim. Change*, 6(1):65–70, 2016.

[80] X. Qu and A. Hall. On the persistent spread in snow-albedo feedback. *Clim. Dynam.*, 42(1-2):69, 2014.

[81] A. Dawson, T.N. Palmer, and S. Corti. Simulating regime structures in weather and climate prediction models. *Geophys. Res. Lett.*, 39(21), 2012.

[82] P. Liu and Coauthors. An MJO simulated by the NICAM at 14- and 7-km resolutions. *Mon. Wea. Rev.*, 137:3254–3268, 2009.

[83] A.F. Prein, W. Langhans, G. Fosser, A. Ferrone, N. Ban, K. Goergen, M. Keller, M. Tlle, O. Gutjahr, F. Feser, E. Brisson, S. Kollet, J. Schmidli, N.P.M. van Lipzig, and R. Leung. A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Rev. Geophys.*, 53(2):323–361, 2015.

[84] C. Volosciuk, D. Maraun, M. Vrac, and M. Widmann. A combined statistical bias correction and stochastic downscaling method for precipitation. *Hydrol. Earth Syst. Sci.*, 21(3):1693–1719, 2017.

[85] D.B. Walton, F. Sun, A. Hall, and S. Capps. A hybrid dynamical-statistical downscaling technique. part i: Development and validation of the technique. *J. Climate*, 28(12):4597–4617, 2015.

[86] D. Maraun, M. Widmann, J.M. Gutierrez, S. Kotlarski, R.E. Chandler, E. Hertig, J. Wibig, R. Huth, and R.A.I. Wilcke. VALUE: A framework to validate downscaling approaches for climate change studies. *Earth's Future*, 3:1–14, 2015.

[87] C. Prudhomme, R.L. Wilby, S. Crooks, A.L. Kay, and N.S. Reynard. Scenario-neutral approach to climate change impact studies: Application to flood risk. *J. Hydrol.*, 390:198–209, 2010.

[88] C. Schär, C. Frei, D. Lüthi, and H.C. Davies. Surrogate climate-change scenarios for regional climate models. *Geophys. Res. Lett.*, 23(6):669–672, 1996.

[89] W. Hazeleger, B.J.J.M. van den Hurk, E. Min, G.J. van Oldenborgh, A.C. Petersen, D.A. Stainforth, E. Vasileiadou, and L.A. Smith. Tales of future weather. *Nat. Clim. Change*, 5:107–113, 2015.

[90] A.M.G. Klein Tank, J.B. Wijngaard, G.P. Können, R. Böhm, G. Demarée, A. Gocheva, M. Mileta, S. Pashiardis, L. Hejkrlik, C. Kern-Hansen, R. Heino, P. Bessemoulin, G. Müller-Westermeier, M. Tzanakou, S. Szalai, T. Pálsdóttir, D. Fitzgerald, S. Rubin, M. Capaldo, M. Maugeri, A. Leitass, A. Bukantis, R. Aberfeld, A.F.V. van Engelen, E. Forland, M. Mietus, F. Coelho, C. Mares, V. Razuvaev, E. Nieplova, T. Cegnar, J.A. López, B. Dahlström, A. Moberg, W. Kirchhofer, A. Ceylan, O. Pachaliuk, L.V. Alexander, and P. Petrovic. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *Int. J. Climatol.*, 22(12):1441–1453, 2002.

[91] J.M. Gutiérrez, R. Cano, A.S. Cofiño, and C. Sordo. Analysis and downscaling multi-model seasonal forecasts in Peru using self-organizing maps. *Tellus A*, 57(3):435–447, 2005.

[92] E.P. Maurer, A.W. Wood, J.C. Adam, D.P. Lettenmaier, and B. Nijssen. A long-term hydrologically-based data set of land surface fluxes and states for the conterminous United States. *J. Climate*, 15(22):3237–3251, 2002.

