

# *Towards process-informed bias correction of climate change simulations*

Article

Accepted Version

Maraun, D., Shepherd, T. G. ORCID: <https://orcid.org/0000-0002-6631-9968>, 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 <https://centaur.reading.ac.uk/73495/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

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

Publisher: Nature Publishing Group

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# Towards process-informed bias correction of climate change simulations

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

<sup>1</sup> University of Graz, Austria

<sup>2</sup> University of Reading, UK

<sup>3</sup> University of Birmingham, UK

<sup>4</sup> University of California, Los Angeles, USA

<sup>5</sup> Institute of Physics of Cantabria (IFCA), Santander, Spain

<sup>6</sup> Max Planck Institute for Meteorology, Hamburg, Germany

<sup>7</sup> Japan-Agency for Marine-Earth Science and Technology (JAMSTEC),  
Yokohama, Japan

<sup>8</sup> Instituto Dom Luiz, University of Lisbon, Portugal

<sup>9</sup> 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.

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

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

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

54 We therefore argue that bias correction should not be dismissed, but that a solid conceptual  
55 and process understanding of climate model biases is required to successfully apply bias  
56 correction. The extent to which biases can be mitigated by post-processing depends on  
57 their origin. We present several examples, discuss their correctability by state-of-the-art bias  
58 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<sup>40,41,25,31,42,43</sup>. 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<sup>44,45,23,11</sup>, but in principle may vary in time<sup>41,40,25,42</sup>. Some authors define a bias as the time independent error component of a model<sup>24,46,47</sup>. 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<sup>6,47,48,23,49</sup> (some exceptions evaluate temporal or spatial dependence<sup>24,27</sup>). 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

98 precipitation fields are essentially unrelated and both fields show different long-term changes,  
99 the quantile mapping looks reasonable for the validation period, for mean and high values  
100 (c,f). The residual bias (g) between corrected model and observation purely stems from the  
101 different trends in both regions. The problem is especially severe for non-parametric quantile  
102 mapping, as demonstrated for the grid box enclosing Venice (h): even though the tempera-  
103 ture and precipitation distributions have completely different shapes, and both distributions  
104 change substantially over time (mean precipitation +28%, mean temperature -0.29K in the  
105 corresponding Southern Ocean grid box), the QQ plot looks reasonable also for the validation  
106 period. In other words: cross-validation of calibrated climatological properties is not able  
107 to identify the absurdity of the chosen example, and is thus not sufficient to evaluate the  
108 performance of bias correction. The reason for the failure is that, in climate modelling, model  
109 and observations are not in synchrony and predictive skill cannot, as in weather forecasting,  
110 be established by cross-validation<sup>26</sup>. The evaluation is restricted to long term distributional  
111 aspects only, and provided the sampling is adequate, cross-validation will merely reproduce  
112 the long-term distribution. But in a non-synchronous setting it is still possible to evaluate  
113 non-calibrated aspects, in particular for the temporal and, if required, spatial dependence  
114 structure. Such an evaluation would yield essential and indispensable information about the  
115 appropriateness of a bias correction.

### 116 3 Bias correction under present conditions

117 [Figure 2 about here.]

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

[Figure 3 about here.]

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

[Figure 4 about here.]

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

## 161 4 Bias correction under climate change conditions

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

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

[Figure 5 about here.]

171 It has been argued that biases in surface weather resulting from circulation biases cannot be  
 172 bias corrected<sup>26,30</sup>. For instance, when the frequency of circulation types is misrepresented,  
 173 bias correction may increase biases for specific circulation types<sup>29</sup>. Here we further show that

174 bias correction in the presence of substantial circulation biases may induce implausible future  
175 signals.

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

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

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

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

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

212 Many cases have been identified where climate models may simulate implausible changes  
213 of large-scale climatic phenomena, because the underlying processes are not realistically rep-  
214 resented. Prominent examples are the representation of ENSO feedbacks<sup>68,69</sup>, the Indian  
215 summer monsoon<sup>70,71,72</sup>, the influence of increased diabatic heating on the intensification of  
216 extratropical cyclones<sup>73</sup>, or European blocking<sup>51</sup>. Current bias correction methods will not



217 succeed in improving these changes, as they result from fundamental climate model errors<sup>30</sup>.  
218 At the regional scale, misrepresented land-surface interactions may result in implausi-  
219 ble climate change trends. For instance, models simulating unrealistically low summer soil  
220 moisture tend to over-represent summer temperature increases<sup>74,75</sup>; similarly the simulated  
221 increase of spring temperature is tightly linked to snow-albedo feedback strength<sup>74</sup>. Further-  
222 more trends may be implausible as a result of inadequately parameterised sub-grid processes.  
223 For instance, there is evidence that the response of summer convective precipitation extremes  
224 to global warming is mis-represented by regional climate models with parameterised convec-  
225 tion<sup>76,77</sup>.

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

241 [Figure 6 about here.]

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

254 Thus, bias correction is trapped in a fundamental dilemma: in situations where the driving  
255 model simulates a credible change a trend preserving bias correction<sup>32,11</sup> is a sensible choice.  
256 In many cases, however, we may have strong evidence that the simulated regional climate  
257 change is implausible - we would like to improve the change. Standard quantile mapping  
258 modifies simulated trends. But as discussed above and demonstrated for the snow albedo

259 feedback, we know that these modifications may not be physically justified. Here, one would  
260 have to assess the raw and modified changes on a case-by-case basis, referring to the relevant  
261 climatic processes and their model representation.

## 262 **5 Ways Ahead**

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

### 275 **5.1 Understanding Model Biases**

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

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

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

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

310 There is realistic hope that further model improvements and increased model resolution  
311 may improve the representation of both local and large-scale processes<sup>81,54,82,58,83</sup>. The re-  
312 sulting reduction in location biases and the increase in credibility of future projections will  
313 render subsequent bias correction a more defensible approach.

## 314 5.2 New Bias Correction Approaches

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

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

325 As laid out above, a misrepresentation of regional feedbacks may result in an implausible  
326 regional climate change signal, and quantile mapping will likely not be able to improve it.  
327 Avenues should be explored to explicitly account for regional-scale processes and feedbacks  
328 for improving the climate change signal in the statistical postprocessing. One such avenue is,  
329 again, process-based bias correction. For instance, summer temperature biases may depend  
330 on temperature because of soil moisture feedbacks. Here it has been suggested to condition  
331 the correction on simulated soil moisture<sup>67</sup>. Another avenue are emulators of high-resolution  
332 RCMs, which simulate a credible climate change signal. For instance, local variations in the  
333 warming signal could be statistically expressed by covariates such as elevation, continentality  
334 or large-scale warming patterns. These expressions can be calibrated across a range of dy-  
335 namically downscaled GCMs, and then applied to statistically downscale the climate change  
336 signal of other GCMs<sup>85</sup>. Such emulators could also be developed for other regional processes  
337 such as convection: measures of stability and moisture convergence could serve as input to  
338 emulate high-resolution convection permitting models. Thereby the representation of extreme

339 events could be improved, a weak point of essentially all statistical post-processing methods  
340 so far.

