Improved bias correction techniques for hydrological simulations of climate change

David W. Pierce[1,*], Daniel R. Cayan[1,2], Edwin P. Maurer[3], John T. Abatzoglou[4], Katherine C. Hegewisch[4]

[1]Division of Climate, Atmospheric Sciences, and Physical Oceanography, Scripps Institution of Oceanography, La Jolla, CA, 92093-0224

[2]U.S. Geological Survey, La Jolla, CA, 92037

[3]Civil Engineering Department, Santa Clara University, Santa Clara, CA, 95053-0563

[4]Department of Geography, University of Idaho, Moscow, ID, 83844

Version 2.0

18 June 2015

*Corresponding Author. Division of Climate, Atmospheric Sciences, and Physical Oceanography, Scripps Institution of Oceanography, La Jolla, CA, 92093-0224.

dpierce@ucsd.edu

Abstract

Global climate model output typically needs to be bias corrected before it can be used for climate change impact studies. Three existing bias correction methods, and a new one developed here, are applied to daily maximum temperature and precipitation from 21 global climate models (GCMs) to investigate how different methods alter the climate change signal of the GCM. The quantile mapping (QM) and cumulative distribution function transform (CDF-t) bias correction methods can significantly alter the GCM's mean climate change signal, with differences of up to 2°C and 30 percentage points for monthly mean temperature and precipitation, respectively. Equidistant quantile matching (EDCDFm) bias correction preserves GCM changes in mean daily maximum temperature, but not precipitation. An extension to EDCDFm termed PresRat is introduced, which generally preserves the GCM changes in mean precipitation. Another problem is that GCMs can have difficulty simulating variance as a function of frequency. To address this, a frequency-dependent bias correction method is introduced that is twice as effective as standard bias correction in reducing errors in the models' simulation of variance as a function of frequency, and does so without making any locations worse, unlike standard bias correction. Lastly, a preconditioning technique is introduced that improves the simulation of the annual cycle while still allowing the bias correction to take account of an entire season's values at once.

1. Introduction

Climate impact assessments can be sensitive to biases in global climate model (GCM) output (IPCC, 2013). For example, precipitation biases degrade hydrological simulations due to the non-linear nature of runoff: a moderate amount of precipitation generates little runoff if the soil can absorb the moisture, while doubling the precipitation generates more than twice the runoff if the moisture storage capacity of the soil is exceeded. This non-linear relationship becomes more extreme in arid regions (Wigley and Jones, 1985). Similarly, temperature biases can influence the partition of precipitation into snow or rain, affecting the snowpack and therefore the timing and magnitude of runoff over the entire year.

For this reason hydrological simulations generally use bias corrected GCM output. Bias correction is often an integral part of downscaling GCM output (e.g., Wood et al. 2002; Maurer et al. 2010). Here however we consider the bias correction step alone. Bias correction is best applied on a spatial scale near the original GCM's spatial resolution (Maraun, 2013), so we examine bias correction on a grid commensurate with the original GCMs.

Many bias correction methods have been used in climate impact studies. One widely used method is quantile mapping (QM; e.g., Panofsky and Brier 1968; Wood et al. 2002; Thrasher et al. 2012), which adjusts a model value by mapping quantiles of the model's distribution onto quantiles of the observations. QM has been applied to climate model output over both the U.S. (e.g., Maurer et al. 2007, 2014) and globally (Thrasher et al. 2012).

Previous studies have shown that QM alters the magnitude and even direction of mean changes projected from the original GCM (Hagemann et al. 2011; Pierce et al. 2013; Maurer and Pierce 2013). This can engender confusion and inconsistent results, for example between bias

corrected GCM output for regional climate studies and unadulterated GCM output evaluated by the IPCC (2007, 2013). If a climate model has too much variability, QM tends to reduce variability on all timescales, including the trend (Pierce et al. 2013; Maurer and Pierce 2013). If the GCM has too little variability, QM tends to increase the trend. Since bias correction is a purely statistical method, it fails to discriminate between the physical processes determining trends associated with anthropogenic forcing and shorter-term fluctuations associated with natural internal climate variability. From this perspective there is little justification for allowing bias correction that primarily addresses problems on synoptic, seasonal and annual timescales to change the trend as well.

Although the correct long-term future trend in climate variables is unknown, as witnessed by the IPCC's adoption of a "one model, one vote" policy for evaluating climate projections, in this work we choose to implement a bias correction scheme that does not alter the original GCM trend. This reduces the disparity between global model studies with a given GCM and regional models based on bias-corrected output from that GCM. Other options for how to interpret the long-term trend in a GCM that has incorrect short-timescale variability await further research.

Other bias correction methods include the cumulative distribution function transform (CDF-t) method (Michelangeli et al. 2009), which assumes that the historical mapping between the model and observed cumulative distribution functions applies to the future period, and equidistant quantile matching (EDCDFm; Li et al. 2010), which preserves the GCM-predicted change at each quantile evaluated additively (i.e., as the future minus historical value). However changes in precipitation are often more usefully evaluated as multiplicative changes, since a fixed amount of precipitation change has different implications in wet and arid regions. We show that EDCDFm alters the GCM-predicted mean precipitation change (evaluated

multiplicatively), and CDF-t alters both the model-predicted temperature and precipitation changes. The first goal of this work is to show that a straightforward extension to EDCDFm, which we term “PresRat” (because it preserves the ratio) can retain the model-predicted future change in mean precipitation evaluated as a ratio (cf. Wang and Chen, 2014).

GCM biases in temporal variance can also pose problems for impact modeling. For example, a model might have too much variability on synoptic timescales yet too little on annual timescales, making it challenging to represent the proper magnitude and spectra of phenomena such as droughts. Although simulations have improved with the CMIP5 models, deficiencies still remain in representing regional variability on interannual to decadal timescales (Sheffield et al., 2013). QM, CDF-t, and EDCDFm do not address this problem. Such biases could influence the simulation of heat waves or flooding events, with consequences for agriculture, ecosystems, droughts, or reservoir simulations. The second goal of this work is to describe a method that reduces frequency-dependent climate model biases.

Lastly, bias correction is typically implemented in a time window, often of about a month long. Choosing an appropriate time window involves compromises between correcting the annual cycle, reducing discontinuities at the edge of the time window, and evaluating extreme values over an entire season. The third goal of this work is to show that a simple preconditioning technique together with iteratively applied bias correction can improve the final corrected seasonal cycle, while still allowing a seasonal time window and reducing discontinuities at the window’s edges.

The rest of this work is structured as follows. In section 2 we describe the observed and model data sources we use to evaluate the bias correction schemes. Section 3 addresses the problem of bias correction altering model-predicted changes, and proposes an extension to the

EDCDFm bias correction scheme that preserves model-predicted mean future changes in precipitation. Section 4 addresses frequency-dependent model biases, documents the extent to which these are seen in the current generation of global climate models, and proposes a method for reducing these biases. Section 5 shows how simple preconditioning together with an iterative bias correction scheme can improve the representation of the annual cycle and reduce bias measured in different windows. A summary and conclusions are given in section 6.

2. Data sources and time periods

2.1 Global climate models

We use daily maximum temperature and precipitation fields from 21 GCMs that participated in the Coupled Model Intercomparison Project, version 5 (CMIP5; Taylor et al., 2012), listed in Table 1. The models used are all those available from the U.S. Bureau of Reclamation (USBR) archive of regridded (1°x1° longitude-latitude) CMIP5 global climate models at the time this work was performed (ftp://gdo-dcp.ucllnl.org/pub/dcp/archive/cmip5/bcca; Maurer et al. 2014). GCM output was obtained from both historical (1950-2005) runs and future (2006-2099) runs using representative concentration pathway 8.5 (RCP8.5).

2.2 Observations

We used observed daily maximum temperature and precipitation data from Maurer et al. (2002), as updated through 2010 (available from http://www.engr.scu.edu/~emaurer/gridded_obs/index_gridded_obs.html). The ultimate source of this gridded product is the NOAA co-operative observer weather stations, with techniques from the PRISM project (Daly et al. 1994) used to adjust observed precipitation values to match

long-term PRISM climatology. The data come on a 1/8° x 1/8° latitude-longitude grid, which we aggregated to the same 1°x1° grid as the GCM output.

2.3 Time periods

The World Meteorological Organization (WMO) recommends that climatological normals be calculated over 30-year periods (Trewin 2007). We follow this guidance by bias correcting GCM values to a 30-yr climatological record of observations, and furthermore by bias correcting contiguous 30-yr segments of climate simulations individually. A different segment length could be used, subject to two opposing considerations: 1) The segments should be long enough to provide a reasonable estimate of the climatological normals, given natural internal climate variability; 2) The segments should be short enough that the statistical characteristics of the variable being downscaled are reasonably stationary over the period being downscaled. We used 30 years as a compromise for these two criteria.

For the future model projections we bias correct the periods 2010-2039, 2040-2069, and 2070-2099 separately. In the results shown below we focus on 2070-2099 as our "future" period. The climatological (historical) period is the last 30 years of the GCMs' historical runs (1976-2005), used for both the models and observations. We bias correct and evaluate the models over the same historical period (1976-2005) so that difference between the bias corrected results and observations is known to be due to the bias correction itself, rather than due to differences in climate between the historical period and an independent verification period (cf. Teutschbein and Seibert, 2012). This differs from, for example, downscaling, where an independent period is typically used to evaluate the downscaled results.

3. Preserving model-predicted mean changes

We evaluate temperature changes as a difference (future minus historical) and precipitation changes as a ratio (future / historical). This is unlike Maurer and Pierce (2013), which evaluated precipitation changes as a difference. However, evaluating precipitation changes as a ratio can be useful since a fixed amount of precipitation change has different implications in an arid region than in a wet region.

The present work explores three approaches to bias correction: preserving the mean model-predicted change, reducing frequency-dependent biases, and preconditioning and reducing biases in different time windows. If all approaches were implemented simultaneously it would be difficult to distinguish the influence of each procedure on the resultant change. In this section we use standard monthly bias correction (all January values are bias corrected together, etc.) excluding frequency-dependent bias correction or preconditioning.

3.1 Effect of QM, CDF-t, and EDCDFm on model-predicted changes

3.1.1 Quantile Mapping

Quantile mapping (QM; Panofsky and Briar, 1968; Wood et al. 2002) bias corrects a model value by changing it to the observed value at the quantile that the model value falls in the model's historical distribution. The process is illustrated schematically in Figure 1a, using CDFs of synthetic gamma distributions to mimic precipitation.

Averaged across the 21 GCMs, QM exaggerates monthly mean model-projected warming (2070-2099 minus 1976-2005) in the Rockies in January and diminishes it in July (Figure 2a). Maurer and Pierce (2013) showed why QM alters the GCM trend when model variance is biased; briefly, if the model's variance is incorrect, QM alters the trend as it corrects the variance. Figure 2a shows multi-model mean values, but the modification in any individual model can be much

174 greater. The RMS spread across the 21 models is shown in Figure 2b. The spread is appreciable
175 using the QM technique, with RMS values of up to 2 °C, and more spread is found in the warmer
176 months.

177 Figure 3 shows a similar analysis for precipitation, evaluated multiplicatively in terms of
178 percentage change. QM tends to make the original model-predicted mean change wetter over the
179 Northwestern U.S. in January and California in July. The RMS spread across models is ~25
180 percentage points in parts of the Northwest in January, and exceeds 60 percentage points in the
181 dry California/Great Basin region in July.

182 *3.1.2 CDF-t*

183 CDF-t bias correction (Michelangeli et al. 2009) finds a transformation that maps the
184 GCM cumulative distribution function (CDF) of a climate variable in the historical period to the
185 observed CDF, then applies that same mapping to the GCM's future CDF. The process is
186 illustrated schematically in Figure 1c. When bias correcting a historical run CDF-t reduces to
187 QM, although the treatment of values off the end of the distribution (discussed below) comes
188 into play.

189 The second columns of Figure 2 and Figure 3 show that CDF-t modifies the original
190 monthly mean temperature projection less than QM, but still on the order of 0.5 °C. CDF-t tends
191 to make the precipitation projections drier, which can be understood in terms of Figure 1c. To
192 produce a point on the bias corrected future distribution (dotted green line) it is necessary that the
193 model historical value at the quantile being bias corrected fall within the range of observed
194 values, as indicated by vector (2) in Figure 1c. As vector (2) progressively moves to the right in
195 Figure 1c, at higher quantiles it becomes impossible to map future changes beyond the maximum
196 observed value. In this event, following Michelangeli et al. (2009), the correction used is that

found at the maximum valid historical value. However in climate projections the precipitation distribution changes shape such that the most extreme events increase preferentially (e.g., IPCC 2007, 2013). In this situation CDF-t uses a correction that falls at a lower quantile and so misses the preferential increase in the highest quantiles.

3.1.3 EDCDFm

EDCDFm (Li et al. 2010) bias corrects a future value *x* that falls at quantile *u* in the future distribution by adding the historical value at *u* to the model predicted change in value at *u*. The process is illustrated schematically in Figure 1b (note the non-linear X axis when considering the length of the Δ vector). When bias correcting a model historical run, EDCDFm reduces to QM.