[93] D.P. van Vuuren, J. Edmonds, M. Kainuma, K. Riahi, A. Thomson, K. Hibbard, G. Hurtt, T. Kram, V. Krey, J.-F. Lamarque, T. Masui, M. Meinshausen, N. Nakicenovic, S.J. Smith, and S.K. Rose. The representative concentration pathways: an overview. *Clim. Change*, 109(1):5–31, 2011.

Corresponding author
Douglas Maraun (douglas.maraun@uni-graz.at)

Acknowledgements
Many of the ideas laid out in this paper have been developed during discussions within the EU COST Action ES1102 “VALUE”, in particular during the workshop “Global climate model biases: causes, consequences and correctability” held at the Max Planck Institute for Meteorology, Hamburg, October 2014.

Author contributions
The paper is a result of a workshop organised by D.M. and S.H. D.M. wrote the first draft of the manuscript with inputs from all authors. D.M., G.Z., J.M.G and D.W. contributed analyses underlying the figures. All authors discussed the content of the manuscript.

List of Figures

1 **Cross-validation problem** | Quantile mapping from ERA40 daily boreal winter (DJF) temperature [°C, Southern Ocean, 45S-55S, 175W-163W] to E-OBS daily precipitation [mm/day, Central Europe, 45N-55N, 5E-17E], calibrated over 1961-1980. **a-c**, mean and **d-f**, 95th percentile over validation period (1981-2000). **a,d**, uncorrected ERA40, **b,e** observations, **c,f** corrected ERA40. **g**, histogram of biases across all grid boxes. **h** QQ-plot for grid box close to Venice (see cross in panel **a**). A QQ-plot plots the quantiles of two distributions against each other, i.e., for two time series, the values are sorted separately and then plotted against each other. The correction function is based on linear interpolation between empirical quantiles with a constant correction for new extreme values. 21

2 **Unrealistic dry spell lengths** | Distribution of dry spell lengths (wet-day threshold 0.1 mm) at **a**, Tafjord (Norway; 7.41° W, 62.23°N, winter), **b**, Constanta (Romania; 28.63° E, 44.22°N, winter), **c**, Sion (Switzerland, 7.33° E, 46.22°N, winter) and **d**, Rome (Italy, 12.58° E, 41.78°N, summer) of MPI-ESM-LR downscaled with CLM to a horizontal resolution of 0.44°, 1971-2000. Black: observations (ECA-D[90]), blue: raw climate model, red: corrected climate model. Long dry spells are typically underrepresented even after a seasonal wet day correction (**a**), although in some cases the correction may improve the overall distribution (**b**). Often, artefacts are introduced for short (**c**) and long (**d**) spells. 22

3 **Non-representative model output** | Daily precipitation bias correction for the GISS-E2-R model against station data at Piura, Peru[91] from 1976-2000. **a**, observations; **b**, raw GCM data; **c**, quantile mapped GCM data; **d**, QQ plot. Grey shading: El Niño events. As the GCM is run in climate mode, simulated events are not synchronised with observations. Even though the quantile mapping perfectly adjusts the simulated distribution, the result is meaningless, as the GCM does not correctly capture the clustering of extreme precipitation during El Niño events. 23

4 **Missing temperature inversions** | Distribution of spring (MAM) daily mean temperature differences between Fresno (∼90m) and Three Rivers (70 km towards the southeast, at ∼400m) in California, US, 1981-2000. Blue: observations (1/8° gridded data[92]), orange: GFDL-CM3 GCM after quantile mapping against observations (scaled by 1/4). In reality, temperature inversions ($\Delta T < 0$) in the Central Valley occur on about 7% of the days. The coarse-resolution GCM does not simulate such inversions. Quantile mapping provides the correct climatological temperature difference, but is by construction unable to produce sub-grid inversions. The correction function was based on parametric Gaussian distributions. 24