### 341 **5.3 Evaluating Bias Correction**

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

### 354 **5.4 Alternative approaches**

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

## 364 **6 Final Remarks**

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

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

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

## 380 References

- 381 [1] B.C. Hewitson, J. Daron, R.G. Crane, M.F. Zermoglio, and C. Jack. Interrogating  
382 empirical-statistical downscaling. *Clim. Change*, 122:539–554, 2014.
- 383 [2] P. Adams, E. Eitland, B. Hewitson, C. Vaughan, R. Wilby, and S. Zebiak. Toward an  
384 ethical framework for climate services. A white paper of the climate services partnership  
385 working group on climate services ethics, 2015.
- 386 [3] G. Flato, J. Marotzke, B. Abiodun, P. Braconnot, S.C. Chou, W. Collins, P. Cox, F. Dri-  
387 ouech, S. Emori, V. Eyring, C. Forest, P. Gleckler, E. Guilyardi, C. Jakob, V. Kattsov,  
388 C. Reason, and M. Rummukainen. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovern-  
389 mental Panel on Climate Change*, chapter Evaluation of Climate Models. Cambridge  
390 University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.
- 391 [4] S. Kotlarski, K. Keuler, O.B. Christensen, A. Colette, M. Déqué, A. Gobiet, K. Goergen,  
392 D. Jacob, D. Lüthi, E. van Meijgaard, G. Nikulin, C. Schär, C. Teichmann, R. Vautard,  
393 K. Warrach-Sagi, and V. Wulfmeyer. Regional climate modelling on European scales: A  
394 joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geosci. Model. Dev. Discuss.*, 7:217–293, 2014.
- 395 [5] E.P. Maurer. Fine-resolution climate projections enhance regional climate change impact  
396 studies. *EOS*, 88(47):504, 2007.
- 397 [6] H. Li, J. Sheffield, and E.F. Wood. Bias correction of monthly precipitation and tem-  
398 perature fields from Intergovernmental Panel on Climate Change AR4 models using  
399 equidistant quantile matching. *J. Geophys. Res.*, 115:D10101, 2010.
- 400 [7] A. Dosio, P. Paruolo, and R. Rojas. Bias correction of the ENSEMBLES high resolution  
401 climate change projections for use by impact models: Analysis of the climate change  
402 signal. *J. Geophys. Res. Atmos.*, 117(D17), 2012.
- 403 [8] S. Hagemann, C. Chen, J.O. Haerter, J. Heinke, D. Gerten, and C. Piani. Impact of  
404 a statistical bias correction on the projected hydrological changes obtained from three  
405 GCMs and two hydrology models. *J. Hydrometeorol.*, 12(4):556–578, 2011.
- 406 [9] A.M.K Stoner, K. Hayhoe, X. Yang, and D.J. Wuebbles. An asynchronous regional  
407 regression model for statistical downscaling of daily climate variables. *Int. J. Climatol.*,  
408 33:2473–2494, 2013.
- 409  
410

- 411 [10] E.H. Girvetz, E.P. Maurer, P.B. Duffy, A. Ruesch, B. Thrasher, and C. Zganjar. Making  
412 climate data relevant to decision making: the important details of spatial and temporal  
413 downscaling. The World Bank, 2013.
- 414 [11] S. Hempel, K. Frieler, L. Warszawski, J. Schewe, and F. Piontek. A trend-preserving  
415 bias correction - the ISI-MIP approach. *Earth Syst. Dynam.*, 4:219–236, 2013.
- 416 [12] E.P. Maurer, L. Brekke, T. Pruitt, B. Thrasher, J. Long, P. Duffy, M. Dettinger,  
417 D. Cayan, and J. Arnold. An enhanced archive facilitating climate impact and adaptation  
418 analysis. *Bull. Amer. Meteorol. Soc.*, 95(7):1011–1019, 2014.
- 419 [13] CORDEX. Bias-adjusted RCM data. [http://www.cordex.org/index.php?option=com\\_content&view=](http://www.cordex.org/index.php?option=com_content&view=article&id=101)  
420 2016.
- 421 [14] S. Gangopadhyay, T. Pruitt, L. Brekke, and D. Raff. Hydrologic projections for the  
422 Western United States. *EOS*, 92(48):441–442, 2011.
- 423 [15] S. Hagemann, C. Chen, D. Clark, S. Folwell, S.N. Gosling, I. Haddeland, N. Hannasaki,  
424 J. Heinke, F. Ludwig, F. Voss, and A. Wiltshire. Climate change impact on available  
425 water resources obtained using multiple global climate and hydrology models. *Earth*  
426 *Syst. Dynam.*, 4:129–144, 2013.
- 427 [16] L. Warszawski, K. Frieler, V. Huber, F. Piontek, O. Serdeczny, and J. Schewe. The Inter-  
428 Sectoral Impact Model Intercomparison Project (ISIMIP): Project framework. *Proc. Nat.*  
429 *Acad. Sci.*, 111(9):3228–3232, 2014.
- 430 [17] D. Cayan, K. Kunkel, C. Castro, A. Gershunov, J. Barsugli, A. Ray, J. Overpeck, M. An-  
431 derson, J. Russell, B. Rajagopalan, I. Rangwala, and P. Duffy. *Assessment of Climate*  
432 *Change in the Southwest United States: A Report Prepared for the National Climate*  
433 *Assessment*, chapter Future climate: Projected average., pages 101–125. Island Press,  
434 2013.
- 435 [18] World Bank. Turn Down the Heat: Climate Extremes, Regional Impacts, and the Case  
436 for Resilience. A report for the World Bank by the Potsdam Institute for Climate Impact  
437 Research and Climate Analytics, 2013.
- 438 [19] A. Georgakakos, P. Fleming, M. Dettinger, C. Peters-Lidard, T.C. Richmond, K. Reck-  
439 how, K. White, and D. Yates. *Climate Change Impacts in the United States: The Third*  
440 *National Climate Assessment*, chapter Water Resources, pages 69–112. U.S. Global  
441 Change Research Program, 2014.
- 442 [20] Worldbank. Climate change knowledge portal. <http://sdwebx.worldbank.org/climateportal/>.
- 443 [21] prepdata. Partnership for Resilience and Preparedness (PREP). [www.prepdata.org](http://www.prepdata.org).
- 444 [22] D. Maraun, F. Wetterhall, A.M. Ireson, R.E. Chandler, E.J. Kendon, M. Widmann,  
445 S. Brienen, H.W. Rust, T.Sauter, M.Thiemeßl, V.K.C. Venema, K.P. Chun, C.M. Good-  
446 ess, R.G. Jones, C. Onof, M. Vrac, and I. Thiele-Eich. Precipitation downscaling under

