Civil Engineering   School of Engineering

12-2015

Improved Bias Correction Techniques for Hydrological Simulations of Climate Change

David W. Pierce

Daniel R. Cayan

Edwin P. Maurer
Santa Clara University, emaurer@scu.edu

John T. Abatzoglou

Katherine C. Hegewisch

Improved Bias Correction Techniques for Hydrological Simulations of Climate Change*

DAVID W. PIERCE

Division of Climate, Atmospheric Sciences, and Physical Oceanography, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California

DANIEL R. CAYAN

Division of Climate, Atmospheric Sciences, and Physical Oceanography, Scripps Institution of Oceanography, University of California, San Diego, and U.S. Geological Survey, La Jolla, California

EDWIN P. MAURER

Civil Engineering Department, Santa Clara University, Santa Clara, California

JOHN T. ABATZOGLOU AND KATHERINE C. HEGEWISCH

Department of Geography, University of Idaho, Moscow, Idaho

(Manuscript received 16 December 2014, in final form 6 June 2015)

ABSTRACT

Global climate model (GCM) 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 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% 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 it does so without making any locations worse, unlike standard bias correction. Last, 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

Corresponding author address: David W. Pierce, Division of Climate, Atmospheric Sciences, and Physical Oceanography, Scripps Institution of Oceanography, University of California, San Diego, Mail Stop 0224, La Jolla, CA 92093-0224.
E-mail: dpierce@ucsd.edu

example, precipitation biases degrade hydrological simulations because of the nonlinear 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 nonlinear 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 United States (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 2014). 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 time scales, including the trend (Pierce et al. 2013; Maurer and Pierce 2014). 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 time scales 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-time-scale 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 time scales yet too little on annual time scales, making it challenging to represent the proper magnitude and spectra of phenomena such as droughts. Although simulations have improved with the models in phase 5 of the Coupled Model Intercomparison Project (CMIP5), deficiencies still remain in representing regional variability on interannual to decadal time scales (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.

Last, 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

TABLE 1. The GCMs used in this work and their originating institutions (see http://www.ametsoc.org/PubsAcronymList for full model expansions).

Model acronym	Model source/institution
ACCESS1.0	Commonwealth Scientific and Industrial Research Organisation (CSIRO) and Bureau of Meteorology, 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 (NCAR), United States
CESM1[biogeochemistry (BGC)]	NCAR, United States
CNRM-CM5	Centre National de Recherches Météorologiques, France
CSIRO Mk3.6.0	Queensland Climate Change Centre of Excellence and CSIRO, Australia
GFDL CM3	Geophysical Fluid Dynamics Laboratory (GFDL), Princeton, New Jersey, United States
GFDL-ESM2G	GFDL, Princeton, New Jersey, United States
GFDL-ESM2M	GFDL, Princeton, New Jersey, United States
INM-CM4	Institute of Numerical Mathematics, Russian Academy of Sciences, Russia
IPSL-CM5A-LR	L'Institut Pierre-Simon Laplace, France
IPSL-CM5A-MR	L'Institut Pierre-Simon Laplace, France
MIROC-ESM	Japan Agency for Marine-Earth Science and Technology (JAMSTEC) and National Institute for Environmental Studies (NIES), Japan
MIROC-ESM-CHEM	JAMSTEC and NIES, Japan
MIROC5	Atmosphere and Ocean Research Institute and NIES, Japan
MPI-ESM-LR	Max Planck Institute for Meteorology, Germany
MPI-ESM-MR	Max Planck Institute for Meteorology, Germany
MRI-CGCM3	Meteorological Research Institute, Japan
NorESM1-M	Norwegian Climate Centre, Norway

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

a. Global climate models

We use daily maximum temperature and precipitation fields from 21 GCMs that participated in 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° × 1° longitude–latitude) global climate models in CMIP5 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 (2006–99) runs using representative concentration pathway 8.5 (RCP8.5).

b. 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 Cooperative 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 $\frac{1}{8}° \times \frac{1}{8}°$ latitude–longitude grid, which we aggregated to the same 1° × 1° grid as the GCM output.

c. Time periods

The World Meteorological Organization (WMO) recommends that climatological normals be calculated over 30-yr 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; and 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 2010–39, 2040–69, and 2070–99 separately. In

the results shown below, we focus on 2070–99 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 (2014), 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 (FDBC) or preconditioning.

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

1) QUANTILE MAPPING

Quantile mapping (Panofsky and Brier 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 Fig. 1a, using cumulative distribution functions (CDFs) of synthetic gamma distributions to mimic precipitation.