5 **Large-scale circulation problems** | **a-c**, FGOALS-g2; **d-f**, MPI-ESM-MR. **a,d**, simulated (colour shading, mm/day) and observed (contour lines at 4 and 6 mm/day) mean winter precipitation 1976-2005. **b,e**, uncorrected mean precipitation averaged over 10W to 20E (vertical red lines in **a** and **d**) from present and future (2070-2099, RCP8.5[93]) simulations. **c,f**, corresponding corrected simulations (the black line by construction equals observed winter precipitation). Precipitation is bias corrected relative to the GPCP climatology (1980-2013). In FGOALS-g2, the storm track is unrealistically far south. As a result, even though the storm track shifts northwards in the future simulation, the corrected precipitation shifts southwards. For MPI-ESM-MR the circulation bias is low, avoiding an unphysical inconsistency between circulation and precipitation shift. The correction function multiplicatively adjusts long-term mean biases. 25

6 **Implausible sub-grid climate change signal** | Spring (MAM) daily mean temperature [°C] in the Sierra Nevada and Central Valley, California, US. **a-c**, present climate (1981-2000 average); **d-f**, simulated change (2081-2100 average minus 1981-2000 average, RCP8.5 scenario[93]). **a,d**, GFDL-CM3 GCM, bilinearly interpolated to 8km grid; **b,e**, corrected GCM (for present by construction identical with observations at 8km horizontal resolution[92]); **c,f**, WRF RCM at 3km horizontal resolution, driven with GFDL-CM3 climate change signal[85]. Whereas the RCM simulates plausible strong elevation-dependent warming (the strongest temperature increase in the Sierra Nevada mountains), the bias correction modulates the GCM change unsystematically and not related to elevation. 26