- 447 climate change. Recent developments to bridge the gap between dynamical models and  
448 the end user. *Rev. Geophys.*, 48:RG3003, 2010.
- 449 [23] C. Teutschbein and J. Seibert. Bias correction of regional climate model simulations for  
450 hydrological climate-change impact studies: Review and evaluation of different methods.  
451 *J. Hydrol.*, 456:12–29, 2012.
- 452 [24] C. Piani, J.O. Haerter, and E. Coppola. Statistical bias correction for daily precipitation  
453 in regional climate models over Europe. *Theor. Appl. Climatol.*, 99(1-2):187–192, 2010.
- 454 [25] S. Vannitsem. Bias correction and post-processing under climate change. *Nonlin. Proc.*  
455 *Geophys.*, 18:911–924, 2011.
- 456 [26] J. Eden, M. Widmann, D. Grawe, and S. Rast. Skill, correction, and downscaling of  
457 GCM-simulated precipitation. *J. Climate*, 25:3970–3984, 2012.
- 458 [27] D. Maraun. Bias correction, quantile mapping and downscaling: Revisiting the inflation  
459 issue. *J. Climate*, 26:2137–2143, 2013.
- 460 [28] D. Maraun and M. Widmann. The representation of location by a regional climate model  
461 in complex terrain. *Hydrol. Earth Syst. Sci.*, 19:3449–3456, 2015.
- 462 [29] N. Addor, M. Rohrer, R. Furrer, and J. Seibert. Propagation of biases in climate models  
463 from the synoptic to the regional scale: Implications for bias adjustment. *J. Geophys.*  
464 *Res.*, 121(5):2075–2089, 2016.
- 465 [30] T.F. Stocker, Q. Dahe, G.-K. Plattner, and M. Tignor. IPCC Workshop on Re-  
466 gional Climate Projections and their Use in Impacts and Risk Analysis Studies.  
467 [https://www.ipcc.ch/pdf/supporting-material/RPW\\_WorkshopReport.pdf](https://www.ipcc.ch/pdf/supporting-material/RPW_WorkshopReport.pdf), 2015.
- 468 [31] U. Ehret, E. Zehe, V. Wulfmeyer, K. Warrach-Sagi, and J. Liebert. Should we apply  
469 bias correction to global and regional climate model data? *Hydrol. Earth Syst. Sci.*, 16,  
470 2012.
- 471 [32] J.O. Haerter, S. Hagemann, C. Moseley, and C. Piani. Climate model bias correction  
472 and the role of timescales. *Hydrol. Earth Syst. Sci.*, 15(3):1065–1079, 2011.
- 473 [33] F. Johnson and A. Sharma. A nesting model for bias correction of variability at multiple  
474 time scales in general circulation model precipitation simulations. *Wat. Resour. Res.*,  
475 48:W01504, 2012.
- 476 [34] P.-A. Michelangeli, M. Vrac, and H. Loukos. Probabilistic downscaling approaches: Ap-  
477 plication to wind cumulative distribution functions. *Geophys. Res. Lett.*, 36(11), 2009.
- 478 [35] C. Piani and J.O. Haerter. Two dimensional bias correction of temperature and precip-  
479 itation copulas in climate models. *Geophys. Res. Lett.*, 39(20):L20401, 2012.
- 480 [36] M. Vrac and P. Friederichs. Multivariate-intervariable, spatial, and temporal-bias cor-  
481 rection. *J. Climate*, 28(1):218–237, 2015.

- 482 [37] A.A.L. Levy, W. Ingram, M. Jenkinson, C. Huntingford, F.H. Lambert, and M. Allen.  
483 Can correcting feature location in simulated mean climate improve agreement on pro-  
484 jected changes? *Geophys. Res. Lett.*, 40:354–358, 2013.
- 485 [38] J.J. Barsugli, G. Guentchev, R. M. Horton, A. Wood, L.O. Mearns, X.-Z. Liang, J.A.  
486 Winkler, K. Dixon, K. Hayhoe, R.B. Rood, L. Goddard, A. Ray, L. Buja, and C. Am-  
487 mann. The practitioner’s dilemma: How to assess the credibility of downscaled climate  
488 projections. *EOS*, 94(46):424–425, 2013.
- 489 [39] C. Wang, L. Zhang, S.-K. Lee, L. Wu, and C.R. Mechoso. A global perspective on CMIP5  
490 climate model biases. *Nat. Clim. Change*, 4:201–205, 2014.
- 491 [40] J.H. Christensen, F. Boberg, O.B. Christensen, and P. Lucas-Picher. On the need for  
492 bias correction of regional climate change projections of temperature and precipitation.  
493 *Geophys. Res. Lett.*, 35:L20709, 2008.
- 494 [41] C.M. Buser, H.R. Künsch, D. Lüthi, M. Wild, and C. Schär. Bayesian multi-model projec-  
495 tion of climate: bias assumptions and interannual variability. *Clim. Dynam.*, 33:849–868,  
496 2009.
- 497 [42] D. Maraun. Nonstationarities of regional climate model biases in European seasonal  
498 mean temperature and precipitation sums. *Geophys. Res. Lett.*, 39:L06706, 2012.
- 499 [43] C. Teutschbein and J. Seibert. Is bias correction of regional climate model (rcm) simula-  
500 tions possible for non-stationary conditions? *Hydrol. Earth Syst. Sci.*, 17(12):5061–5077,  
501 2013.
- 502 [44] F. Johnson and A. Sharma. Accounting for interannual variability: A comparison of  
503 options for water resources climate change impact assessments. *Wat. Resour. Res.*,  
504 47:W04508, 2011.
- 505 [45] P. Berg, H. Feldmann, and H.-J. Panitz. Bias correction of high resolution regional  
506 climate model data. *J. Hydrol.*, 448-449:80–92, 2012.
- 507 [46] C. Chen, J.O. Haerter, S. Hagemann, and C. Piani. On the contribution of statistical  
508 bias correction to the uncertainty in the projected hydrological cycle. *Geophys. Res.*  
509 *Lett.*, 38:L20403, 2011.
- 510 [47] A. Dosio and P. Paruolo. Bias correction of the ENSEMBLES high resolution climate  
511 change projections for use by impact models: Evaluation on the present climate. *J.*  
512 *Geophys. Res. Atmos.*, 116(D16), 2011.
- 513 [48] M. J. Themeßl, A. Gobiet, and A. Leuprecht. Empirical-statistical downscaling and  
514 error correction of daily precipitation from regional climate models. *Int. J. Climatol.*,  
515 31:1530–1544, 2011.