EDCDFm preserves the GCM-predicted median change evaluated additively, but not necessarily the mean change since the quantile at which the mean falls can change in the future. However, for daily maximum temperature, GCM-predicted changes are generally a weak function of quantile in the neighborhood of the mean value, so EDCDFm preserves the model-predicted change in mean temperature to within a few hundredths of a degree C (third column of Figure 2).

As expected, EDCDFm does not preserve GCM-predicted fractional changes, i.e., (future model value − historical model value)/(historical model value). At every quantile EDCDFm preserves the numerator of this ratio, but in the process of bias correction substitutes the observed value for the historical model value in the denominator, changing the ratio. This is illustrated in the third column of Figure 3. EDCDFm alters the original model-predicted mean precipitation change by more than 30 percentage points in the dry (rain shadow) parts of the

Northwest U.S. This will happen particularly when there are both large biases and large changes in the upper quantiles of a skewed precipitation distribution.

3.2 Bias correction that preserves model-predicted mean changes

Given the same GCM input, QM, EDCDFm, and CDF-t produce different future temperature and precipitation fields, and it is not obvious which one is correct. QM assumes that the historical model error in value at a given *value* is preserved in the future (arrow (2) in Figure 1a), EDCDFm assumes that the historical model error in value at a given *quantile* is preserved in the future (Δ in Figure 1c), and CDF-t assumes that the historical model error in *quantile* at a given quantile is preserved in the future (arrow (2) in Figure 1b). (The "missing" version of this quartet of bias correction methods, which would assume that the historical model error in quantile at a given *value* is preserved in the future, could also be constructed.)

Here we explore an alternative assumption: that the GCM-predicted mean change is preserved in the bias corrected future projections. EDCDFm already preserves model-predicted mean change in temperature (evaluated additively) for all practical purposes, so we adopt it for temperature. However an amended form is required for precipitation since we evaluate its changes multiplicatively. If the predicted GCM value x falls at quantile u, then the bias corrected precipitation value is the historical value at u multiplied by the model-predicted change at u evaluated as a ratio (i.e., model future precipitation / model historical precipitation). This preserves the model-predicted median (not mean) change evaluated multiplicatively. In fact, Li et al. (2010) do this for a small number (~0.3%) of grid points that otherwise are "problematic" when bias correcting precipitation additively, although they did not explore the implications of preserving a model-predicted mean future precipitation change. Also, Wang and Chen (2014) adopt this ratio-based approach for bias correcting precipitation, although their stated reason is to

avoid the negative precipitation values that might arise when using additive factors. This scheme cannot be applied at quantiles with no precipitation, in which case we set the model-predicted change ratio to 1.

Applying EDCDFm with model-predicted change ratios is only part of the solution to preserving the original model-predicted mean change, because the quantile at which the mean falls can change between the historical and future period if the shape of the distribution changes. Although this results in negligible errors in temperature, precipitation distributions are more skewed and GCMs can show significantly varying projections of future change as a function of quantile. However, the mean precipitation change can be preserved exactly if the bias corrected value is multiplied by a correction factor $K = \langle x \rangle / \langle \hat{x} \rangle$, where x is the change (expressed as a ratio) in mean precipitation from the GCM, $\hat{x}$ is the change in mean precipitation following bias correction, and brackets indicate that the mean is taken over all days in the temporal window (monthly here).

The treatment of zero-precipitation days is an important consideration for regional climate change (Polade et al. 2014). At each grid cell we calculate a location-specific zero-precipitation threshold, τ, such that applying τ makes the model's number of zero-precipitation days match observations over the historical period. We require τ >= 0.01 mm/day to avoid the possibility of very small denominators in the model-predicted change ratio. Current GCMs tend to precipitate too frequently, often at daily amounts above 0.01 mm, so this limit is rarely invoked. The GCM-predicted future fraction of zero-precipitation days, Z_{gf}, is calculated using τ with the GCM's original (non-bias corrected) future time series. The model data is then bias corrected, and the smallest Z_{gf} fraction of precipitation values are set to zero. This preserves the model-predicted change in fraction of non-precipitating days, even if it increases. However if the

model has a strong dry bias, so that it has many more zero-precipitation days than observed, the model predicted change in zero precipitation days may not be preserved since there is no way to know which of the extra zero-precipitation days should be set to a positive value.

We call the combination of using the model-predicted change ratio, the treatment of zero precipitation days outlined above, and the final correction factor, the PresRat bias-correction method because it preserves the mean GCM-predicted future mean precipitation change evaluated as a ratio. Figure 1d includes results from PresRat applied to the synthetic example data (purple line).

Corrections that PresRat requires to maintain the model-predicted mean precipitation change are second order, arising from changes in the percentile at which the mean falls combined with differing model-predicted changes at different percentiles, and so tend to be modest. Figure 4 shows K for four different months averaged across all 21 GCMs. In any given month using the model change ratio alone tends to alter the model-predicted mean change by less than 5% in most of the region. In some places though, especially California in the summer, PresRat requires substantial corrections to preserve the model-predicted mean change.

By construction, PresRat preserves the model-projected mean precipitation change almost exactly (rightmost column of Figure 3). Discrepancies only arise due to problems with the model's number of zero-precipitation days, as noted above.

In summary, both temperature and precipitation can be bias corrected using methods that preserve GCM-predicted future mean changes. Doing so helps minimize confusion and inconsistent results between downscaled regional climate simulations and global model analyses, such as in the IPCC reports (2007, 2013). This also means that model-predicted mean changes

can be subsequently downscaled if desired (cf. Wood et al. 2002, who remove the mean GCM change before downscaling and then add it back afterwards).

4. Frequency Dependent Bias Correction

4.1 Overview

The effect of bias correction on model-predicted trends is a special case of the effect of bias correction on variability evaluated at long (multidecadal) timescales. We now address the more general question of model biases at different timescales and how to reduce them.

Details of our spectral approach are given in Appendix A. In brief, the model variance is compared to observations in 100 logarithmically spaced frequency bins. A digital filter is then applied in frequency space to make the model spectrum better match observations. One caveat is that we do not consider frequency-dependent biases in different seasons or months, only as a whole over the entire time period. This potentially means that it is not feasible to expect a removal of biases across all timescales of interest by this technique (e.g., bias correcting 2-10 day timescale temperature biases in winter and summer separately).

Since we bias correct in 30-yr periods (section 2.3), the PresRat method will preserve model-predicted mean changes at periods of 30 years and longer in the future projection. Accordingly we consider, at most, periods from two days (the Nyquist frequency given daily model output) to 30 years. This interval is further refined to two days to 11 years in light of our spectral analysis technique (Appendix A).

4.2 Frequency dependent model errors

Figure 5 shows the observed (1976-2005) distribution of variance in daily maximum temperature across frequencies (labeled using equivalent periods; left column), and the multi-

309 model mean errors in representing this distribution (middle column). The right column shows
310 multi-model RMSE (i.e., at each point, the spread of values across the 21 models). The
311 frequency-dependent bias correction is based on normalized spectra (spectral values divided by
312 the variance of the original time series) so that it leaves the overall variance unaltered. Therefore
313 at every location the values in the left hand column summed across frequency bands totals 100%.
314 The annual cycle (9-15 months band) dominates daily maximum temperature variability
315 over almost all of the conterminous U.S., containing on average 62% of the variance. The main
316 exceptions are along the California coast, Florida, and in a strip of the central U.S. downwind of
317 the Rockies, where higher frequencies (< 9 months) contribute more than elsewhere.
318 Models allocate less of the total variance to periods shorter than 9 months than observed.
319 In the 10-30 day band, the mean error reaches −9% (not shown). The proportion of variance in
320 the annual cycle is represented with little mean error and spread across models. Conversely,
321 models allocate more of the total variance to periods longer than 30 months, with nearly ~40%
322 more variance than observed, and the spread across models is large. However, the fraction of
323 total variance in these long time scales is small (< 1%).
324 Figure 6 shows the same analysis using daily precipitation. Periods between 2 and 10
325 days contain the majority of the variance (~62%). The exception is the west coast, where 10 day
326 to 9 month variability is nearly as important, and the annual cycle contains > 7% of the total
327 variance. The models have a 5-10% mean bias towards too much short-period (2-10 day)
328 variability along the west coast and upper Midwest, and too little variability in the southern Great
329 Plains and the Gulf coast. Model-simulated precipitation variability at 30 months or longer
330 accounts for an anomalously large proportion of the total variance in the southeastern U.S., and
331 an anomalously small proportion in the Pacific Northwest. Such errors could arise from, for

example, misrepresentations of the frequency, strength, or teleconnections of ENSO or other low-frequency modes of natural climate variability. Rupp et al. (2013) also found that models overestimate temperature variance and underestimate precipitation variance at timescales longer than a year in the Pacific Northwest. Disagreements across the models are large at these longer periods.

4.3 Frequency dependent bias correction

To reduce the frequency-dependent model biases, the ratio σ of the model's variance spectrum to the observed variance spectrum in the historical run is computed in each of the 100 logarithmically spaced frequency bins. The model time series is then transformed to frequency space, and the amplitude of the Fourier components are multiplied by $\sigma(f)^{-1/2}$ (the square root accounts for the fact that variance is proportional to the amplitude of the Fourier components squared). The result is then transformed back to the time domain. Basing the corrections on the historical run means that model-predicted future changes in the spectrum are retained, but assumes (like all statistical approaches) that model errors in the historical period are present in the future simulation as well. A more detailed illustration of the frequency-dependent bias correction process is given in the supplementary material, section S1.

Even standard bias correction techniques such as QM, EDCDFm, and CDF-t alter the spectra of the time series they are applied to. To isolate the effect of the frequency dependent bias correction (FDBC), we first present results using only FDBC, then examine combined results using FDBC and standard bias correction.

Example results of the FDBC using daily maximum temperature from the CCSM4 GCM are illustrated at a location in central Nevada (hot, dry) and a location in western Washington State (cool, wet) in the top panels of Figure 7. The error in the model's representation of the

355 spectrum of variability decreases substantially after FDBC is applied (i.e., green circles in the
356 right column of Figure 7 are much closer to 1).

357 It is useful to define an RMSE metric appropriate for ratios, which we designate as log-
358 RMSE to differentiate it from standard RMSE measures more appropriate to differences. Let
359 $\epsilon = \ln \sigma$; then

$$\text{log-RMSE} \equiv \exp(\langle \epsilon^2 \rangle^{1/2}) - 1 \quad (1)$$

360 where the angle brackets indicate the mean over the logarithmically spaced frequency values.
361 This expression treats equal ratios of error equally (i.e., the model having twice the observed
362 variance produces the same error as the observations having twice the model's variance), and the
363 final −1 makes a perfect result (model variance equal observed, so $\sigma = 1$) give a log-RMSE of 0.
364 In general, if the model values are incorrect (on average across log-spaced frequencies) by a
365 factor of σ, then the log-RMSE is $\sigma - 1$. These log-RMSE values are indicated in the right
366 column of Figure 7. When we refer to log-RMSE below, we specifically mean the model's error
367 in reproducing the distribution of variance across frequencies, as illustrated in Figure 7.

368 Precipitation is more difficult to correct in frequency space than temperature because it
369 cannot have negative values, which limits the adjustments FDBC can produce. There are also
370 days with zero precipitation, and to avoid exacerbating the models' drizzle problems (Sun et al.
371 2006; Dai 2006) we leave unmodified any values less than 1 mm/day. In dry areas this can leave
372 few days for FDBC to operate upon.

373 Precipitation results at the two example locations are shown in the bottom panels of
374 Figure 7. CCSM4 shows a much stronger than observed annual cycle at the hot dry location,
375 likely related to the coarse model overestimating winter precipitation in the Sierra Nevada rain

shadow. The log-RMSE values show that despite the limitations inherent in correcting precipitation, errors decrease after FDBC.

The multi-model ensemble average log-RMSE for daily maximum temperature is shown in the top row of Figure 8 both before (left column) and after (middle column) FDBC. The models' spectra systematically disagree with the observations, particularly along the west coast and in a band extending north from northern Texas. Before FDBC the mean log-RMSE is 0.50; after FDBC the log-RMSE drops to 0.11.

Results for daily precipitation are shown in the bottom row of Figure 8. The models do worse in the Rocky Mountains and Great Basin than elsewhere. As expected for the reasons given above, precipitation is less easily corrected than temperature; the mean log-RMSE for precipitation drops by less than a factor of 2 after FDBC.

The histograms in Figure 8 (right column) show the difference between each grid cells's corrected and original log-RMSE, pooled across every location and model. On average FDBC decreases the log-RMSE for daily maximum temperature by 0.39, and no locations are worse. Even for precipitation, which shows less improvement than temperature, the correction virtually always decreases the log-RMSE.

Histograms of the amplitude of the corrections pooled across all models and locations are shown in Figure 9. Any day's maximum temperature is changed less than 3°C about 95% of the time, although rarely the changes can exceed 4°C. The change in precipitation is less than 40% or 1.5 mm day^{-1} about 95% of the time, although on rare occasion can be more than 50% or 2.5 mm day^{-1}. Since FDBC operates on normalized spectra, altering the distribution of variance across frequencies without altering the overall variance, the mean changes are approximately zero.