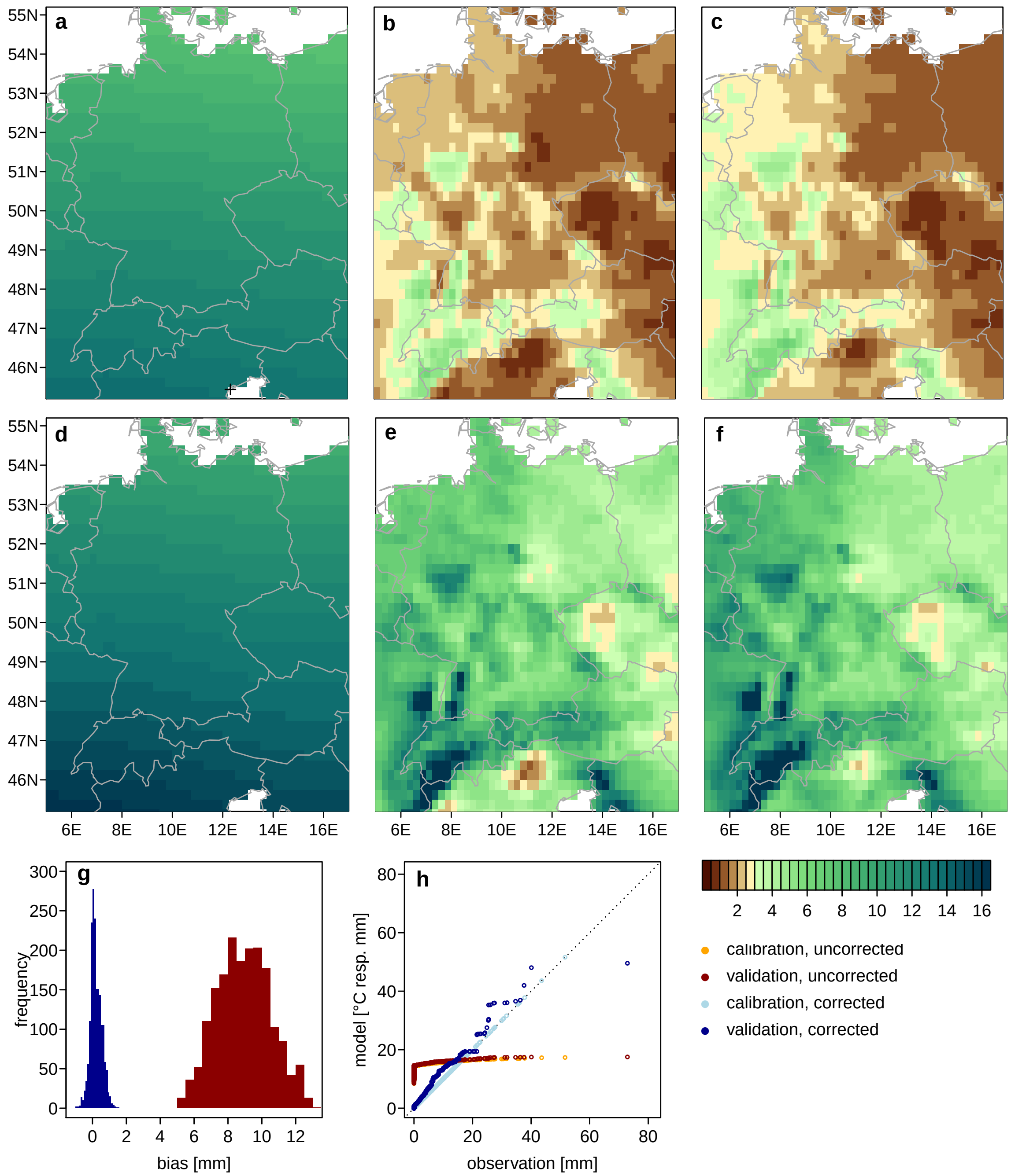

Figure 1: **Cross-validation problem** | Quantile mapping from ERA40 daily boreal winter (DJF) temperature [°C, Southern Ocean, 45S-55S, 175W-163W] to E-OBS daily precipitation [mm/day, Central Europe, 45N-55N, 5E-17E], calibrated over 1961-1980. **a-c**, mean and **d-f**, 95th percentile over validation period (1981-2000). **a,d**, uncorrected ERA40, **b,e** observations, **c,f** corrected ERA40. **g**, histogram of biases across all grid boxes. **h** QQ-plot for grid box close to Venice (see cross in panel **a**). A QQ-plot plots the quantiles of two distributions against each other, i.e., for two time series, the values are sorted separately and then plotted against each other. The correction function is based on linear interpolation between empirical quantiles with a constant correction for new extreme values.