- 516 [49] L. Gudmundson, J.B. Bremnes, J.E. Haugen, and T. Engen-Skaugen. Technical Note:  
517 Downscaling RCM precipitation to the station scale using statistical transformations a  
518 comparison of methods. *Hydrol. Earth Syst. Sci.*, 16:3383–3390, 2012.
- 519 [50] J. Rajczak, S. Kotlarski, and C. Schär. Does quantile mapping of simulated precipitation  
520 correct for biases in transition probabilities and spell lengths? *J. Climate*, 29:1605–1615,  
521 2016.
- 522 [51] G. Masato, B.J. Hoskins, and T. Woollings. Winter and summer northern hemisphere  
523 blocking in CMIP5 models. *J. Climate*, 26:7044–7059, 2013.
- 524 [52] T. Woollings. Dynamical influences on European climate: An uncertain future. *Phil.*  
525 *Trans. R. Soc. A*, 368:3733–3756, 2010.
- 526 [53] G. Zappa, L.C. Shaffrey, and K.I. Hodges. The ability of CMIP5 models to simulate  
527 North Atlantic extratropical cyclones. *J. Climate*, 26:5379–5396, 2013.
- 528 [54] P. Davini, J. von Hardenberg, S. Corti, H.M. Christensen, S. Juricke, A. Subramanian,  
529 P.A.G. Watson, A. Weisheimer, and T.N. Palmer. Climate SPHINX: evaluating the  
530 impact of resolution and stochastic physics parameterisations in the EC-Earth global  
531 climate model. *Geosci. Model Dev.*, 10(3):1383–1402, 2017.
- 532 [55] F. Pithan, T.G. Shepherd, G. Zappa, and I. Sandu. Climate model biases in jet streams,  
533 blocking and storm tracks resulting from missing orographic drag. *Geophys. Res. Lett.*,  
534 43:7231–7240, 2016. doi:10.1002/2016GL069551.
- 535 [56] A. van Niekerk, J.F. Scinocca, and T.G. Shepherd. The modulation of stationary waves,  
536 and their response to climate change, by parameterized orographic drag. *J. Atmos. Sci.*,  
537 74:2557–2574, 2017. doi:10.1175/JAS-D-17-0085.1.
- 538 [57] M. Ashfaq, C.B. Skinner, and N.S. Diffenbaugh. Influence of SST biases on future climate  
539 change projections. *Clim. Dynam.*, 36:1303–1319, 2011.
- 540 [58] A.A. Scaife, D. Copsey, C. Gordon, C. Harris, T. Hinton, S. Keeley, A. O’Neill,  
541 M. Roberts, and K. Williams. Improved Atlantic winter blocking in a climate model.  
542 *Geophys. Res. Lett.*, 38:L23703, 2011.
- 543 [59] S.P.E. Keeley, R.T. Sutton, and L.C. Shaffrey. The impact of North Atlantic sea surface  
544 temperature errors on the simulation of North Atlantic European region climate. *Q. J.*  
545 *R. Meteorol. Soc.*, 138:1774–1783, 2012.
- 546 [60] T.G. Shepherd. Atmospheric circulation as a source of uncertainty in climate change  
547 projections. *Nat. Geosci.*, 7:703–708, 2014.
- 548 [61] A. Hall. Projecting regional change. *Science*, 346(6216):1461–1462, 2014.
- 549 [62] A. Colette, R. Vautard, and M. Vrac. Regional climate downscaling with prior statistical  
550 correction of the global climate forcing. *Geophys. Res. Lett.*, 39(13):L13707, 2012.

- 551 [63] E.P. Maurer and D.W. Pierce. Bias correction can modify climate model simulated  
552 precipitation changes without adverse effect on the ensemble mean. *Hydrol. Earth Syst.*  
553 *Sci.*, 18(3):915–925, 2014.
- 554 [64] A. Gobiet, M. Suklitsch, and G. Heinrich. The effect of empirical-statistical correction  
555 of intensity-dependent model errors on the temperature climate change signal. *Hydrol.*  
556 *Earth Syst. Sci.*, 19:4055–4066, 2015.
- 557 [65] D.W. Pierce, D.R. Cayan, E.P. Maurer, J.T. Abatzoglou, and K.C. Hegewisch. Im-  
558 proved Bias Correction Techniques for Hydrological Simulations of Climate Change. *J.*  
559 *Hydrometeorol.*, 16(6):2421–2442, 2015.
- 560 [66] M.B. Switanek, P.A. Troch, C.L. Castro, A. Leuprecht, H.-I. Chang, R. Mukherjee, and  
561 E.M.C. Demaria. Scaled distribution mapping: a bias correction method that preserves  
562 raw climate model projected changes. *Hydrol. Earth Syst. Sci.*, 21:2649–2666, 2017.
- 563 [67] O. Bellprat, S. Kotlarski, D. Lüthi, and C. Schär. Physical constraints for temperature  
564 biases in climate models. *Geophys. Res. Lett.*, 40:4042–4047, 2013.
- 565 [68] H. Bellenger, E. Guilyardi, J. Leloup, M. Lengaigne, and J. Vialard. ENSO representation  
566 in climate models: from CMIP3 to CMIP5. *Clim. Dynam.*, 42:1999–2018, 2014.
- 567 [69] L. Chen, T. Li, and Y. Yu. Causes of strengthening and weakening of ENSO amplitude  
568 under global warming in four CMIP5 models. *J. Climate*, 28:3250–3274, 2015.
- 569 [70] M. Collins, R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao,  
570 W.J. Gutowski, T. Johns, G. Krinner, M. Shongwe, C. Tebaldi, A.J. Weaver, and  
571 M. Wehner. *Climate Change 2013: The Physical Science Basis. Contribution of Work-*  
572 *ing Group I to the Fifth Assessment Report of the Intergovernmental Panel on Cli-*  
573 *mate Change*, chapter Long-term Climate Change: Projections, Commitments and Ir-  
574 *reversibility*. Cambridge University Press, Cambridge, United Kingdom and New York,  
575 NY, USA, 2013.
- 576 [71] H. Annamalai, J. Hafner, K.P. Sooraj, and P. Pillai. Global warming shifts the monsoon  
577 circulation, drying South Asia. *J. Climate*, 26:2701–2718, 2013.
- 578 [72] M.K. Roxy, K. Ritika, P. Terray, R. Murtugudde, K. Ashok, and B.N. Goswami. Drying  
579 of indian subcontinent by rapid indian ocean warming and a weakening land-sea thermal  
580 gradient. *Nat. Commun.*, 6:7423, 2015.
- 581 [73] J. Willison, W.A. Robinson, and G.M. Lackmann. North atlantic storm-track sensitivity  
582 to warming increases with model resolution. *J. Climate*, 28:4513–4524, 2015.
- 583 [74] A. Hall, X. Qu, and J.D. Neelin. Improving predictions of summer climate change in the  
584 united states. *Geophys. Res. Lett.*, 35:L01702, 2008.
- 585 [75] J.H. Christensen and F. Boberg. Temperature dependent climate projection deficiencies  
586 in CMIP5 models. *Geophys. Res. Lett.*, 39:L24705, 2012.