399 *4.3.1 Combined effects of standard and frequency-dependent bias correction*

400 The frequency dependent bias correction (FDBC) is implemented using normalized
401 spectra so that the overall variance of the input time series are unchanged, since the technique is
402 intended to be used in conjunction with standard bias correction. We evaluated FDBC in
403 conjunction with quantile mapping (QM) since we want to compare the bias corrected results to
404 observations, which are only available over the historical period. This in turn restricts this
405 analysis to QM since the other bias correction methods differ from QM only in the future period.

406 For daily maximum temperature, the models' domain-average log-RMSE is 0.50 (Figure
407 8, upper left). Using QM alone decreases this to 0.35, while using FDBC alone decreases this to
408 0.11. The best results are obtained by using QM followed by FDBC, which not only preserves
409 the decrease in log-RMSE, but makes no points in the domain worse. QM alone worsens the log-
410 RMSE at 9.6% of the grid cells.

411 For daily precipitation, the models' domain-average log-RMSE is 0.49, which drops to
412 0.36 using QM alone, and 0.28 using FDBC alone. Using QM followed by FDBC gives the best
413 result, a log-RMSE of 0.24. In this case 1.3% of the grid cells end up having a worse log-RMSE,
414 which is still much better than the 22.9% of grid cells that are worsened by QM alone or the
415 4.5% of cells worsened by FDBC followed by QM. This small but consistent superiority when
416 applying QM before FDBC is the reason we perform the operations in this order.

417 To evaluate the effect of FDBC on runoff in a hydrological simulation, we used the VIC
418 hydrological model (Liang et al. 1994), configured for the western U.S. and forced over the
419 period 1950-1999 with four sources of daily temperatures and precipitation: 1) observations
420 (Livneh, 2013); 2) the CCSM4 GCM; 3) CCSM4 fields bias corrected using QM (since this is a
421 historical simulation); 4) CCSM4 fields with QM and FDBC. We define the model error in

simulating runoff variability in a frequency band as the log (base 10) of the ratio of the spectral power of runoff found using the GCM forcing fields to the spectral power found using the observations. An error of +1 means the model has 10x too much spectral power in a given frequency band, while -1 means 10x too little power. Figure 10a shows that when driven by CCSM4 fields, VIC overestimates low-frequency runoff variance by more than an order of magnitude over much of the interior southwest, a result of CCSM4's overly strong precipitation in the region. Bias correction (Figure 10b) improves the simulation markedly, while FDBC (Figure 10c) improves it somewhat more. Averaged across points in the domain, the mean error after bias correction is greatest at highest frequencies (Figure 10d, black line), and FDBC reduces the mean error at nearly all frequencies (red line), and overall by about a factor of 2 compared to bias correction alone.

5. Pre-conditioning and iterative bias correction

Bias correction is typically applied in a time window. For example, it can be applied monthly, so all January values are bias corrected together, then all February values, etc., as in Wood et al. (2002) and Maurer et al. (2010). However monthly bias correction of daily data potentially has discontinuities at the edges of the time window (e.g., Jan 31 is corrected using information from Jan 1, which is 30 days away, but no information from Feb 1, which is only one day away). To reduce these discontinuities Thrasher et al. (2010) use a moving-window approach, where bias correction is applied on a single day-of-year at a time using pooled values from a surrounding 31 day time window as training data for better sampling.

A drawback to using a time window of a month is that many weather extremes can occur anytime over a multi-month season. For example, the 20 highest values of California-averaged

444 daily precipitation over the period 1930-2002 have occurred as early as November and as late as
445 February, while extreme hot days have occurred as early as June and as late as September.
446 Ideally, the largest model value would be bias corrected to the largest observed value even if the
447 maximum fell at the beginning of the season in the observations and the end of the season in the
448 model. This argues for using a time window that is no narrower than a multi-month season if the
449 extremes are distributed over a season. (Of course, if the variable being bias corrected truly does
450 have all its extreme values fall in a single month of the year, then a single-month time window is
451 appropriate.) A more complete illustration of the problems obtained when using a 31-day sliding
452 time window is given in the supplementary material, section S2.

453 In this work we apply bias correction over a 91-day window, chosen to be wide enough
454 to encompass seasonal weather phenomena. To address the issue of discontinuities at the edges
455 of time windows, we iteratively apply the bias correction two additional times, with windows of
456 181 and 365 days, respectively. This ensures that every day is bias corrected with at least some
457 information from adjoining days no matter where it falls in the initial 91-day window. A similar
458 approach, dubbed “nested” bias correction, was adopted by Johnson and Sharma (2012),
459 although they used it for a different purpose than is done here. We use fixed, non-overlapping
460 time windows rather than moving ones to avoid the complications of matching quantiles in
461 datasets with greatly different sizes. For example, consider the case described above of bias
462 correcting a single central day-of-year using training data from the surrounding 31-day window,
463 and the whole processes is moved through the year. In a 50-year record the training data will
464 consist of 50*31=1550 days while the data to be corrected will consist of only 50 days. It is not
465 straightforward to match the most extreme event in a 50-event record to the most extreme event
466 in a 1550-event record.

The disadvantage to using a season-long time window is that the correction of the annual cycle worsens. Bias correction techniques such as QM, CDF-t, EDCDFm, and PresRat cannot rearrange the input time series' corresponding rank time series (i.e., the time series of the rank of each value, where rank 1 is largest value in the time series, etc.). Instead, they change the association of ranks to values. Fixing a distorted simulation of the annual cycle requires rearranging the rank time series. For example, imagine that January is climatologically colder than February (the average rank of February days is less than the average January rank), but the model has this relationship reversed. Fixing this error requires rearranging the rank time series.

The traditional approach to this problem is to apply bias correction in a relatively narrow time window. For example, using a simple monthly window ensures that the monthly means will be correct. However this does not address the discontinuities at the edges of the time window, nor the desirability of including all extreme values over an entire season when remapping the model distribution to the observed distribution.

In our bias correction process, we precede the primary bias correction with a simple "preconditioning" step designed to correct the annual cycle. The bias correction can then be applied to a time series that has a rank order consistent with the observed annual cycle. For precipitation, every day's value is multiplied by the ratio of the observed to model climatological value for that day of the year, where the climatologies are calculated over the historical period to allow changes in the future. For temperature, the preconditioning operates on the daily anomaly with respect to the period being downscaled. The model anomaly is multiplied by the ratio of the observed to model climatological standard deviation for that day (calculated over the historical period so it can change in the future), then added to the observed climatological value for that day (thus adjusting the annual cycle) plus the model-projected change in climatological value for

that day (to allow for future temperature changes). Since estimating a daily climatology from 30-year records is noisy, the daily values are cubic spline interpolated between 15-day averages. This preconditioning is a basic form of bias correction, but would be unsatisfactory if applied alone since it corrects only on the mean value and, for temperature, the variance. Following the preconditioning by QM, CDF-t, EDCDFm, or PresRat addresses extreme values as well, which are of great societal importance.

The effects of preconditioning on the annual cycle are illustrated using the CCSM4 GCM in Figure 11, which shows the RMSE difference between the observed and model simulated annual cycle of daily precipitation at each grid cell over the period 1976-2005. (The analogous figure for daily maximum temperature, which typically has a stronger annual cycle than precipitation, is shown in supplementary material Fig. S1.) Values are normalized by the annual mean at each point so that errors in arid and wet regions can be more easily compared. To reduce noise, the annual cycles are filtered with a 31-day boxcar filter before the RMSE is calculated. The original model has appreciable errors in the annual cycle (panel a), which are reduced with a simple monthly bias correction (panel b). Correcting a day at a time based on statistics of a surrounding 31-day window yields the least error (panel c). Using either a single 91-day window or our iterative approach with 91, 181, and 365-day windows gives mediocre results since the wide windows are less able to correct errors in the annual cycle, as described above (panels d and e). However, preconditioning helps substantially (panel f), giving a result with less error than monthly BC although somewhat more than with the sliding central day/31-day window approach.

The annual cycle is important, but many societal impacts are affected more by extreme events. Figure 12 shows a scatterplot of sorted daily precipitation values in the CCSM4 GCM

513 and observations at a point in the central Sierra Nevada (37.5 °N, −119.5 °W; 1976-2005). In a
514 perfect model values would fall along the diagonal (grey). Before bias correction (panel a), the
515 model under-represents the strongest events by a factor of 2. Simple monthly bias correction
516 (panel b) and using the central day in a 31-day wide sliding window (panel c) improve the
517 representation considerably, but still with errors. Using a wide bias correction window gives
518 good agreement between the observed and model-simulated extrema (panels d and e).
519 Preconditioning, which addresses the annual cycle rather than the extremes, has little effect on
520 this measure (panel f).

521 Summary statistics of the modeled representation of extremes at every grid cell can be
522 obtained by fitting a line between the top 5 observed and model extremes (dashed red lines in
523 Figure 12). The slopes and intercepts of the lines at all locations can then be mapped (Figure 13).
524 A perfect model representation of extremes would give a slope of 1 and intercept of 0. By this
525 measure, the original model (panels a, b) has appreciable errors in its representation of daily
526 extremes, as does the model after bias correction using either simple monthly BC (panels c, d) or
527 BC using a central day in a sliding 31-day window (panels e, f). Using a wider, 91-day window
528 improve the representation considerably (panels g, h), and iterating over the 91, 181, and 365-
529 day windows gives excellent agreement between the model and observations (panels i, j).

530 In summary, bias correction techniques that map one distribution to another are not
531 optimally suited for correcting the annual cycle. The traditional solution of applying the
532 correction in time windows of about a month is not necessarily a good fit with weather extremes,
533 which in many locations can occur anytime in a multi-month season. To get around this problem,
534 we use a simple preconditioning step that improves the representation of the annual cycle along
535 with a relatively wide (91-day) time window for bias correction, and iterate the bias correction

twice (181- and 365-day windows) to reduce discontinuities at the edges of the window. The overall result yields a representation of the annual cycle that is superior to simple monthly bias correction and a distribution of extremes that agrees well with observations over the training period.

6. Summary and Conclusions

GCMs generally produce biased simulations of variables such as temperature and precipitation. It is necessary to remove these biases before using the model-simulated fields in applications that have non-linear sensitivities to biases, such as land surface or hydrological modeling.

The choice of bias correction method is particularly important in climate change impact studies since bias correction can alter GCM projected mean changes. We demonstrate that quantile mapping (QM; Panofsky and Brier, 1968) or the CDF transform method (CDF-t; Michelangeli et al., 2009) can alter the original GCM-projected monthly mean change by up to 2°C when bias correcting temperature and 30 percentage points when bias correcting precipitation. This introduces a source of uncertainty comparable to uncertainty from emission scenarios in some cases. The EDCDFm method (Li et al., 2010) preserves GCM changes in mean temperature, but not changes in mean precipitation measured multiplicatively (as a ratio or percentage change). We introduced an extension to EDCDFm for precipitation termed PresRat that preserves the model-projected percentage change in mean precipitation by using a model-predicted change ratio (as in Wang and Chen 2014), but also a final correction factor and a zero-precipitation threshold that makes the modeled number of zero-precipitation days match observations. However none of the bias correction techniques, PresRat included, can preserve the

558 model-predicted mean precipitation change in locations that are so dry there are insufficient
559 precipitation days to bias correct.
560 We also examined the more general issue of the models' representation of variance
561 across a range of timescales, and introduced a frequency-dependent bias correction method that
562 reduces inaccuracies in the GCMs' spectra. As a group, the 21 GCMs apportion too little
563 variability of daily maximum temperature to times scales between 10 and 90 days and too much
564 to time scales longer than 30 months. The models' simulation of daily precipitation variability
565 was more mixed, but at long timescales (> 30 months) they show more variability than observed
566 in the Gulf coast region and less than observed in the Pacific Northwest. These problems can be
567 reduced by a frequency-dependent bias correction implemented as digital filter in the frequency
568 domain. This is one step towards addressing time-dependent model biases, an important subject
569 that has many implications for impacts such as droughts and heat waves. We implement the
570 frequency dependent bias correction as a separate step following the EDCDFm or PresRat bias
571 correction, which means this step could be combined with any other existing bias correction
572 method (such as quantile mapping or CDF-t) as well. However the current implementation
573 operates on the entire time series of daily values, so frequency-dependent errors on the seasonal
574 or monthly timescale can persist under some circumstances.
575 Traditional bias correction is done in a time window, often of about a month, to reduce
576 errors in the annual cycle. However in many locations weather extremes can occur sometime
577 during a multi-month season, which argues for using a time window on the order of a season in
578 such places. A simple preconditioning technique has been shown to yield a good simulation of
579 the seasonal cycle even when using a season-wide time window. The end result captures both the
580 extremes of the time series and the annual cycle.

This study has not addressed whether bias correction *should* be applied at any particular location given that model-observational disagreements are influenced by natural climate variability, which can be large and affect climate means over years to decades (e.g., Maraun et al. 2010; Deser et al. 2012). Although this is an interesting question, in this work we have followed the common practice of applying bias correction to the GCMs at all locations to bring them into agreement with a pre-selected recent climatological period.