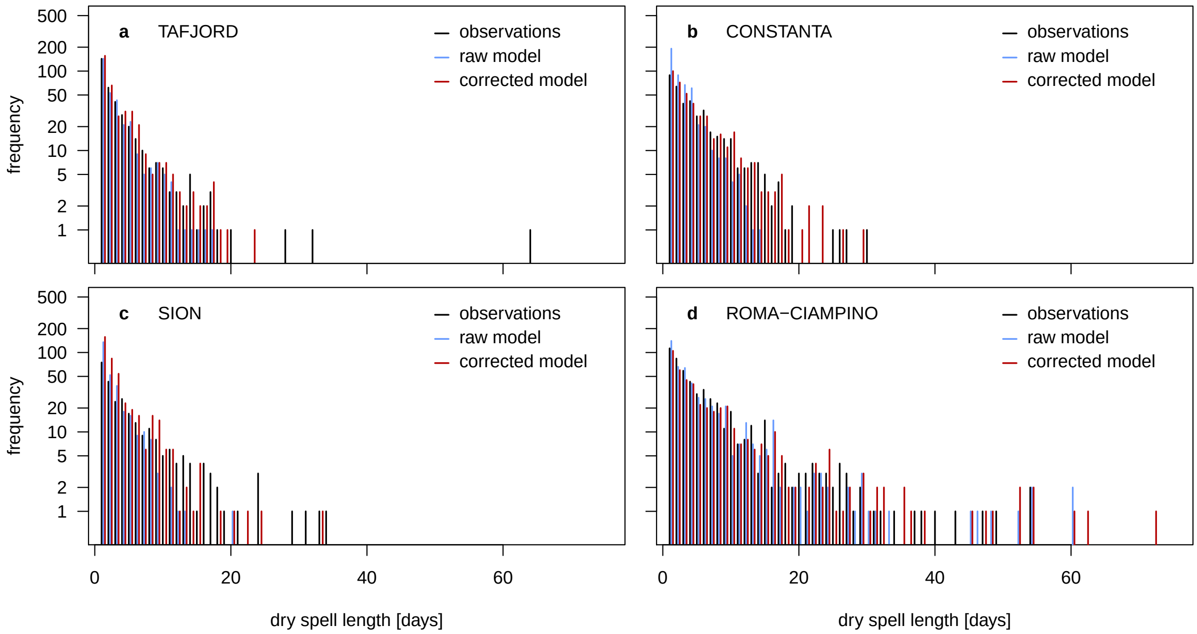

Figure 2: **Unrealistic dry spell lengths** | Distribution of dry spell lengths (wet-day threshold 0.1 mm) at **a**, Tafjord (Norway; 7.41° W, 62.23°N, winter), **b**, Constanta (Romania; 28.63° E, 44.22°N, winter), **c**, Sion (Switzerland, 7.33° E, 46.22°N, winter) and **d**, Rome (Italy, 12.58° E, 41.78°N, summer) of MPI-ESM-LR downscaled with CLM to a horizontal resolution of 0.44°, 1971-2000. Black: observations (ECA-D[90]), blue: raw climate model, red: corrected climate model. Long dry spells are typically underrepresented even after a seasonal wet day correction (**a**), although in some cases the correction may improve the overall distribution (**b**). Often, artefacts are introduced for short (**c**) and long (**d**) spells.