- 587 [76] E.J. Kendon, N.M. Roberts, H.J. Fowler, M.J. Roberts, S.C. Chan, and C.A. Senior.  
588 Heavier summer downpours with climate change revealed by weather forecast resolution  
589 model. *Nat. Clim. Change*, 4:570–576, 2014.
- 590 [77] E.P. Meredith, D. Maraun, V.A. Semenov, and W. Park. Evidence for added value of  
591 convection permitting models for studying changes in extreme precipitation. *J. Geophys.*  
592 *Res. Atmos.*, 120:12,500–12,513, 2015.
- 593 [78] M. Collins, R.E. Chandler, P.M. Cox, J.M. Huthnance, J. Rougier, and D.B. Stephenson.  
594 Quantifying future climate change. *Nat. Clim. Change*, 2(6):403–409, 2012.
- 595 [79] I.R. Simpson, R. Seager, M. Ting, and T.A. Shaw. Causes of change in northern hemi-  
596 sphere winter meridional winds and regional hydroclimate. *Nat. Clim. Change*, 6(1):65–  
597 70, 2016.
- 598 [80] X. Qu and A. Hall. On the persistent spread in snow-albedo feedback. *Clim. Dynam.*,  
599 42(1-2):69, 2014.
- 600 [81] A. Dawson, T.N. Palmer, and S. Corti. Simulating regime structures in weather and  
601 climate prediction models. *Geophys. Res. Lett.*, 39(21), 2012.
- 602 [82] P. Liu and Coauthors. An MJO simulated by the NICAM at 14- and 7-km resolutions.  
603 *Mon. Wea. Rev.*, 137:3254–3268, 2009.
- 604 [83] A.F. Prein, W. Langhans, G. Fosser, A. Ferrone, N. Ban, K. Goergen, M. Keller, M. Tlle,  
605 O. Gutjahr, F. Feser, E. Brisson, S. Kollet, J. Schmidli, N.P.M. van Lipzig, and R. Le-  
606 ung. A review on regional convection-permitting climate modeling: Demonstrations,  
607 prospects, and challenges. *Rev. Geophys.*, 53(2):323–361, 2015.
- 608 [84] C. Volosciuk, D. Maraun, M. Vrac, and M. Widmann. A combined statistical bias  
609 correction and stochastic downscaling method for precipitation. *Hydrol. Earth Syst. Sci.*,  
610 21(3):1693–1719, 2017.
- 611 [85] D.B. Walton, F. Sun, A. Hall, and S. Capps. A hybrid dynamical-statistical downscaling  
612 technique. part i: Development and validation of the technique. *J. Climate*, 28(12):4597–  
613 4617, 2015.
- 614 [86] D. Maraun, M. Widmann, J.M. Gutierrez, S. Kotlarski, R.E. Chandler, E. Hertig,  
615 J. Wibig, R. Huth, and R.A.I. Wilcke. VALUE: A framework to validate downscaling  
616 approaches for climate change studies. *Earth’s Future*, 3:1–14, 2015.
- 617 [87] C. Prudhomme, R.L. Wilby, S. Crooks, A.L. Kay, and N.S. Reynard. Scenario-neutral  
618 approach to climate change impact studies: Application to flood risk. *J. Hydrol.*, 390:198–  
619 209, 2010.
- 620 [88] C. Schär, C. Frei, D. Lüthi, and H.C. Davies. Surrogate climate-change scenarios for  
621 regional climate models. *Geophys. Res. Lett.*, 23(6):669–672, 1996.

- 622 [89] W. Hazeleger, B.J.J.M. van den Hurk, E. Min, G.J. van Oldenborgh, A.C. Petersen, D.A.  
623 Stainforth, E. Vasileiadou, and L.A. Smith. Tales of future weather. *Nat. Clim. Change*,  
624 5:107–113, 2015.
- 625 [90] A.M.G. Klein Tank, J.B. Wijngaard, G.P. Können, R. Böhm, G. Demarée, A. Gocheva,  
626 M. Mileta, S. Pashiardis, L. Hejkrlik, C. Kern-Hansen, R. Heino, P. Bessemoulin,  
627 G. Müller-Westermeier, M. Tzanakou, S. Szalai, T. Pálsdóttir, D. Fitzgerald, S. Ru-  
628 bin, M. Capaldo, M. Maugeri, A. Leitass, A. Bukantis, R. Aberfeld, A.F.V. van Engelen,  
629 E. Forland, M. Miletus, F. Coelho, C. Mares, V. Razuvaev, E. Nieplova, T. Cegnar,  
630 J.A. López, B. Dahlström, A. Moberg, W. Kirchhofer, A. Ceylan, O. Pachaliuk, L.V.  
631 Alexander, and P. Petrovic. Daily dataset of 20th-century surface air temperature and  
632 precipitation series for the European Climate Assessment. *Int. J. Climatol.*, 22(12):1441–  
633 1453, 2002.
- 634 [91] J.M. Gutiérrez, R. Cano, A.S. Cofiño, and C. Sordo. Analysis and downscaling multi-  
635 model seasonal forecasts in Peru using self-organizing maps. *Tellus A*, 57(3):435–447,  
636 2005.
- 637 [92] E.P. Maurer, A.W. Wood, J.C. Adam, D.P. Lettenmaier, and B. Nijssen. A long-term  
638 hydrologically-based data set of land surface fluxes and states for the conterminous United  
639 States. *J. Climate*, 15(22):3237–3251, 2002.
- 640 [93] D.P. van Vuuren, J. Edmonds, M. Kainuma, K. Riahi, A. Thomson, K. Hibbard,  
641 G. Hurtt, T. Kram, V. Krey, J.-F. Lamarque, T. Masui, M. Meinshausen, N. Naki-  
642 cenovic, S.J. Smith, and S.K. Rose. The representative concentration pathways: an  
643 overview. *Clim. Change*, 109(1):5–31, 2011.

#### 644 **Corresponding author**

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

#### 646 **Acknowledgements**

647  
648 Many of the ideas laid out in this paper have been developed during discussions within the EU  
649 COST Action ES1102 “VALUE”, in particular during the workshop “Global climate model  
650 biases: causes, consequences and correctability” held at the Max Planck Institute for Meteo-  
651 rology, Hamburg, October 2014.  
652  
653

#### 654 **Author contributions**

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

658 **List of Figures**

659 1 **Cross-validation problem** | Quantile mapping from ERA40 daily boreal winter (DJF) temperature [ $^{\circ}\text{C}$ , 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

670 2 **Unrealistic dry spell lengths** | Distribution of dry spell lengths (wet-day threshold 0.1 mm) at **a**, Tafjord (Norway;  $7.41^{\circ}\text{W}$ ,  $62.23^{\circ}\text{N}$ , winter), **b**, Constanta (Romania;  $28.63^{\circ}\text{E}$ ,  $44.22^{\circ}\text{N}$ , winter), **c**, Sion (Switzerland,  $7.33^{\circ}\text{E}$ ,  $46.22^{\circ}\text{N}$ , winter) and **d**, Rome (Italy,  $12.58^{\circ}\text{E}$ ,  $41.78^{\circ}\text{N}$ , summer) of MPI-ESM-LR downscaled with CLM to a horizontal resolution of  $0.44^{\circ}$ , 1971-2000. Black: observations (ECA-D<sup>90</sup>), 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

680 3 **Non-representative model output** | Daily precipitation bias correction for the GISS-E2-R model against station data at Piura, Peru<sup>91</sup> 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

688 4 **Missing temperature inversions** | Distribution of spring (MAM) daily mean temperature differences between Fresno ( $\sim 90\text{m}$ ) and Three Rivers (70 km towards the southeast, at  $\sim 400\text{m}$ ) in California, US, 1981-2000. Blue: observations ( $1/8^{\circ}$  gridded data<sup>92</sup>), 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