In the end, as global climate model results continue to be applied to investigate phenomena that are sensitive to model biases, bias correction will become an ever more important step. The bias correction methods outlined here can improve these simulations, giving a clearer picture of future climate conditions for a variety of applications.

Acknowledgements

This work was sponsored by the California Energy Commission under contract CEC-500-10-041. Additional support for DWP and DC came from the USGS through the Southwest Climate Science Center, and from NOAA through the California Nevada Climate Applications Project (CNAP) Regional Integrated Science Applications (RISA) program. We also thank Ethan Gutmann of NCAR and an anonymous reviewer for thoughtful comments on an earlier version of this work, and the U.S. Bureau of Reclamation for making available the library of regridded climate projections at http://gdo-dcp.ucllnl.org.

Appendix A: Details of spectral approach

Ghil et al. 2002 review some of the numerous techniques that are available to compute variance spectra. Many newer methods have been developed to identify narrow-band signals

against a background of noise. However, in this work we are also concerned with the power in the broad parts of the spectrum that might in other applications be considered simply “noise”. This variability represents weather and climate fluctuations that affect hydrology and ecosystems across a wide range of time scales. Accordingly we use relatively wide bandwidths and employ the Jenkins and Watts (1969) method of computing variance spectra as the Fourier transformation of the autocovariance function. We require at least 40 degrees of freedom in the spectral estimates, which given 30 years of daily data and a Parzen lag window, means truncating the autocovariance function after 1020 lags (Jenkins and Watts 1969). Following the Jenkins and Watts recommendations the number of frequencies is set to twice the number of lags (2040), so the first non-zero frequency corresponds to a period of ~11 yrs. Longer periods are unresolved, and the frequency-dependent bias correction does not alter their relative proportion of variance.

With over 2000 frequencies spanning from 2 days to 11 years it is useful to reduce the number of frequencies at which the model error is corrected to avoid spurious over-fitting. Accordingly, the frequency-dependent model errors are calculated in a reduced set of 100 frequency bins of equal width in the logarithm of frequency. This means that higher frequency bins have multiple samples. All periods shorter than ~80 days have at least 5 samples per bin, reaching 140 samples at a period of 2 days. Averaging in bins therefore reduces the uncertainty in the spectral estimates for periods shorter than ~80 days.

Von Storch and Zwiers (2001) note the problems in interpreting spectral plots on a logarithmic frequency axis, since the displayed area under the spectrum is no longer proportional to the variance. It is possible to maintain the property of being a spectral density if the spectral value is multiplied by frequency, or if the plotted values are integrated (as opposed to averaged)

625 across constant widths of the logarithmic frequency axis. However these approaches change the
626 angle of a plotted spectrum (for example, a white spectrum is then no longer flat), which can be
627 confusing. To avoid this potentially misleading situation, values shown here are simply averaged
628 in frequency so that the spectra appear similar to what is typically found in the literature (i.e., a
629 white spectrum is flat).

630 # References

631 Dai, A., 2006: Precipitation characteristics in eighteen coupled climate models. J. Climate, 19,
632 4605-4630.

633 Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A Statistical-Topographic Model for Mapping
634 Climatological Precipitation over Mountanious Terrain. J. Appl. Meteor., 33, 140-158.

635 Deser, C., Knutti, R., Solomon, S., and Phillips, A. S., 2012: Communication of the role of
636 natural variability in future North American climate, Nature Clim. Change, 2, 775-779.

637 Ghil, M., M. R. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, M. E. Mann, A. W. Robertson,
638 A. Saunders, Y. Tian, F. Varadi, and P. Yiou, 2002: Advanced spectral methods for
639 climatic time series. Rev. Geophys., 40, doi:10.1029/2000RG000092.

640 Hagemann, S., C. Chen, J. O. Haerter, J. Heinke, D. Gerten, and C. Piani, 2011.: Impact of a
641 Statistical Bias Correction on the Projected Hydrological Changes Obtained from Three
642 GCMs and Two Hydrology Models, J. Hydrometeorology, 12, 556-578,
643 10.1175/2011jhm1336.1.

644 IPCC, 2007: Climate Change 2007: Impacts, Adaptation, and Vulnerability. Contribution of
645 Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on

Climate Change, M. L. Parry, O. F. Canziani, J. P. Palutikof, P. J. van der Linden, and C. E. Hanson, Eds., Cambridge University Press, Cambridge, UK, 976 pp.

IPCC, 2013: Climate Change 2013: The physical science basis. Working group I contribution to the IPCC fifth assessment report, T. Stocker, Q. Dahe, and G-K Plattner, coordinating lead authors. Available from http://www.ipcc.ch/report/ar5/wg1/.

Jenkins, G. M., and D. G. Watts, 1969: Spectral analysis and its applications. Holden-Day, 525 pp.

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

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I, 2010.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 48, RG3003, 10.1029/2009rg000314.

Maraun, D., 2013: Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue, J. Climate, 26, 2137-2143, 10.1175/jcli-d-12-00821.1.

Maurer, E. P., A. W. Wood, J. C. Adam, and D. P. Lettenmaier, 2002: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. J. Climate, 15, 3237-3251.

667 Maurer, E. P., L. Brekke, T. Pruitt, and P. B. Duffy, 2007: Fine-resolution climate projections
668 enhance regional climate change impact studies. EOS, Transactions of the American
669 Geophysical Union, 88, 504, doi:10.1029/2007EO470006.
670 Maurer, E. P., H. G. Hidalgo, T. Das, M. D. Dettinger and D. R. Cayan, 2010: The utility of
671 daily large-scale climate data in the assessment of climate change impacts on daily
672 streamflow in California. Hydrol. Earth Syst. Sci., 14, 1125-1138, doi:10.5194/hess-14-
673 1125-2010.
674 Maurer, E. P., T. Das, and D. R. Cayan, 2013: Errors in climate model daily precipitation and
675 temperature output: time invariance and implications for bias correction. Hydrol. Earth
676 Syst. Sci., 17, 2147-2159. Doi:10.519/hess-17-2147-2013.
677 Maurer, E. P., and D. W. Pierce, 2013: Bias correction can modify climate model-simulated
678 precipitation changes without adverse effect on the ensemble mean. Hydrol. Earth Syst.
679 Sci. Discuss., 10, 11585-11611, doi:10.5194/hessd-10-11585-2013.
680 Maurer, E. P., Brekke, L., Pruitt, T., Thrasher, B., Long, J., Duffy, P. B., Dettinger, M. D.,
681 Cayan, D., and Arnold, J, 2014: An enhanced archive facilitating climate impacts
682 analysis, Bulletin of the American Meteorological Society, doi:10.1175/BAMS-D-1113-
683 00126.00121 *in press*.
684 Michelangeli, P.-A., M. Vrac, and H. Loukos, 2009: Probabilistic downscaling approaches:
685 Application to wind cumulative distribution functions. Geophys. Res. Lett., 36, L11708,
686 doi:10.1029/2009GL038401.
687 Panofsky, H. A. and Brier, G. W.: Some Applications of Statistics to Meteorology, The
688 Pennsylvania State University, University Park, PA, USA, 224 pp., 1968.

689 Pierce, D.W., T. Das, D.R. Cayan, E.P. Maurer, N.L. Miller, Y. Bao, M. Kanamitsu, K.
690 Yoshimura, M.A. Snyder, L.C. Sloan, G. Franco and M. Tyree, 2013: Probabilistic
691 estimates of future changes in California temperature and precipitation using statistical
692 and dynamical downscaling. Clim. Dyn., 40, 839-856. doi:10.1007/s00382-012-1337-9.
693 Polade, S. D., D. W. Pierce, D. R. Cayan, A. Gershunov, and M. D. Dettinger, 2014: The key
694 role of dry days in regional climate change. *In preparation.*
695 Rupp, D. E., J. T. Abatzoglou, K. C. Hegewisch, and P. W. Mote, 2013: Evaluation of CMIP5
696 20th century climate simulations for the Pacific Northwest U.S. J. Geophys. Res. Atmos,
697 *in press.*
698 Sheffield, J., Langenbrunner, B., Meyerson, J. E., Neelin, J. D., Camargo, S. J., Fu, R., Hu, Q.,
699 Jiang, X., Karnauskas, K. B., Kim, S. T., Kumar, S., Kinter, J., Maloney, E. D., Mariotti,
700 A., Pan, Z., Ruiz-Barradas, A., Nigam, S., Seager, R., Serra, Y. L., Sun, D.-Z., Wang, C.,
701 Yu, J.-Y., Johnson, N., Xie, S.-P., Zhang, T., and Zhao, M., 2013: North American
702 Climate in CMIP5 Experiments. Part II: Evaluation of Historical Simulations of Intra-
703 Seasonal to Decadal Variability, J. Climate, 10.1175/jcli-d-12-00593.1.
704 Sun, Y., S. Solomon, A. Dai, and R. W. Portmann, 2006: How often does it rain? J. Climate, 19,
705 916-934.
706 Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An Overview of CMIP5 and the experiment
707 design. Bull. Am. Met. Soc., 93, 485-498, doi:410.1175/BAMS-D-1111-00094.00091.
708 Thrasher, B., E. P. Maurer, C. McKellar, and P. B. Duffy, 2012: Technical Note: Bias correcting
709 climate model simulated daily temperature extremes with quantile mapping. Hydrol.
710 Earth Syst. Sci., 16, 3309-3314. Doi:10.5194/hess-16-3309-2012.

711 Trewin, B. C, 2007: The role of climatological normals in a changing climate. World
712 Meteorological Organization Technical Directive WCDMP-No. 61 and WMO-TD No.
713 1377. Geneva, Switzerland. Available at
714 http://www.wmo.int/datastat/documents/WCDMPNo61_1.pdf.
715 Teutschbein, C., and J. Seibert, 2012: Bias correction of regional climate model simulations for
716 hydrological climate-change impact studies: Review and evaluation of different methods.
717 J. Hydrol., 456-457, 12-29.
718 von Storch, H., and F. W. Zwiers, 2001: Statistical analysis in climate research. Cambridge
719 University Press, 484 pp.
720 Wang, L., and W. Chen, 2014: A CMIP5 multimodel projection of future temperature,
721 precipitation, and climatological drought in China. Int. J. Climatol., 34, 2059-2078.
722 Wood, A. W., E. P. Maurer, A. Kumar, and D. P. Lettenamier, 2002: Long-range experimental
723 hydrologic forecasting for the eastern United States. J. Geophys. Res. D., 107, 4429,
724 doi:10.1029/2001JD000659, 2002.
725 Wood, A. W., Leung, L. R., Sridhar, V., and Lettenmaier, D. P., 2004: Hydrologic implications
726 of dynamical and statistical approaches to downscaling climate model outputs, Climatic
727 Change, 62, 189-216.
728 Wigley, T. M. L., and Jones, P. D.: Influences of precipitation changes and direct CO2 effects on
729 streamflow, Nature, 314, 149-152, 1985.

730 **Figure Captions**

731 Figure 1. Cumulative distribution functions (CDFs) of synthetic daily precipitation data
732 schematically illustrating how each bias correction method constructs the model's bias corrected

733 future CDF (green dotted/dashed lines). The solid blue, grey, and red lines are the same in all
734 panels and show the observed (1976-2005), model historical (1976-2005), and model future
735 (2070-2099) CDFs, respectively. The example point being corrected is X=30 mm/day, which
736 falls at the 0.56 quantile in the model future distribution (dotted orange line). a) Quantile
737 mapping (QM): starting at the point to be corrected, go vertically to the grey line (1),
738 horizontally to the blue line (2), and vertically to the original percentile (3). b) Equidistant CDF-
739 matching (EDCDFm): at the quantile of the point being corrected, compute the offset from the
740 model historical value to the model future value (Δ), then add Δ to the observed value at the
741 percentile being corrected (1). c) The CDF-transform (CDF-t) method; starting at the point to be
742 corrected, go horizontally to the grey line (1), vertically to the blue line (2), and horizontally to
743 the original value (3). d) Final results from all 3 bias correction methods (dotted/dashed green
744 lines), along with the PresRat method (solid purple line) for comparison. Note that the X axis
745 uses a square root transformation and the Y axis uses an inverse error function ("probability
746 plot") transformation.

747 Figure 2. a) Ensemble averaged across all 21 models, the mean difference between the
748 bias-corrected and the original GCM-predicted changes (2070-2100 minus 1976-2005) in daily
749 maximum temperature (°C). b) RMS spread of the differences between the bias-corrected and
750 original GCM-predicted temperature changes across the 21 GCMs. Values are shown for 2
751 months (rows) and 3 bias correction methods (columns).

752 Figure 3. As in Figure 2, but for precipitation, in units of percentage points.

753 Figure 4. Correction factors, *K*, for the PresRat scheme that are necessary to preserve
754 model-predicted changes (2070-2099 vs. 1976-2005) in mean precipitation, illustrated for four
755 months. Values are averaged across 21 GCMs. White areas are within 5% of unity.

Figure 5. Left column: proportion (%) of total variance of daily maximum temperature that falls in the frequency band whose period is indicated in the panel title, from observations over the period 1976-2005. Note that the color range varies substantially by frequency band. Middle column: the multi-model mean error (%) for the same quantity in the GCMs, relative to the observations. Right column: the multi-model RMSE (%).