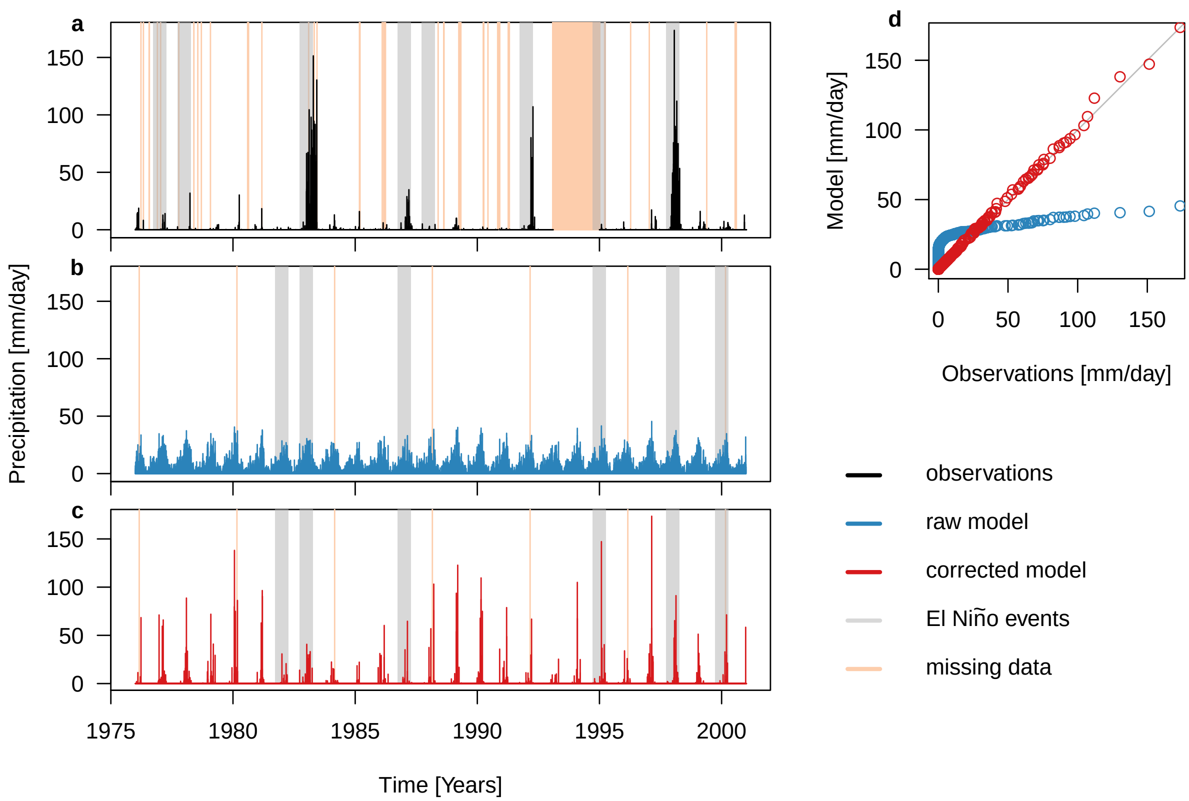

Figure 3: **Non-representative model output** | Daily precipitation bias correction for the GISS-E2-R model against station data at Piura, Peru[91] from 1976-2000. **a**, observations; **b**, raw GCM data; **c**, quantile mapped GCM data; **d**, QQ plot. Grey shading: El Niño events. As the GCM is run in climate mode, simulated events are not synchronised with observations. Even though the quantile mapping perfectly adjusts the simulated distribution, the result is meaningless, as the GCM does not correctly capture the clustering of extreme precipitation during El Niño events.

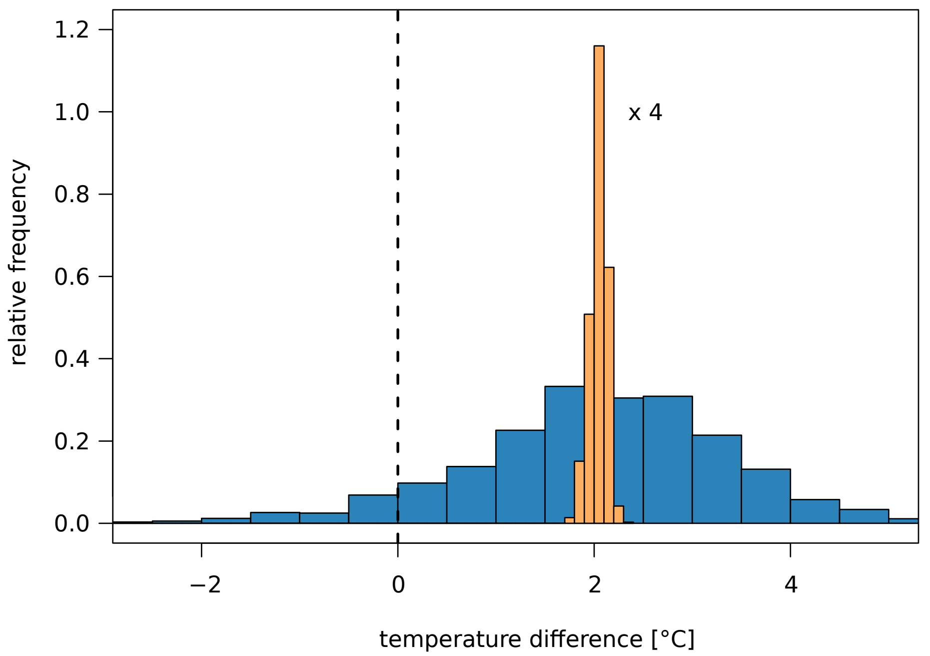

Figure 4: **Missing temperature inversions** | Distribution of spring (MAM) daily mean temperature differences between Fresno (∼90m) and Three Rivers (70 km towards the south-east, at ∼400m) in California, US, 1981-2000. Blue: observations (1/8° gridded data[92]), orange: GFDL-CM3 GCM after quantile mapping against observations (scaled by 1/4). In reality, temperature inversions ($\Delta T < 0$) in the Central Valley occur on about 7% of the days. The coarse-resolution GCM does not simulate such inversions. Quantile mapping provides the correct climatological temperature difference, but is by construction unable to produce sub-grid inversions. The correction function was based on parametric Gaussian distributions.