Averaged across the 21 GCMs, QM exaggerates monthly mean model-projected warming (2070–99 minus 1976–2005) in the Rockies in January and diminishes it in July (Fig. 2a). Maurer and Pierce (2014) 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 multimodel mean values, but the modification in any individual model can be much greater. The RMS spread across the 21 models is shown in Fig. 2b. The spread is appreciable using the QM technique, with RMS values of up to 2°C, and more spread is found in the warmer months.

Figure 3 shows a similar analysis for precipitation, evaluated multiplicatively in terms of percentage change. QM tends to make the original model-predicted mean change wetter over the northwestern United States in January and California in July. The RMS spread across models is ~25% points in parts of the Northwest in January and exceeds 60% points in the dry California–Great Basin region in July.

2) CDF-t

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

The CDF-t results in Figs. 2 and 3 show that CDF-t modifies the original monthly mean temperature projection less than QM, but still on the order of 0.5°C. CDF-t tends to make the precipitation projections drier, which can be understood in terms of Fig. 1c. To produce a point on the bias-corrected future distribution (green dotted line), it is necessary that the model historical value at the quantile being bias corrected falls within the range of observed values, as indicated by arrow "2" in Fig. 1c. As arrow "2" progressively moves to the right in Fig. 1c, at higher quantiles it becomes impossible to map future changes beyond the maximum 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) 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 Fig. 1b (note the nonlinear x axis when considering the length of the arrow "Δ"). When bias correcting a model historical run, EDCDFm reduces to QM.
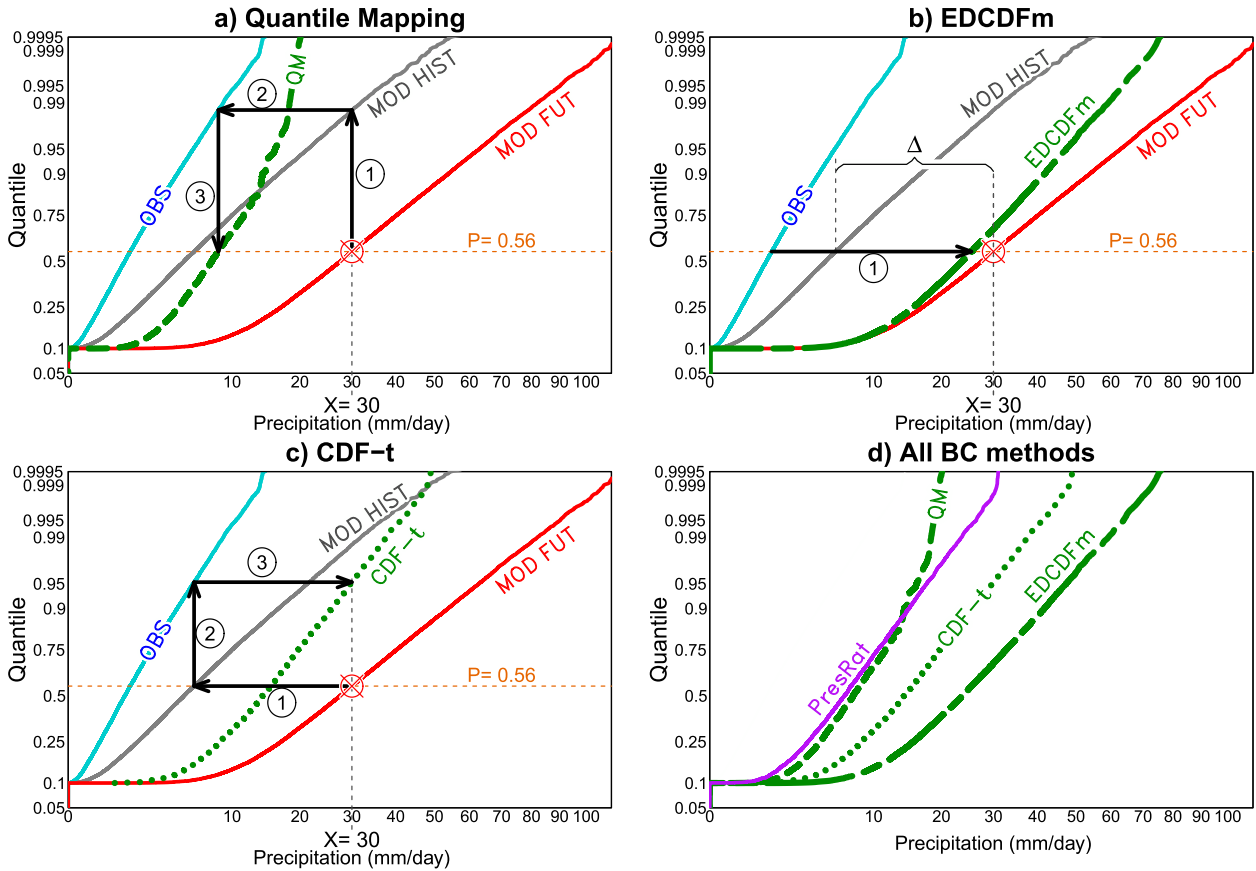
FIG. 1. CDFs of synthetic daily precipitation data schematically illustrating how each bias correction method constructs the model's bias-corrected future CDF (green dotted and dashed lines). The solid blue, gray, and red lines show the observed (1976–2005), model historical (1976–2005), and model future (2070–99) CDFs, respectively. The example point being corrected is $X = 30$ mm day^{-1}, which falls at the 0.56 quantile in the model future distribution (dotted orange line). (a) QM, starting at the point to be corrected, goes vertically to the gray line (indicated by a "1"), horizontally to the blue line (indicated by a "2"), and vertically to the original percentile (indicated by a "3"). (b) EDCDFm at the quantile of the point being corrected computes the offset from the model historical value to the model future value (indicated by a "Δ"), then adds Δ to the observed value at the percentile being corrected (indicated by a "1"). (c) The CDF-t method, starting at the point to be corrected, goes horizontally to the gray line (indicated by a "1"), vertically to the blue line (indicated by a "2"), and horizontally to the original value (indicated by a "3"). (d) Final results from all three bias correction methods (green dotted and dashed 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.