Figure 6. As in Figure 5, but for daily precipitation.

Figure 7. For daily maximum temperature (top panels) and precipitation (lower panels): the left column shows normalized spectra from observations (red line), CCSM4 (blue line), and CCSM4 after frequency-dependent bias correction (green dots and line). Right column: Ratio of CCSM4 spectral power to observations before (blue line) and after (green dots and line) frequency-dependent bias correction. Values are shown at a hot, dry location in central Nevada (39.5, −116.5), and a cool, wet location between Seattle and Portland (46.5, −122.5), as indicated in the panel titles.

Figure 8. For daily maximum temperature (top row) and precipitation (bottom row), the multi-model ensemble average log-RMSE in simulating the observed distribution of variance across frequency, both before the frequency-dependent bias correction (left column) and after (middle column). Right: histograms of how the frequency-dependent bias correction changes the log-RMSE, taken over all models and all locations.

Figure 9. Histograms of how much the frequency-dependent bias correction alters the daily temperature (left, °C) and precipitation (right two panels). The precipitation results are given both as the fraction change (%) and absolute change (mm/day). Results are shown for all the models across all points in the conterminous U.S.

778 Figure 10. Analysis of runoff simulated by the VIC hydrological model with various
779 meteorological forcing fields. a) Error in simulated spectral power of runoff (at a period of 30
780 yrs) when VIC is forced with temperature and precipitation fields from the CCSM4 GCM, where
781 error is defined as log10(power using GCM forcing / power using observed forcing). b) Same as
782 panel a, but using bias corrected (BC) GCM forcing fields. c) same as panel a, but using BC and
783 frequency dependent bias corrected (FDBC) fields. d) Domain-averaged mean error as a function
784 of frequency; black line is for BC forcing fields, dashed red line is for BC+FDBC forcing fields.
785 Figure 11. RMS error (% of climatological annual mean value) in the annual cycle of
786 precipitation (smoothed with a 31-day boxcar filter) as simulated by the CCSM4 model, using
787 various bias correction approaches. a) Original model (no bias correction). b) Simply monthly
788 BC. c) Single central day corrected based on statistics of a sliding 31-day window. d) 91-day
789 window. e) Iterative BC with 91-, 181-, and 365-day windows, but no preconditioning. f)
790 Iterative BC with preconditioning.
791 Figure 12. Scatterplot of sorted daily precipitation values, observed versus model with the
792 following bias correction applied: a) Original model data (no bias correction). b) Simple monthly
793 BC. c) Single central day corrected based on statistics of a sliding 31-day window. d) 91-day
794 window. e) Iterative BC with 91-, 181-, and 365-day windows, but no preconditioning. f)
795 Iterative BC with preconditioning. The dashed red line shows the best fit least-squares line based
796 on the 5 largest values. Model data are from the CCSM4 GCM at a point in the Sierra Nevada
797 (37.5 °N, −119.5 °W), over the period 1976-2005.
798 Figure 13. The slope (left column) and intercept (right column) of the best fit least-
799 squares line between the top 5 observed and modeled (CCSM4) extreme events for different bias
800 correction approaches as indicated in the panel titles.

801

Abbreviation	Model source/institution
Access1-0	Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM), Australia
Bcc-csm1-1	Beijing Climate Center, China
Bnu-esm	Beijing Normal University, China
CanESM2	Canadian Centre for Climate Modelling and Analysis, Canada
CCSM4	National Center for Atmospheric Research, USA
ECSM1-BGC	National Center for Atmospheric Research, USA
CNRM-CM5	Centre National de Recherches Meteorologiques, France
CSIRO-Mk3.6.0	QCCCE & Commonwealth Scientific and Industrial Research Organization, Australia
GFDL-CM3	Geophysical Fluid Dynamics Laboratory, Princeton, USA
GFDL-ESM2G	Geophysical Fluid Dynamics Laboratory, Princeton, USA
GFDL-ESM2M	Geophysical Fluid Dynamics Laboratory, Princeton, USA
INMCM4	Institute of Numerical Mathematics Russian Academy of Sciences, Russia
IPSL-CM5a-LR	Institut Pierre-Simon Laplace, France
IPSL-CM5a-MR	Institut Pierre-Simon Laplace, France
MIROC-ESM	Japan Agency for Marine-Earth Science and Technology, and National Inst. For Environ. Studies, Japan
MIROC-ESM-CHEM	Japan Agency for Marine-Earth Science and Technology, and National Inst. For Environ. Studies, Japan
MIROC5	Atmosphere and Ocean Research Institute and Nat. Inst. For Environ. Studies, Japan
MPI-ESM-LR	Max Planck Institute for Meteorology, Germany
MPI-ESM-MR	Max Planck Institute for Meteorology, Germany
MRI-CGCM3	Meteorological Rsearch Institute, Japan
NorESM1-m	Norwegian Climate Centre

802 Table 1. The GCMs used in this work and their originating institutions.

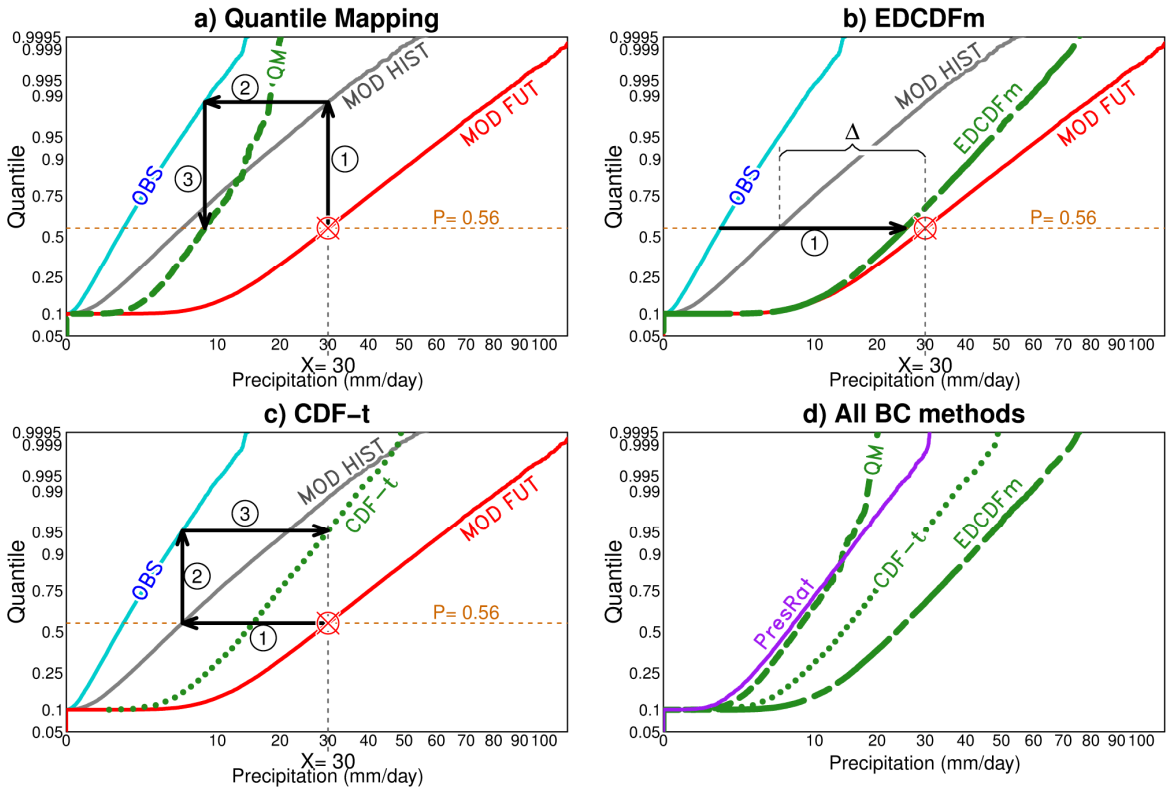


Figure 1. Cumulative distribution functions (CDFs) of synthetic daily precipitation data schematically illustrating how each bias correction method constructs the model's bias corrected future CDF (green dotted/dashed lines). The solid blue, grey, and red lines are the same in all panels and show the observed (1976-2005), model historical (1976-2005), and model future (2070-2099) CDFs, respectively. The example point being corrected is X=30 mm/day, which falls at the 0.56 quantile in the model future distribution (dotted orange line). a) Quantile mapping (QM): starting at the point to be corrected, go vertically to the grey line (1), horizontally to the blue line (2), and vertically to the original percentile (3). b) Equidistant CDF-matching (EDCDFm): at the quantile of the point being corrected, compute the offset from the model historical value to the model future value (Δ), then add Δ to the observed value at the percentile being corrected (1). c) The CDF-transform (CDF-t) method; starting at the point to be corrected, go horizontally to the grey line (1), vertically to the blue line (2), and horizontally to the original value (3). d) Final results from all 3 bias correction methods (dotted/dashed green lines), along with the PresRat method (solid purple line) for comparison. Note that the X axis uses a square root transformation and the Y axis uses an inverse error function ("probability plot") transformation.

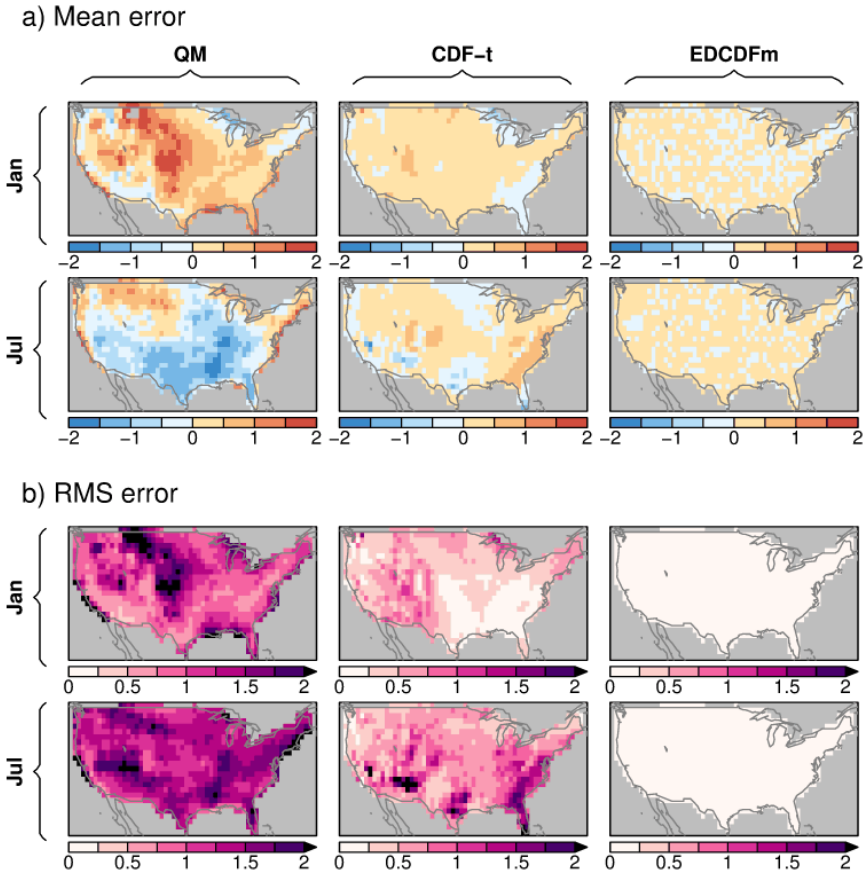


Figure 2. a) Ensemble averaged across all 21 models, the mean difference between the bias-corrected and the original GCM-predicted changes (2070-2100 minus 1976-2005) in daily maximum temperature (°C). b) RMS spread of the differences between the bias-corrected and original GCM-predicted temperature changes across the 21 GCMs. Values are shown for 2 months (rows) and 3 bias correction methods (columns).