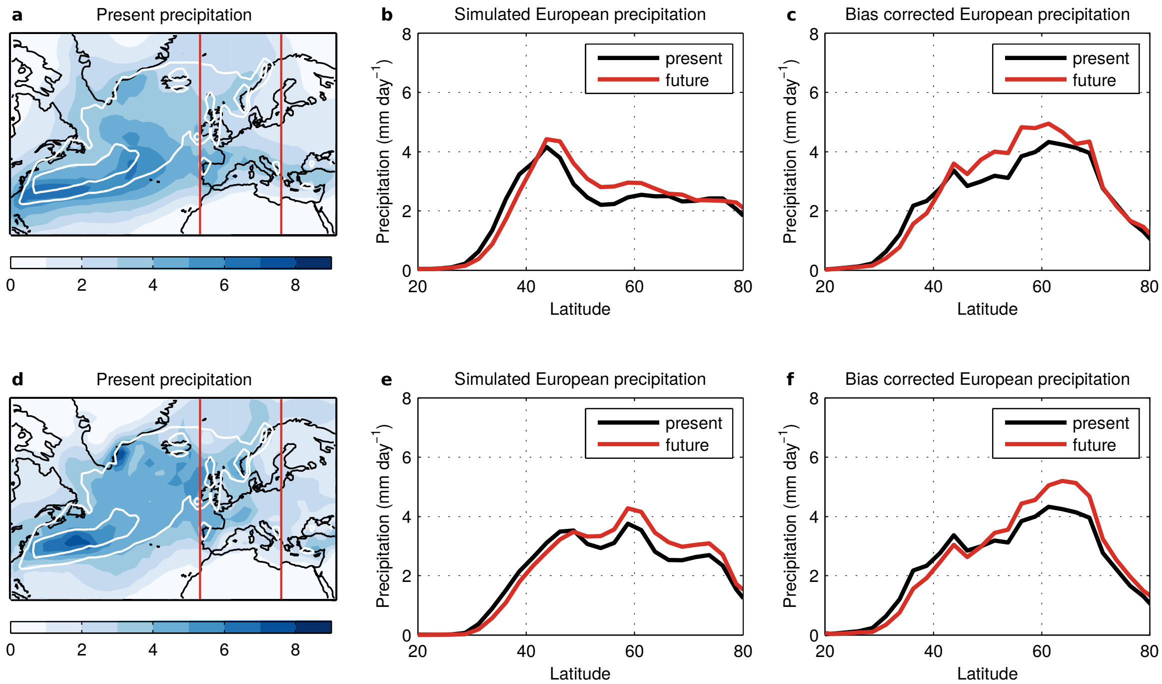

Figure 5: **Large-scale circulation problems** | **a-c**, FGOALS-g2; **d-f**, MPI-ESM-MR. **a,d**, simulated (colour shading, mm/day) and observed (contour lines at 4 and 6 mm/day) mean winter precipitation 1976-2005. **b,e**, uncorrected mean precipitation averaged over 10W to 20E (vertical red lines in **a** and **d**) from present and future (2070-2099, RCP8.5[93]) simulations. **c,f**, corresponding corrected simulations (the black line by construction equals observed winter precipitation). Precipitation is bias corrected relative to the GPCP climatology (1980-2013). In FGOALS-g2, the storm track is unrealistically far south. As a result, even though the storm track shifts northwards in the future simulation, the corrected precipitation shifts southwards. For MPI-ESM-MR the circulation bias is low, avoiding an unphysical inconsistency between circulation and precipitation shift. The correction function multiplicatively adjusts long-term mean biases.