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 Celsius (Fig. 2, right).

As expected, EDCDFm does not preserve GCM-predicted fractional changes, that is, (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 Fig. 3. EDCDFm alters the original model-predicted mean precipitation change by more than 30% points in the dry (rain shadow) parts of the northwestern United States. This will happen particularly when there are both large biases and large changes in the upper quantiles of a skewed precipitation distribution.

b. 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

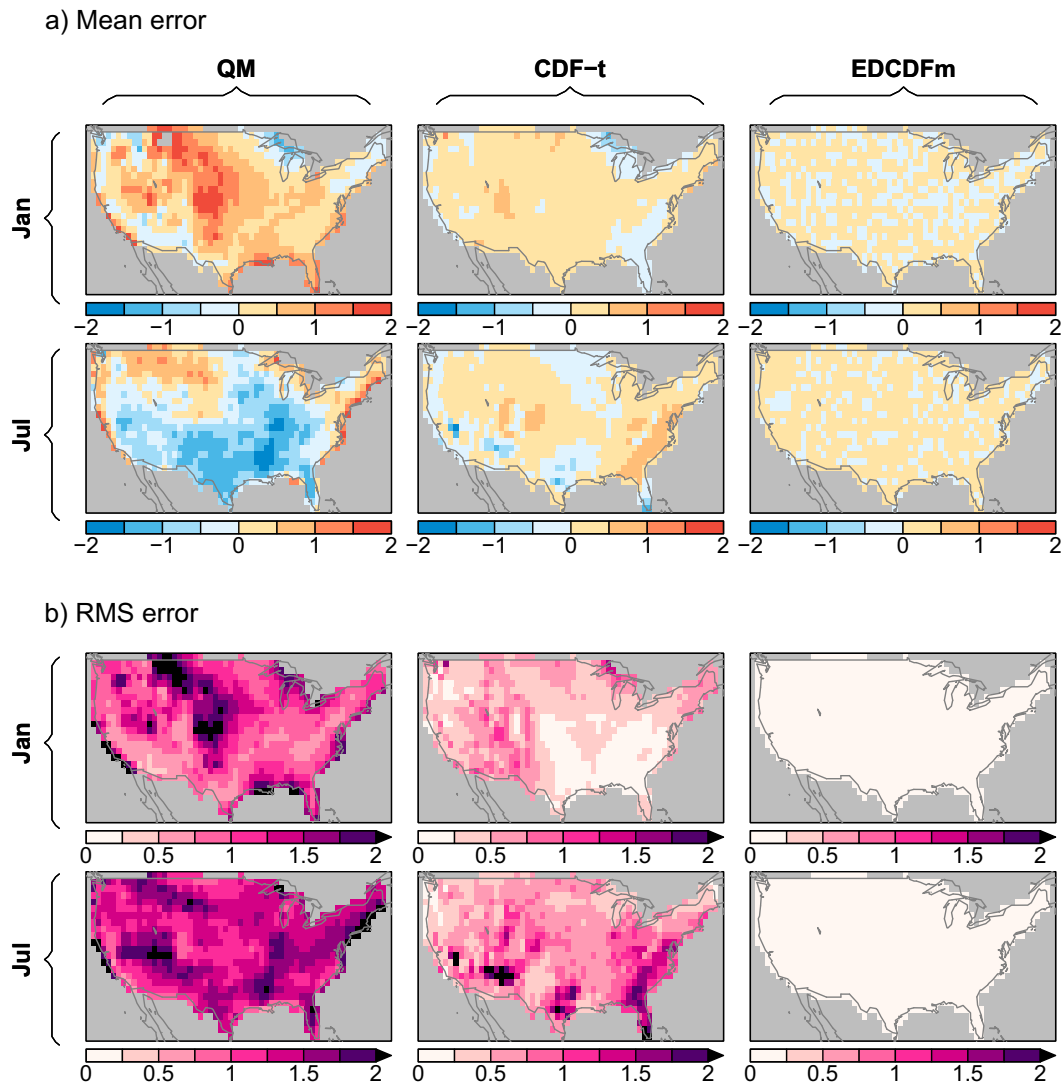a) Mean error

b) RMS error

FIG. 2. (a) Ensemble averaged across all 21 models, the mean difference between the bias-corrected and the original GCM-predicted changes (2070–99 minus 1976–2005) in daily max 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 two months (rows) and three bias correction methods (columns).