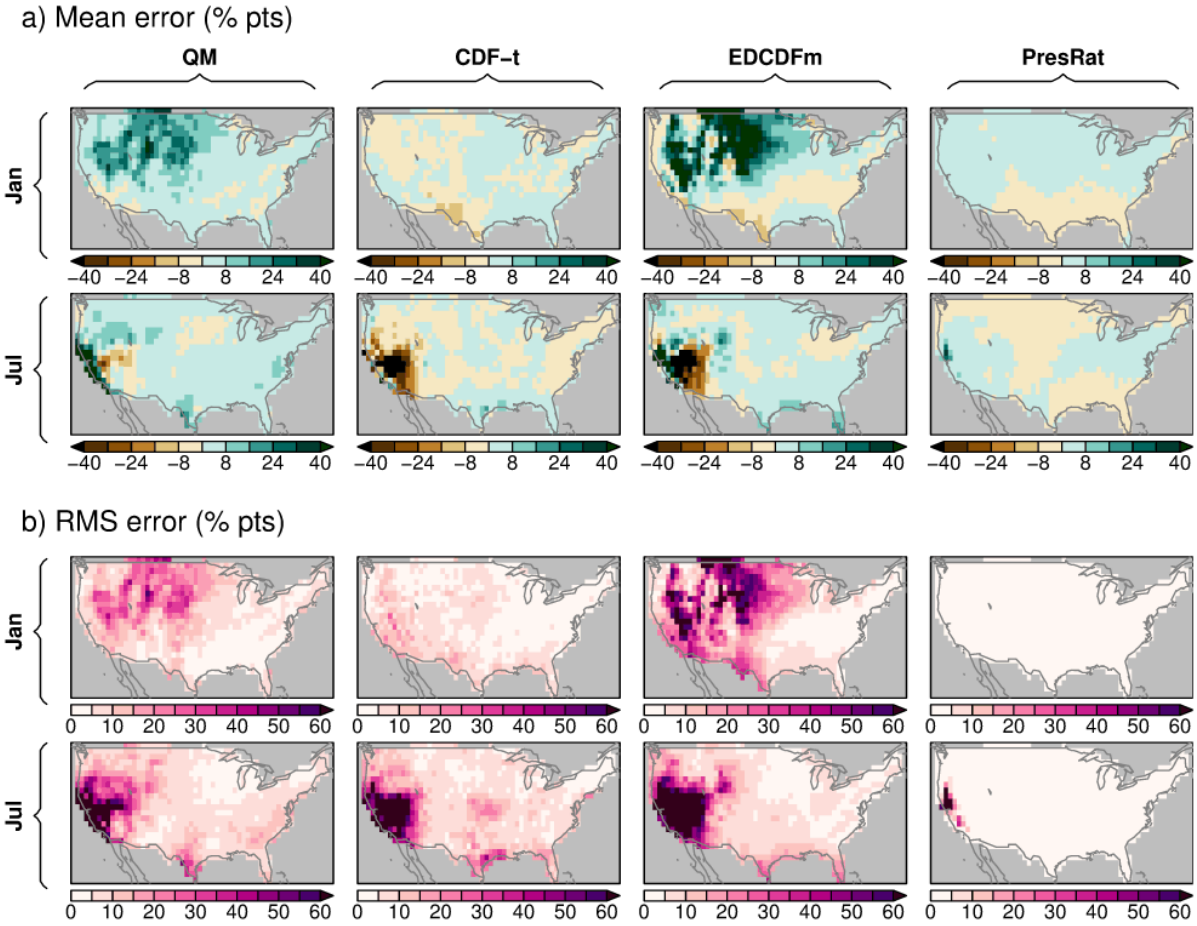


Figure 3. As in Figure 2, but for precipitation, in units of percentage points.

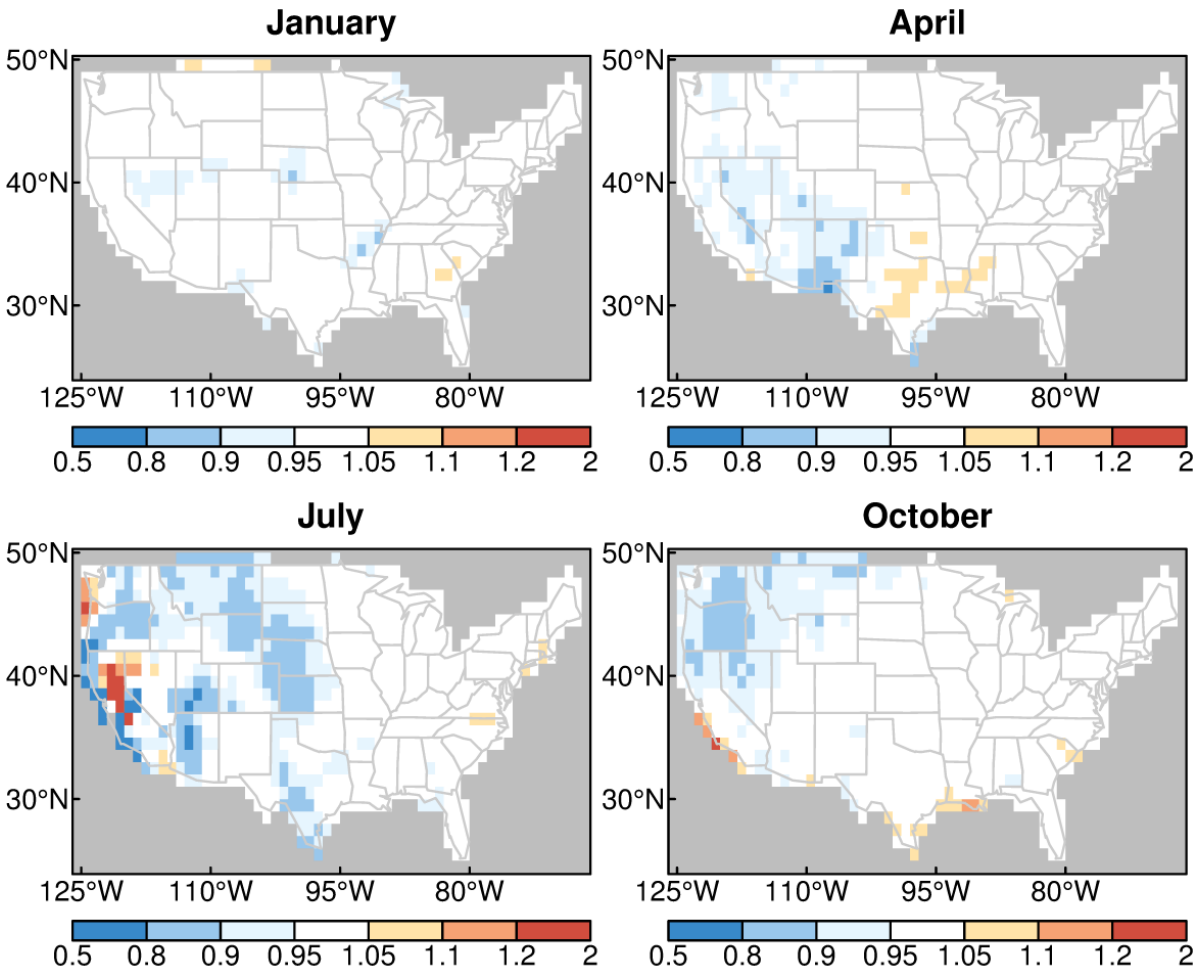
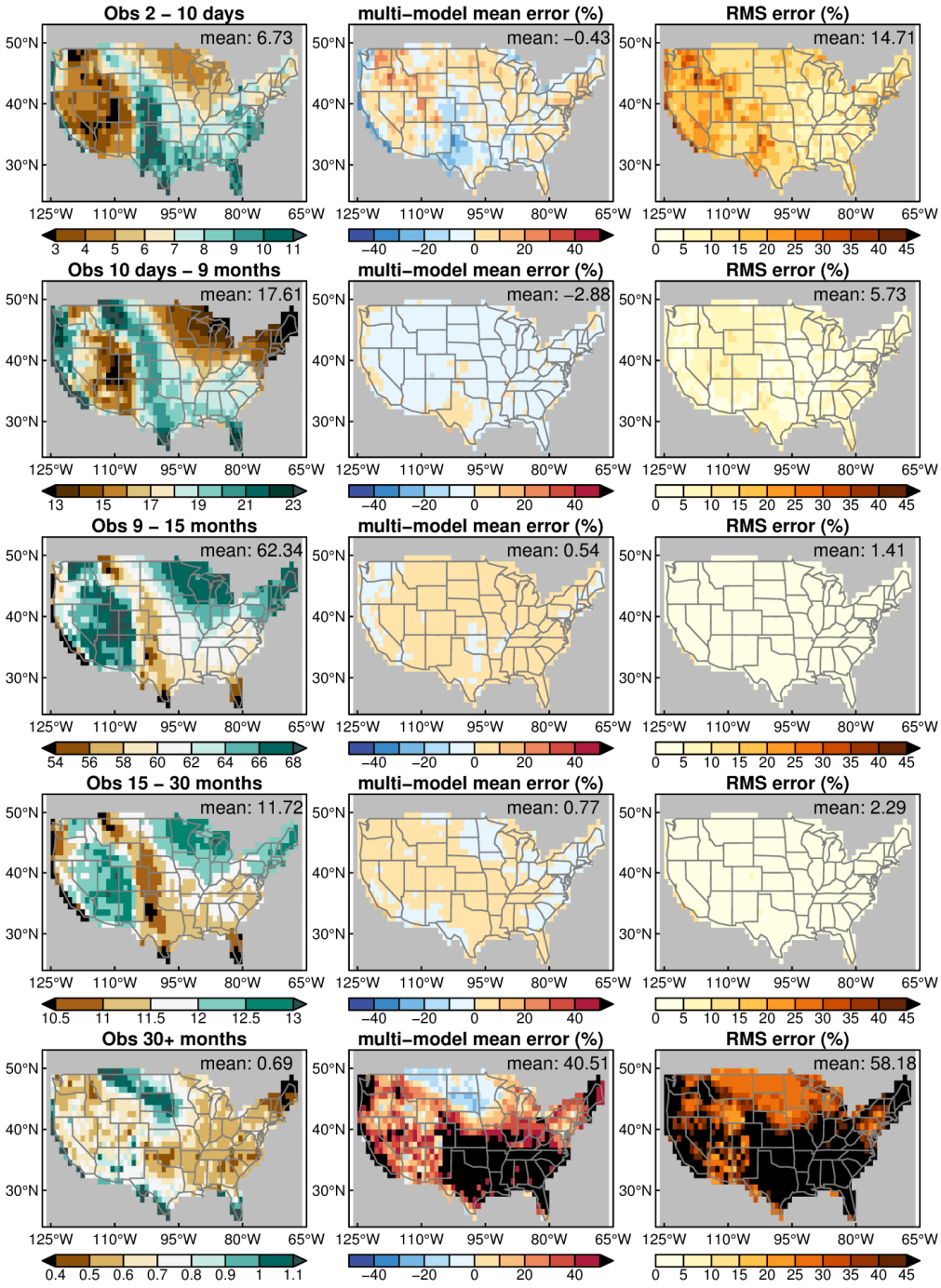


Figure 4. Correction factors, *K*, for the PresRat scheme that are necessary to preserve model-predicted changes (2070-2099 vs. 1976-2005) in mean precipitation, illustrated for four months. Values are averaged across 21 GCMs. White areas are within 5% of unity.



837 .

838 Figure 5. Left column: proportion (%) of total variance of daily maximum temperature
839 that falls in the frequency band whose period is indicated in the panel title, from observations
840 over the period 1976-2005. Note that the color range varies substantially by frequency band.
841 Middle column: the multi-model mean error (%) for the same quantity in the GCMs, relative to
842 the observations. Right column: the multi-model RMSE (%).

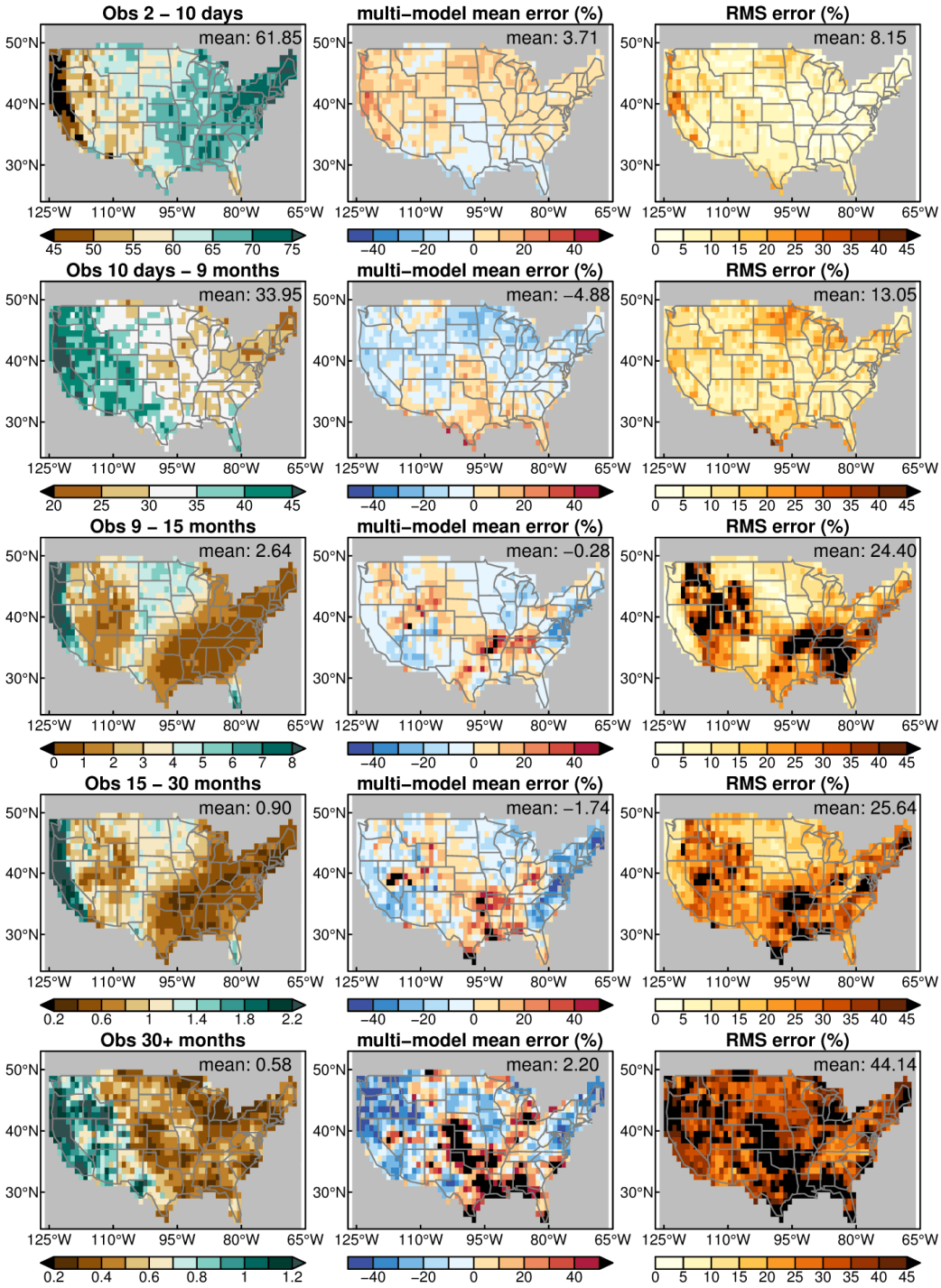


Figure 6. As in Figure 5, but for daily precipitation.