698	5	<b>Large-scale circulation problems</b>   <b>a-c</b> , FGOALS-g2; <b>d-f</b> , MPI-ESM-MR. <b>a,d</b> , simulated (colour shading, mm/day) and observed (contour lines at 4 and 6 mm/day) mean winter precipitation 1976-2005. <b>b,e</b> , uncorrected mean precipitation averaged over 10W to 20E (vertical red lines in <b>a</b> and <b>d</b> ) from present and future (2070-2099, RCP8.5 <sup>93</sup> ) simulations. <b>c,f</b> , 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
699			
700			
701			
702			
703			
704			
705			
706			
707			
708			
709			
710			
711	6	<b>Implausible sub-grid climate change signal</b>   Spring (MAM) daily mean temperature [°C] in the Sierra Nevada and Central Valley, California, US. <b>a-c</b> , present climate (1981-2000 average); <b>d-f</b> , simulated change (2081-2100 average minus 1981-2000 average, RCP8.5 scenario <sup>93</sup> ). <b>a,d</b> , GFDL-CM3 GCM, bilinearly interpolated to 8km grid; <b>b,e</b> , corrected GCM (for present by construction identical with observations at 8km horizontal resolution <sup>92</sup> ); <b>c,f</b> , WRF RCM at 3km horizontal resolution, driven with GFDL-CM3 climate change signal <sup>85</sup> . 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
712			
713			
714			
715			
716			
717			
718			
719			
720			
721			

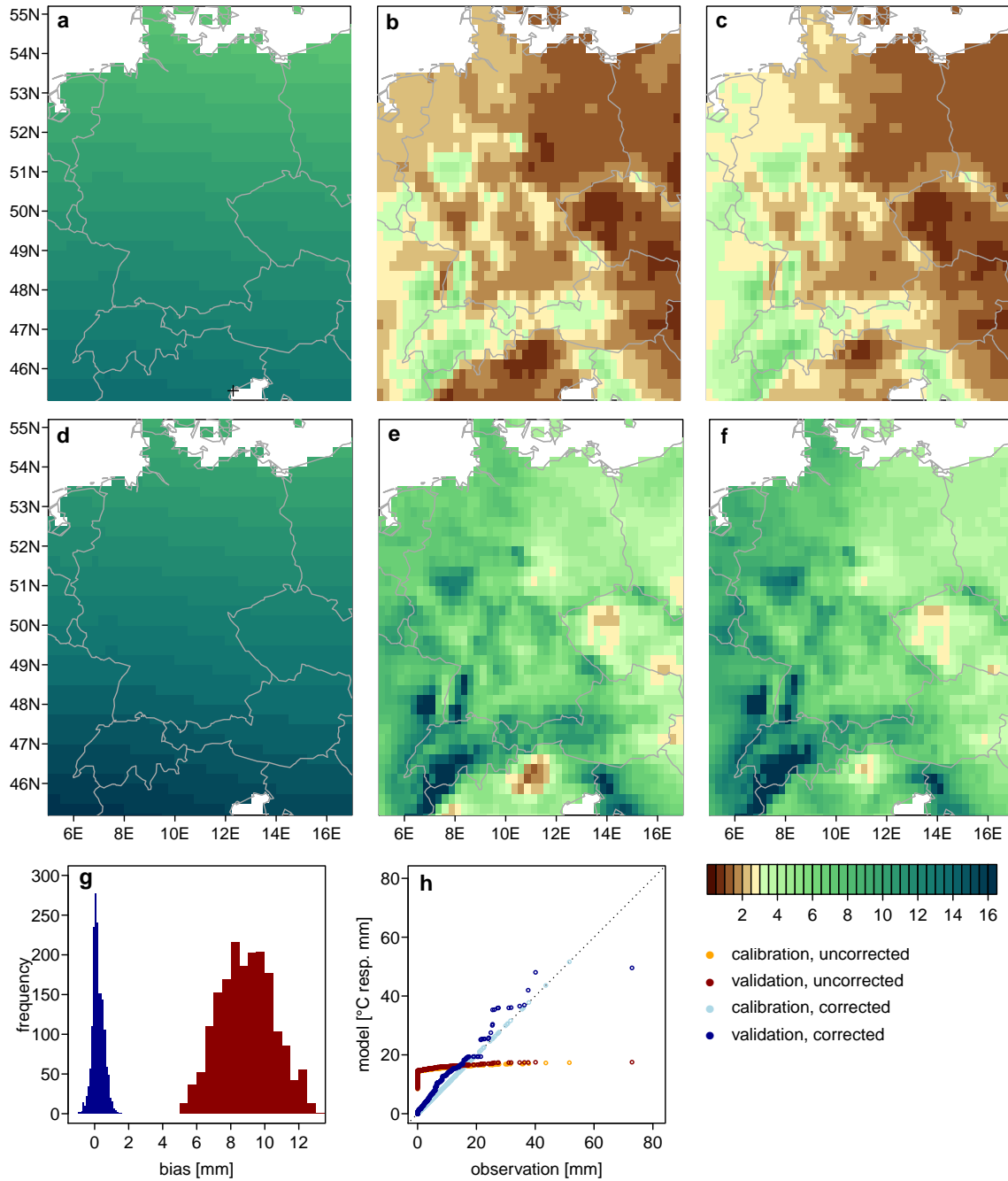


Figure 1: **Cross-validation problem** | Quantile mapping from ERA40 daily boreal winter (DJF) temperature [ $^{\circ}\text{C}$ , Southern Ocean, 45S-55S, 175W-163W] to E-OBS daily precipitation [mm/day, Central Europe, 45N-55N, 5E-17E], calibrated over 1961-1980. **a-c**, mean and **d-f**, 95th percentile 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.