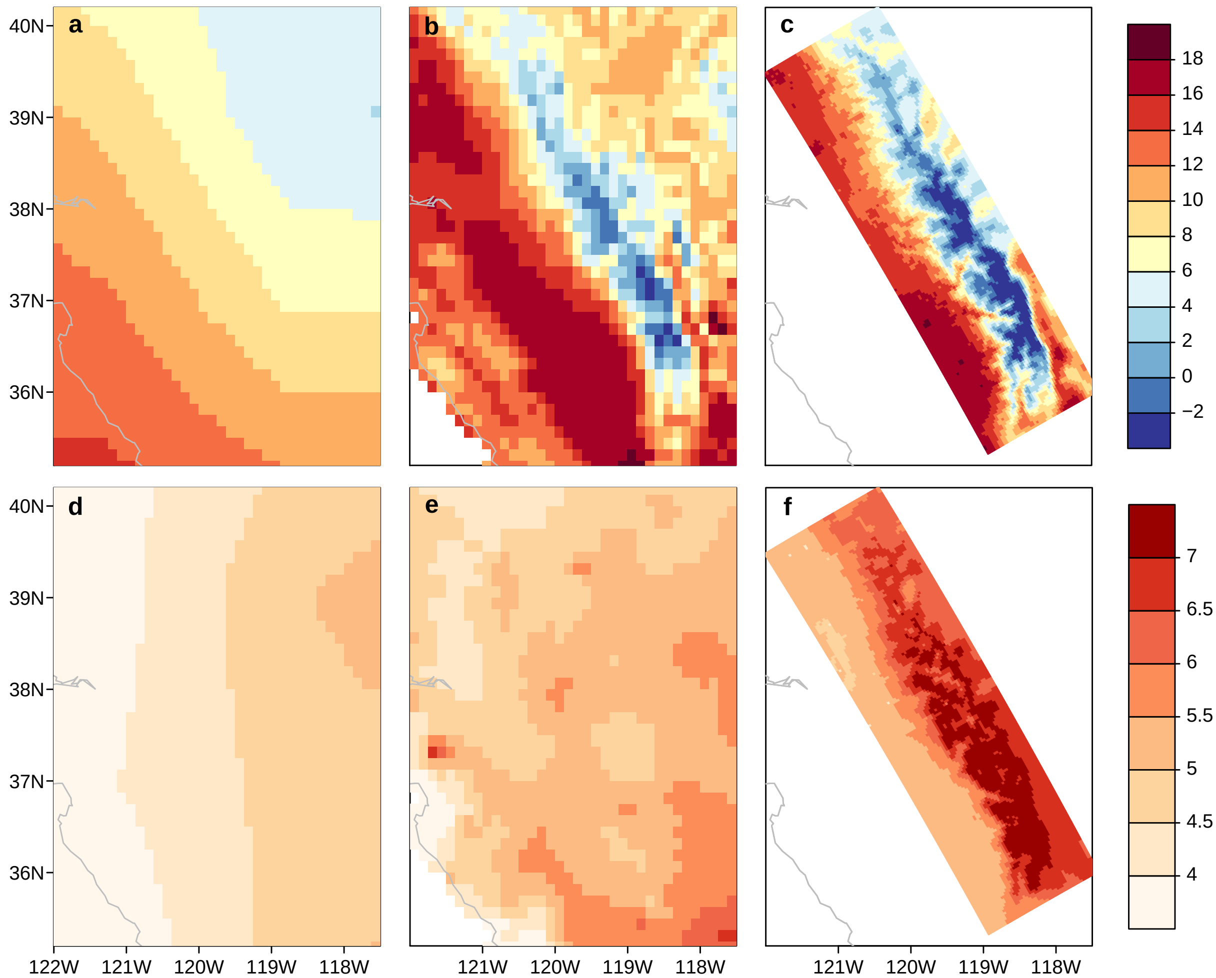

Figure 6: **Implausible sub-grid climate change signal** | Spring (MAM) daily mean temperature [°C] in the Sierra Nevada and Central Valley, California, US. **a-c**, present climate (1981-2000 average); **d-f**, simulated change (2081-2100 average minus 1981-2000 average, RCP8.5 scenario[93]). **a,d**, GFDL-CM3 GCM, bilinearly interpolated to 8km grid; **b,e**, corrected GCM (for present by construction identical with observations at 8km horizontal resolution[92]); **c,f**, WRF RCM at 3km horizontal resolution, driven with GFDL-CM3 climate change signal[85]. Whereas the RCM simulates plausible strong elevation-dependent warming (the strongest temperature increase in the Sierra Nevada mountains), the bias correction modulates the GCM change unsystematically and not related to elevation.