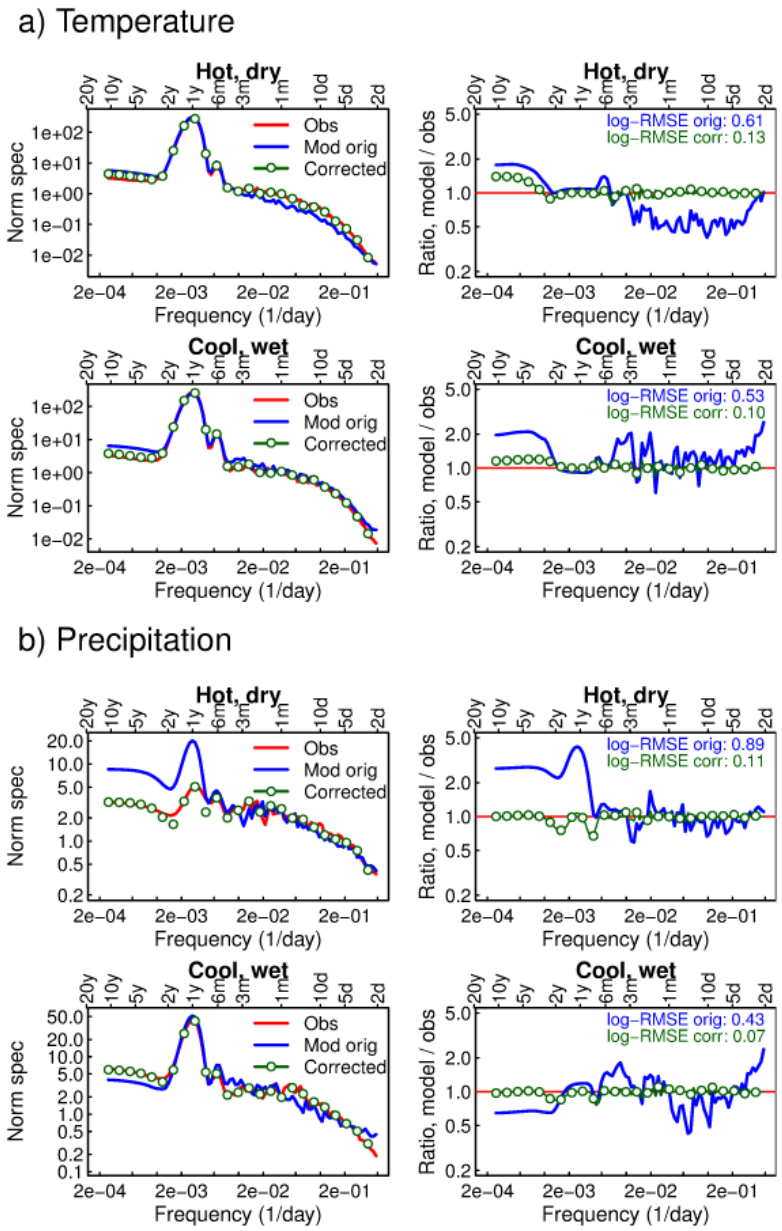


Figure 7. For daily maximum temperature (top panels) and precipitation (lower panels): the left column shows normalized spectra from observations (red line), CCSM4 (blue line), and CCSM4 after frequency-dependent bias correction (green dots and line). Right column: Ratio of CCSM4 spectral power to observations before (blue line) and after (green dots and line) frequency-dependent bias correction. Values are shown at a hot, dry location in central Nevada (39.5, −116.5), and a cool, wet location between Seattle and Portland (46.5, −122.5), as indicated in the panel titles.

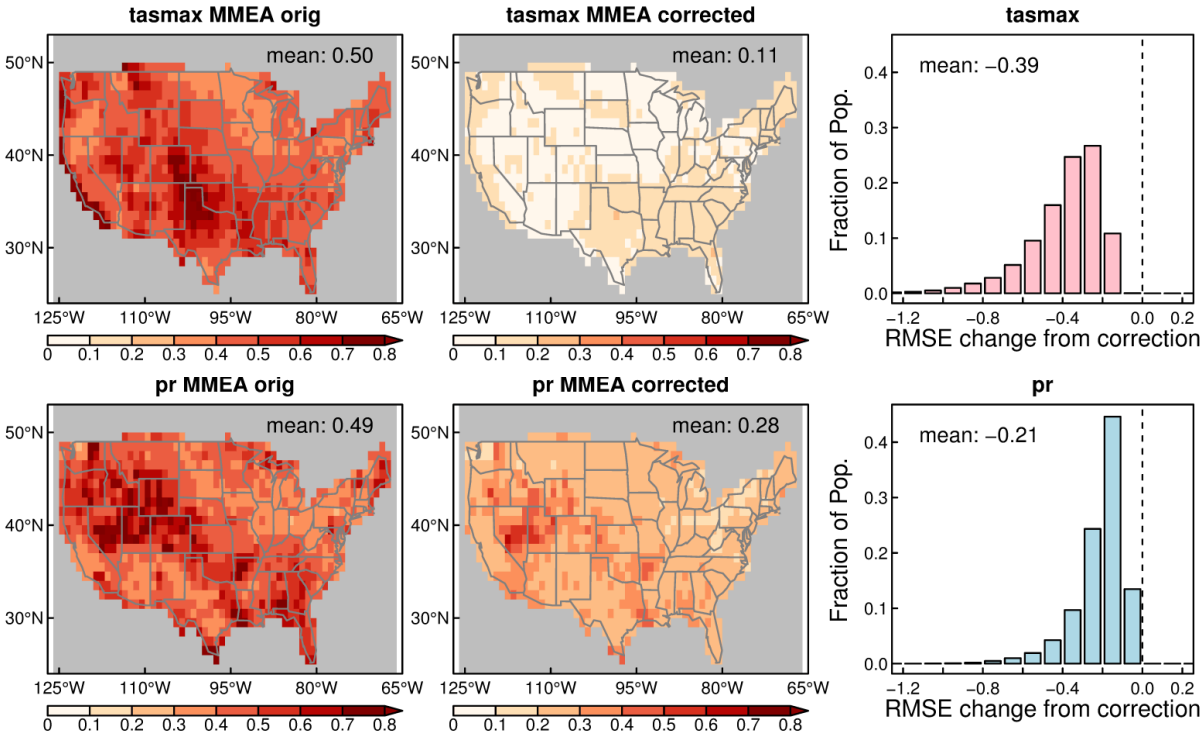
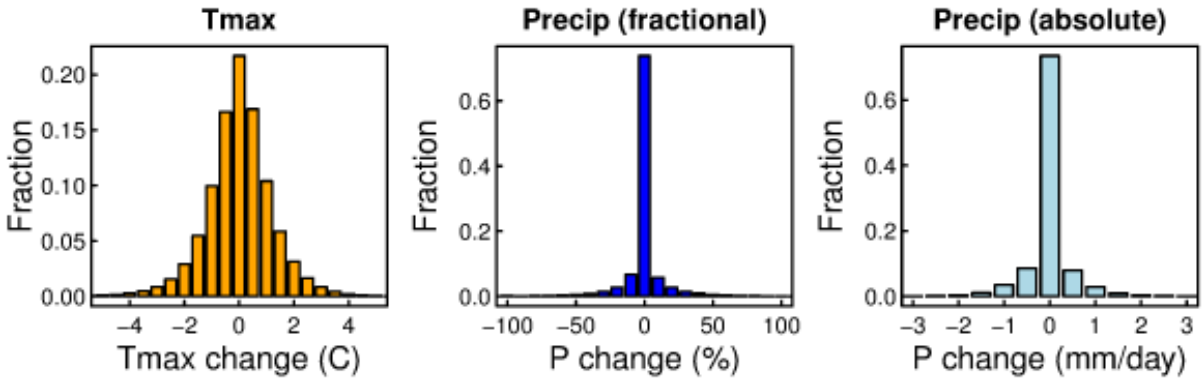


Figure 8. For daily maximum temperature (top row) and precipitation (bottom row), the multi-model ensemble average log-RMSE in simulating the observed distribution of variance across frequency, both before the frequency-dependent bias correction (left column) and after (middle column). Right: histograms of how the frequency-dependent bias correction changes the log-RMSE, taken over all models and all locations.

/cir1/cmip5_regrid/get_rmse_all_models.R



862

863 Figure 9. Histograms of how much the frequency-dependent bias correction alters the
864 daily temperature (left, °C) and precipitation (right two panels). The precipitation results are
865 given both as the fraction change (%) and absolute change (mm/day). Results are shown for all
866 the models across all points in the conterminous U.S.

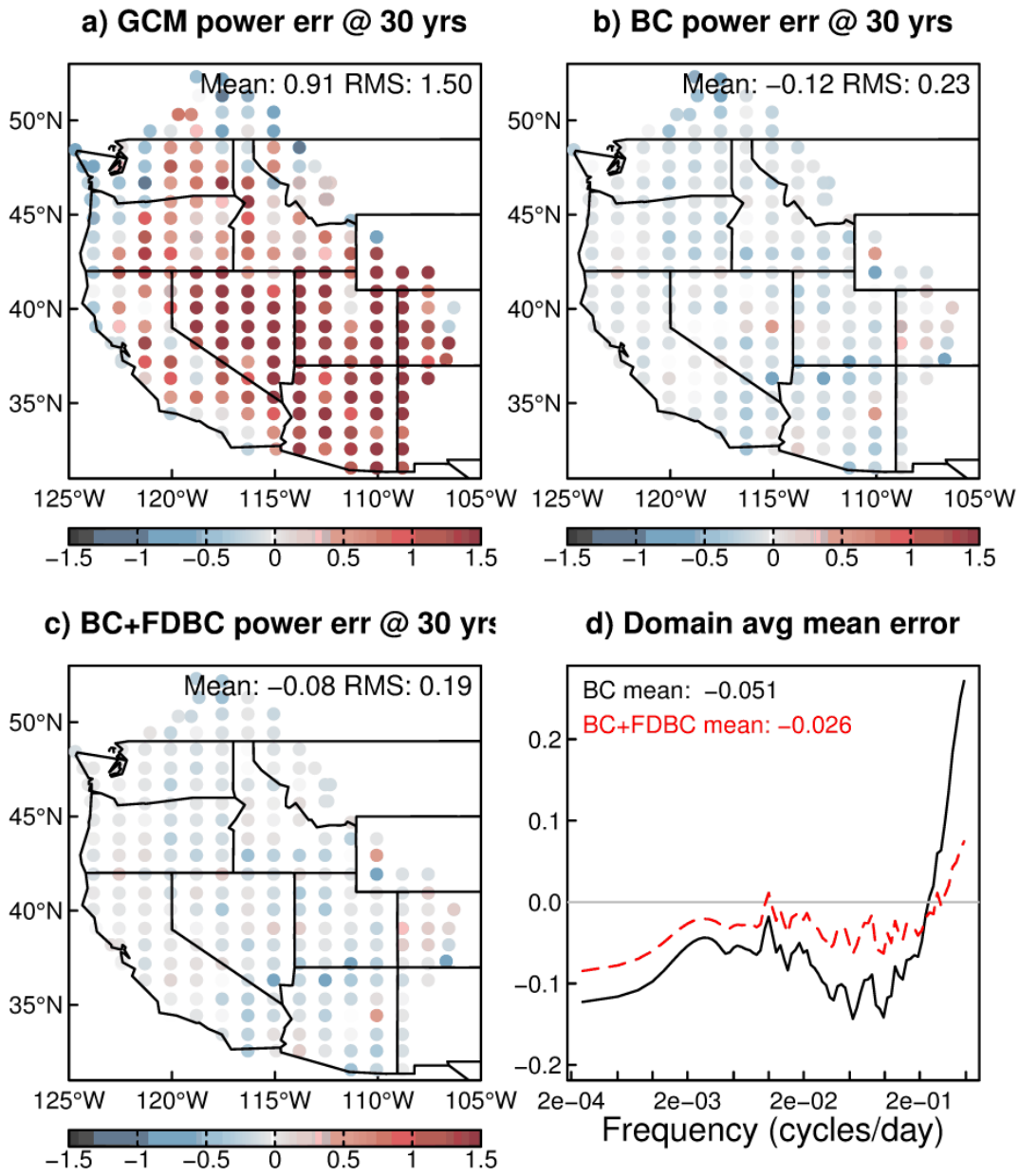


Figure 10. Analysis of runoff simulated by the VIC hydrological model with various meteorological forcing fields. a) Error in simulated spectral power of runoff (at a period of 30 yrs) when VIC is forced with temperature and precipitation fields from the CCSM4 GCM, where error is defined as log10(power using GCM forcing / power using observed forcing). b) Same as panel a, but using bias corrected (BC) GCM forcing fields. c) same as panel a, but using BC and frequency dependent bias corrected (FDBC) fields. d) Domain-averaged mean error as a function of frequency; black line is for BC forcing fields, dashed red line is for BC+FDBC forcing fields.