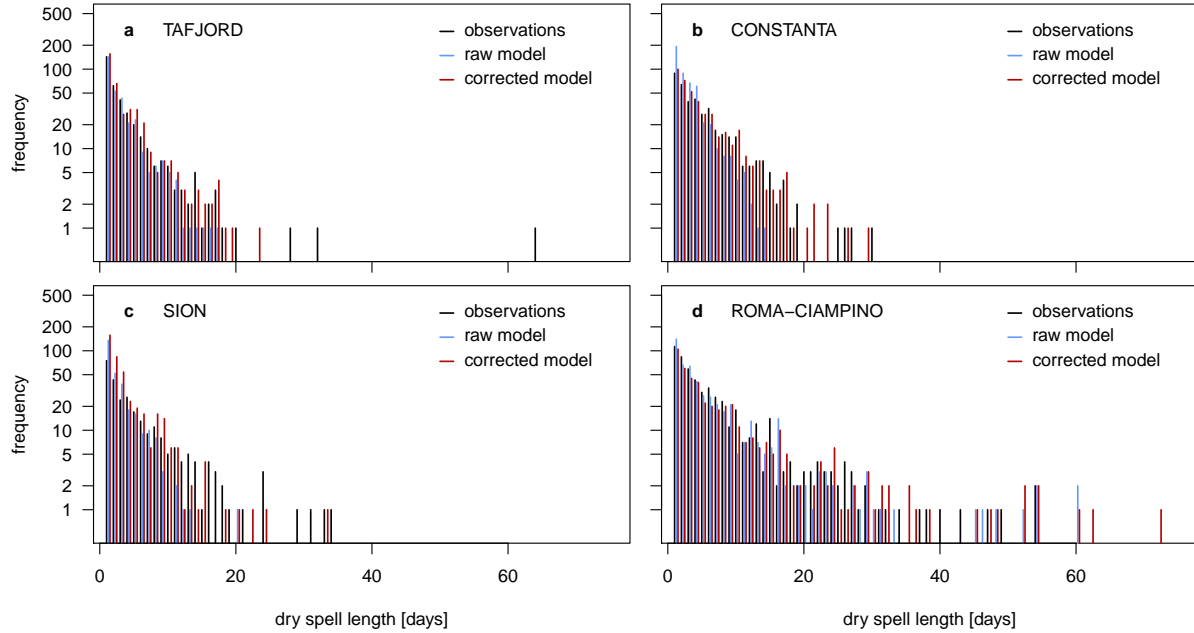


Figure 2: **Unrealistic dry spell lengths** | Distribution of dry spell lengths (wet-day threshold 0.1 mm) at **a**, Tafjord (Norway; 7.41° W, 62.23°N, winter), **b**, Constanta (Romania; 28.63° E, 44.22°N, winter), **c**, Sion (Switzerland, 7.33° E, 46.22°N, winter) and **d**, Rome (Italy, 12.58° E, 41.78°N, summer) of MPI-ESM-LR downscaled with CLM to a horizontal resolution of 0.44°, 1971-2000. Black: observations (ECA-D<sup>90</sup>), 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.



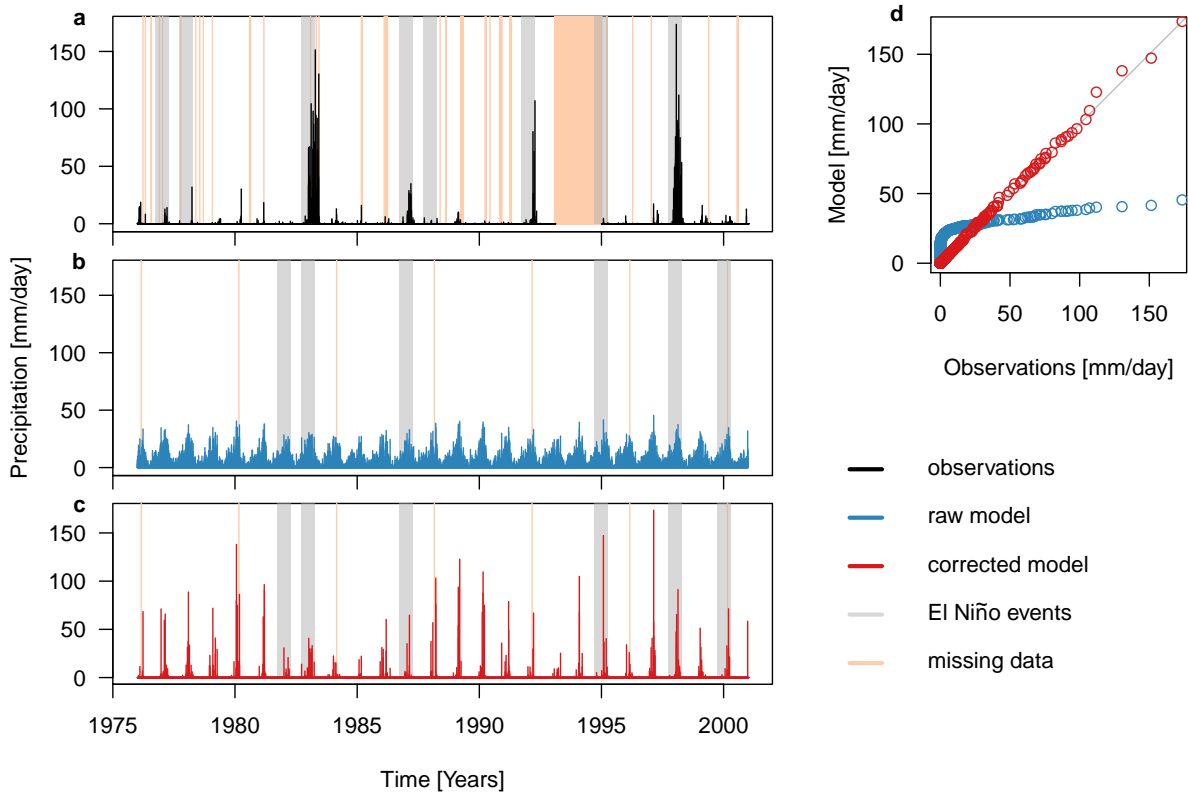


Figure 3: **Non-representative model output** | Daily precipitation bias correction for the GISS-E2-R model against station data at Piura, Peru<sup>91</sup> from 1976-2000. **a**, observations; **b**, raw GCM data; **c**, quantile mapped GCM data; **d**, QQ plot. Grey shading: El Niño events. As the 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.

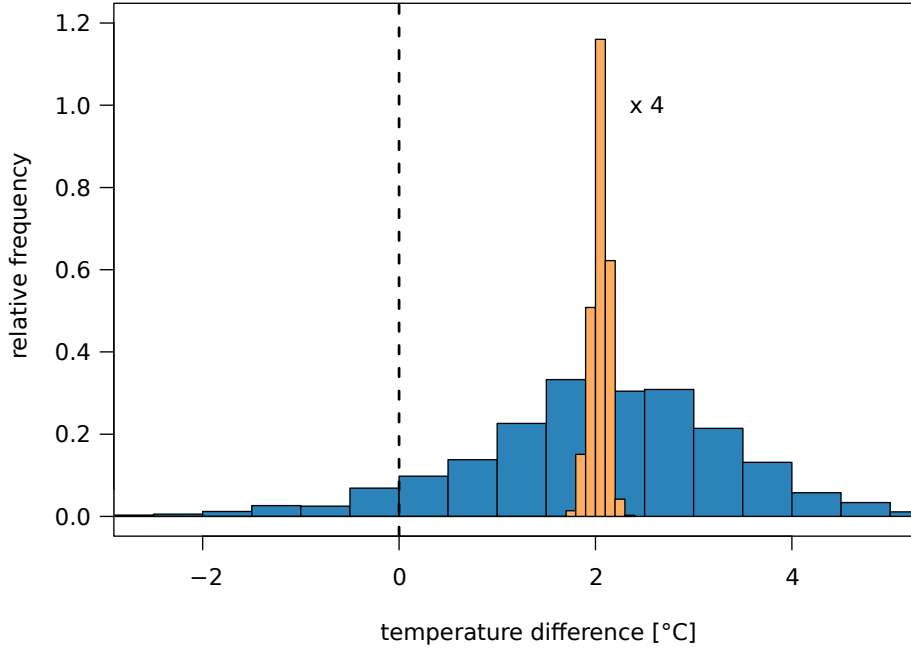


Figure 4: **Missing temperature inversions** | Distribution of spring (MAM) daily mean temperature differences between Fresno ( $\sim 90\text{m}$ ) and Three Rivers (70 km towards the south-east, at  $\sim 400\text{m}$ ) in California, US, 1981-2000. Blue: observations ( $1/8^\circ$  gridded data<sup>92</sup>), 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.

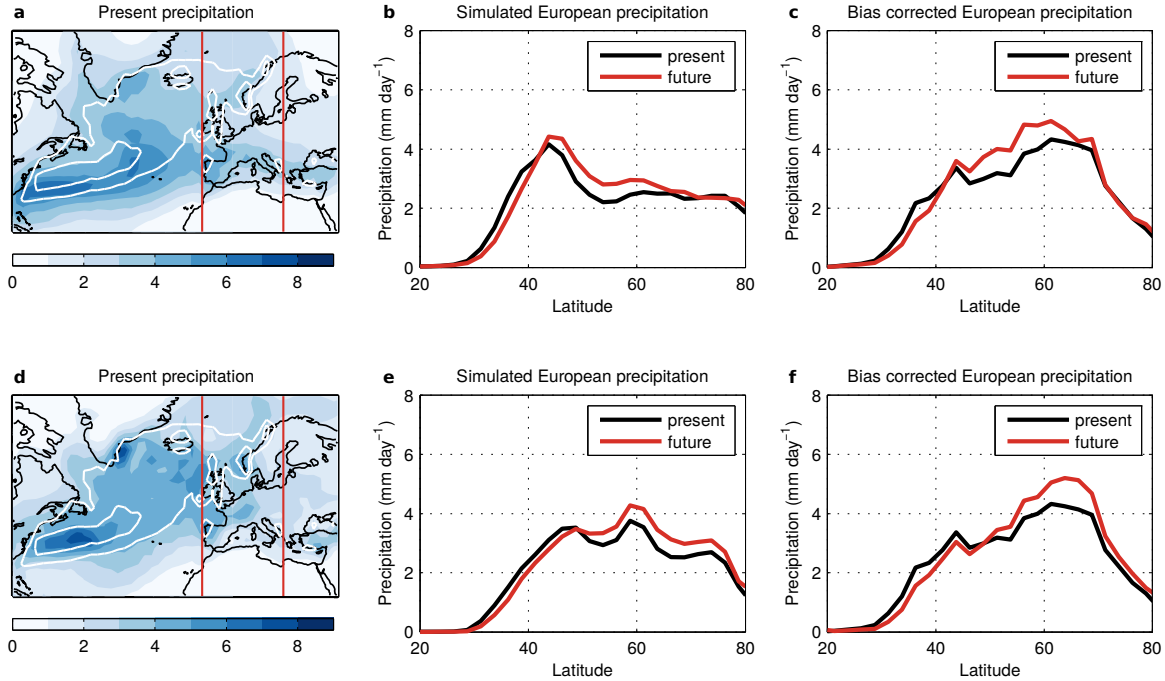


Figure 5: **Large-scale circulation problems** | a-c, FGOALS-g2; d-f, MPI-ESM-MR. a,d, simulated (colour shading, mm/day) and observed (contour lines at 4 and 6 mm/day) mean winter precipitation 1976-2005. b,e, uncorrected mean precipitation averaged over 10W to 20E (vertical red lines in a and d) from present and future (2070-2099, RCP8.5<sup>93</sup>) 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.

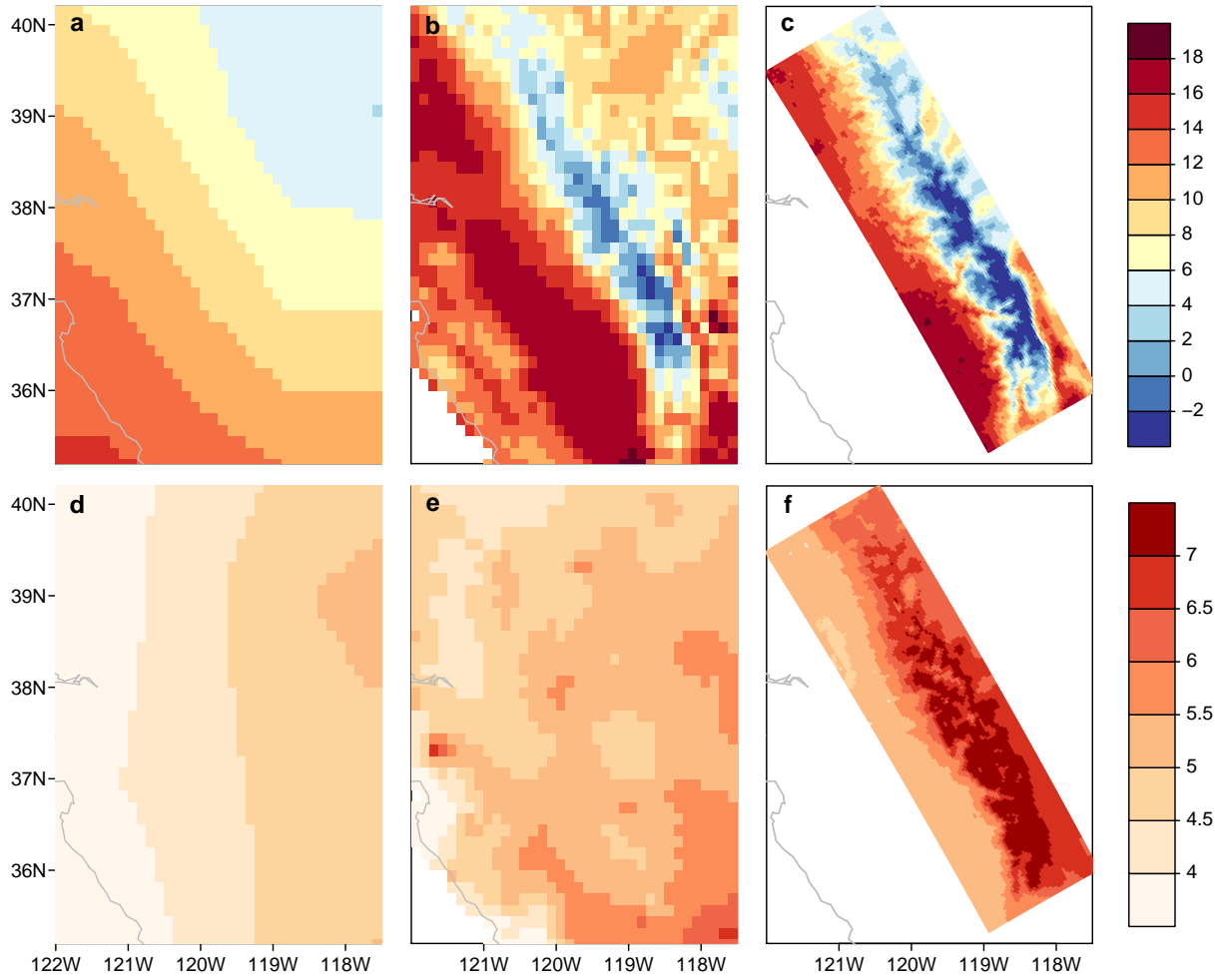


Figure 6: **Implausible sub-grid climate change signal** | Spring (MAM) daily mean temperature [ $^{\circ}\text{C}$ ] in the Sierra Nevada and Central Valley, California, US. **a-c**, present climate (1981-2000 average); **d-f**, simulated change (2081-2100 average minus 1981-2000 average, RCP8.5 scenario<sup>93</sup>). **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<sup>92</sup>); **c,f**, WRF RCM at 3km horizontal resolution, driven with GFDL-CM3 climate change signal<sup>85</sup>. 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.