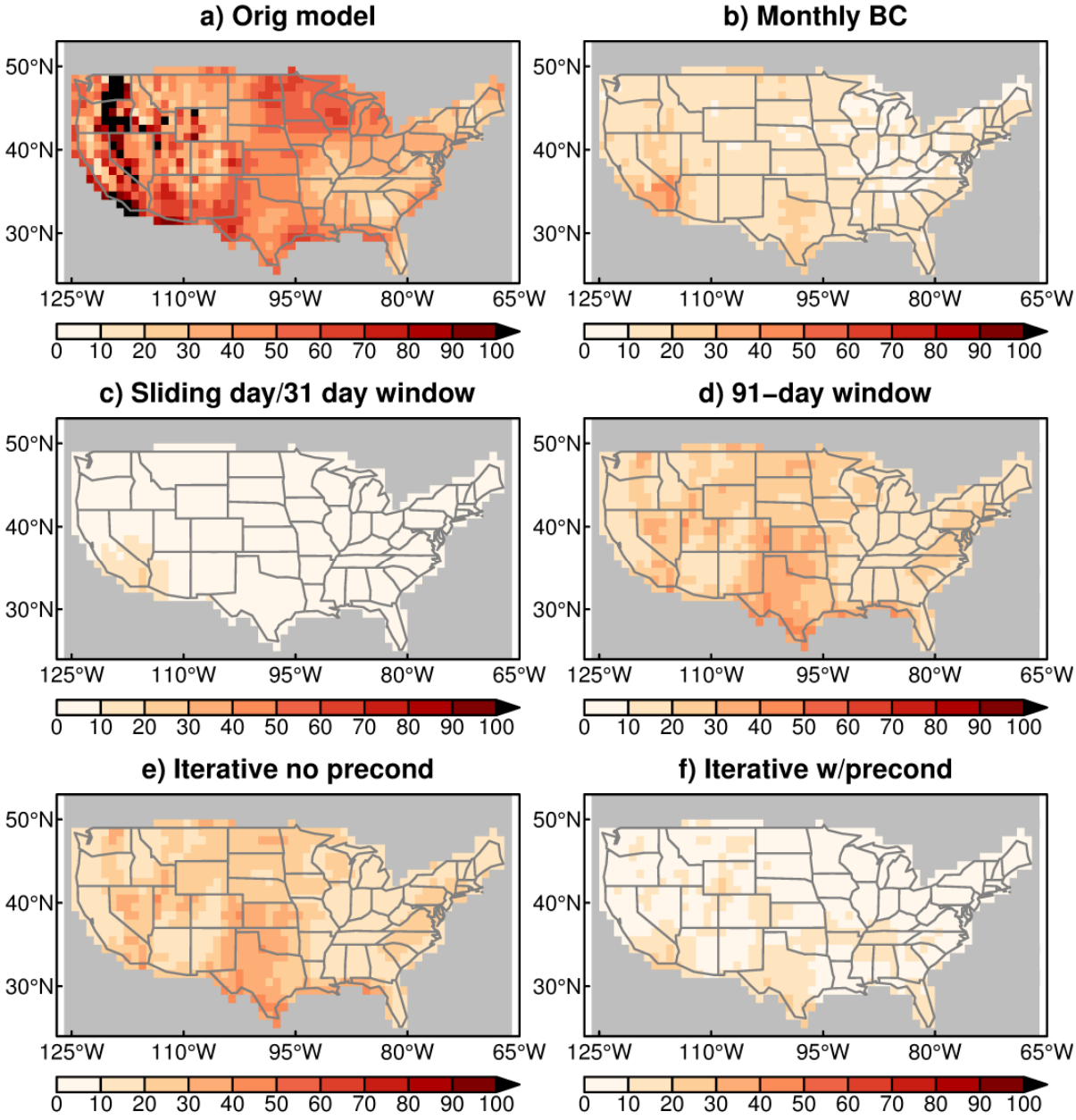
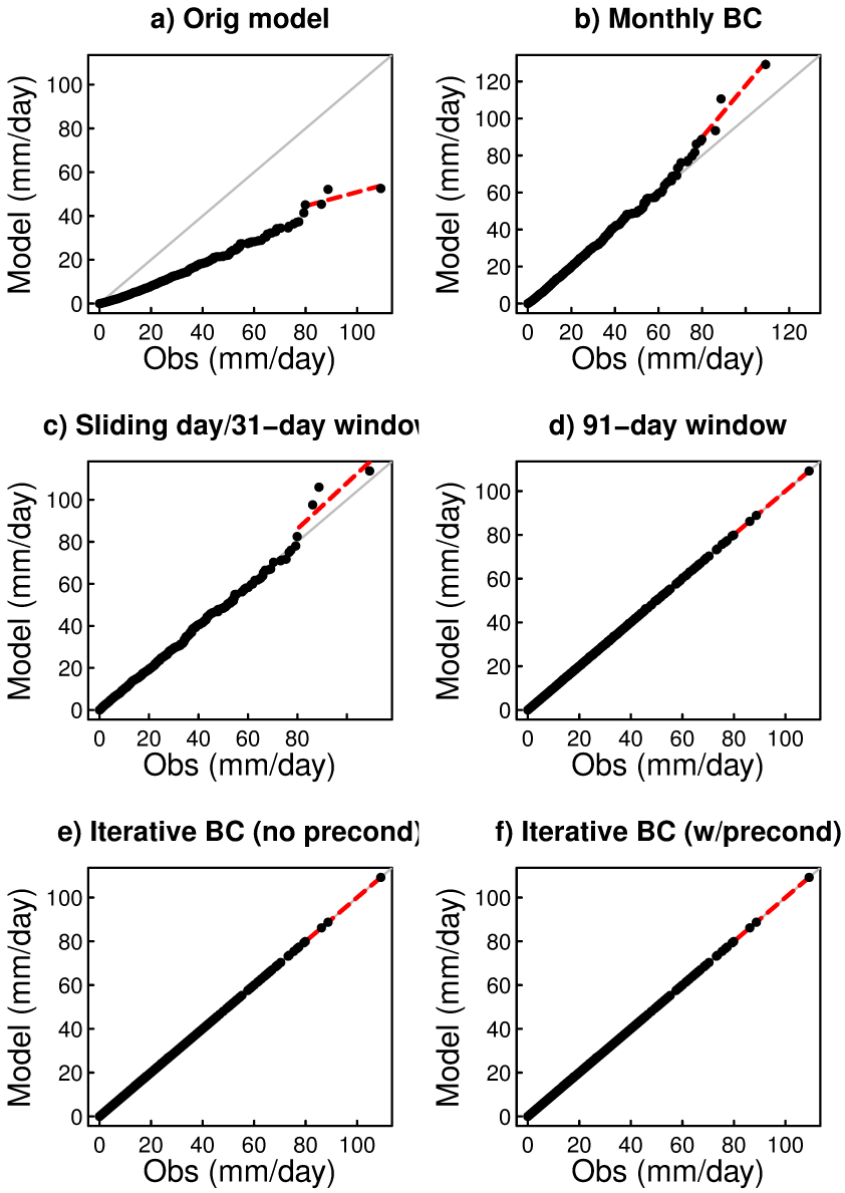


Figure 11. RMS error (% of climatological annual mean value) in the annual cycle of precipitation (smoothed with a 31-day boxcar filter) as simulated by the CCSM4 model, using various bias correction approaches. a) Original model (no bias correction). b) Simply monthly BC. c) Single central day corrected based on statistics of a sliding 31-day window. d) 91-day window. e) Iterative BC with 91-, 181-, and 365-day windows, but no preconditioning. f) Iterative BC with preconditioning.



882

883 Figure 12. Scatterplot of sorted daily precipitation values, observed versus model with the
884 following bias correction applied: a) Original model data (no bias correction). b) Simple monthly
885 BC. c) Single central day corrected based on statistics of a sliding 31-day window. d) 91-day
886 window. e) Iterative BC with 91-, 181-, and 365-day windows, but no preconditioning. f)
887 Iterative BC with preconditioning. The dashed red line shows the best fit least-squares line based
888 on the 5 largest values. Model data are from the CCSM4 GCM at a point in the Sierra Nevada
889 (37.5 °N, −119.5 °W), over the period 1976-2005.

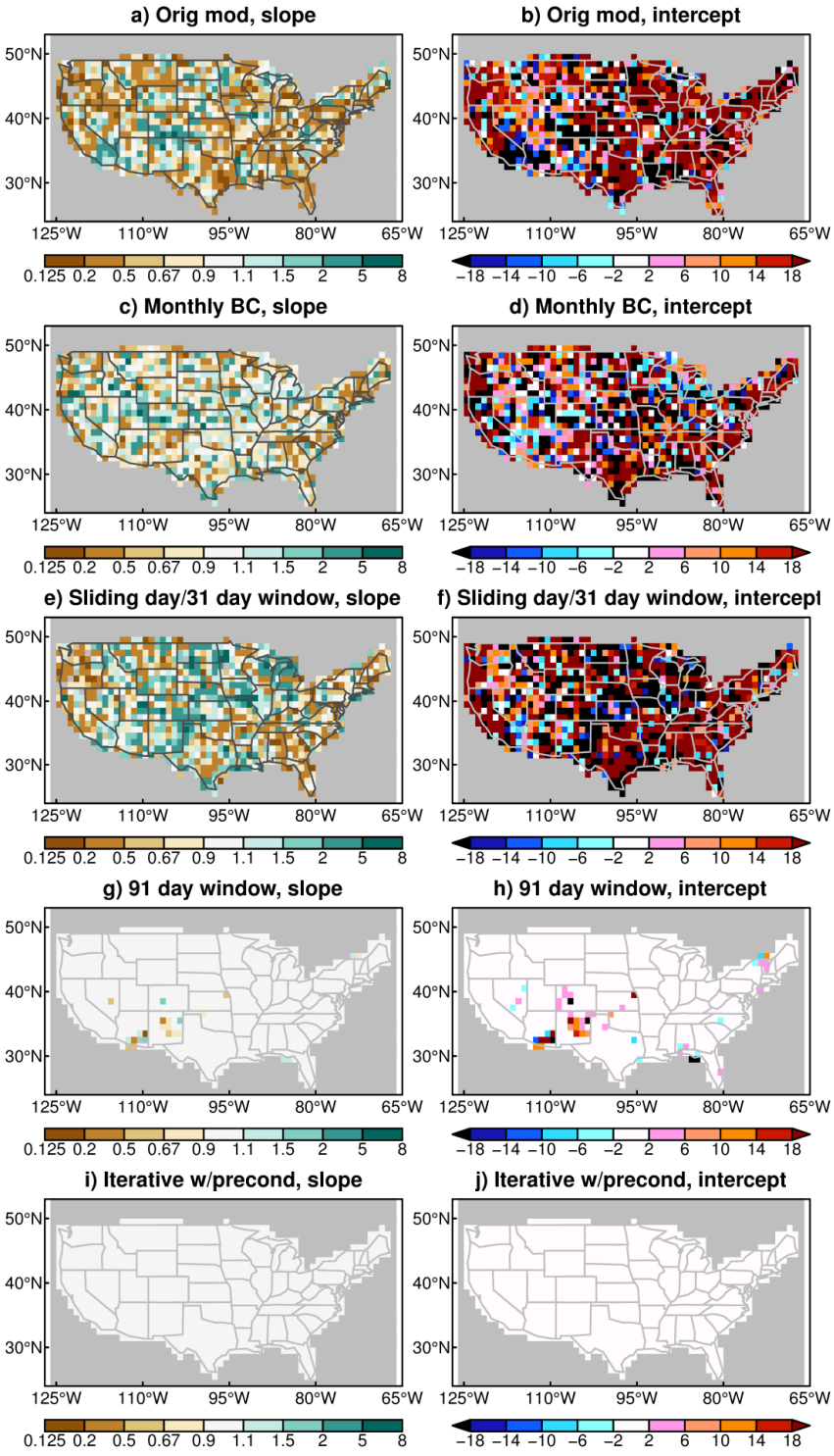


Figure 13. The slope (left column) and intercept (right column) of the best fit least-squares line between the top 5 observed and modeled (CCSM4) extreme events for different bias correction approaches as indicated in the panel titles.