"2" in Fig. 1a), EDCDFm assumes that the historical model error in value at a given quantile is preserved in the future ("Δ" in Fig. 1b), and CDF-t assumes that the historical model error in quantile at a given quantile is preserved in the future (arrow "2" in Fig. 1c). (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
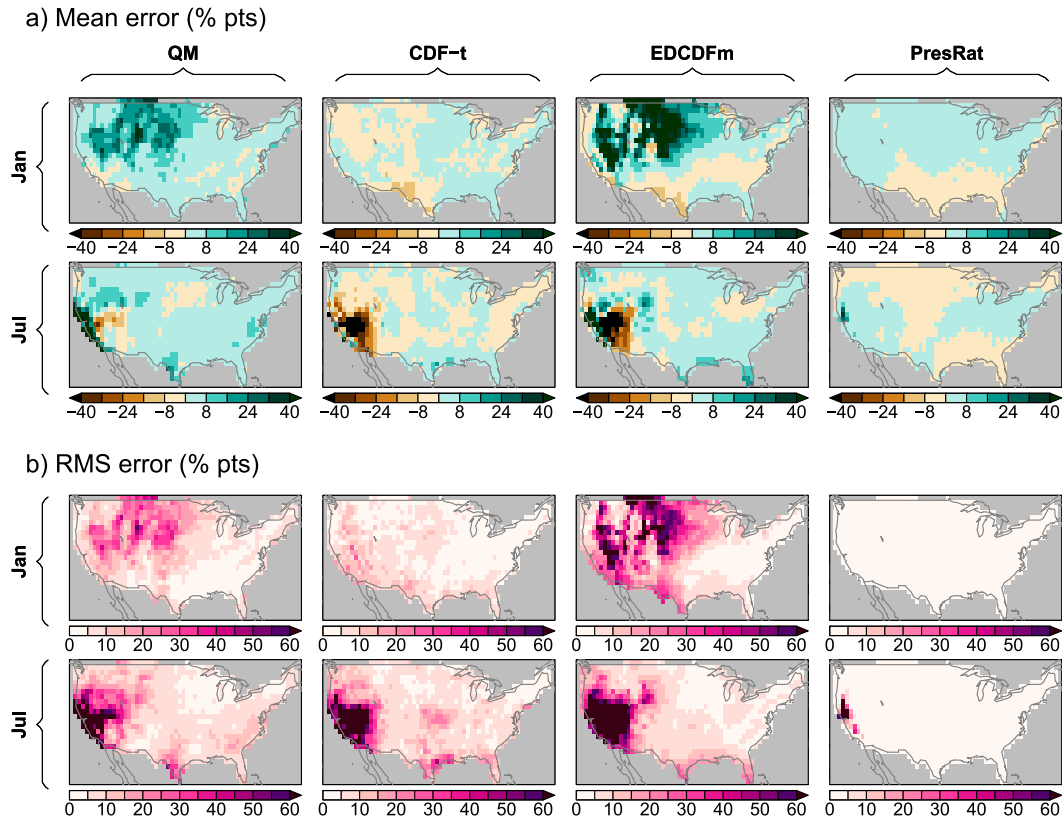
a) Mean error (% pts)

b) RMS error (% pts)

FIG. 3. As in Fig. 2, but for precipitation with four bias correction methods and units of percentage points.

(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 preserve 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 angle 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 $\tau \geq 0.01 \, \text{mm} \, \text{day}^{-1}$ 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 (not bias corrected) future time series. The model data are then bias corrected, and the smallest Z_{gf} of precipitation values are set to zero. This preserves the model-predicted change in fraction of nonprecipitating 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
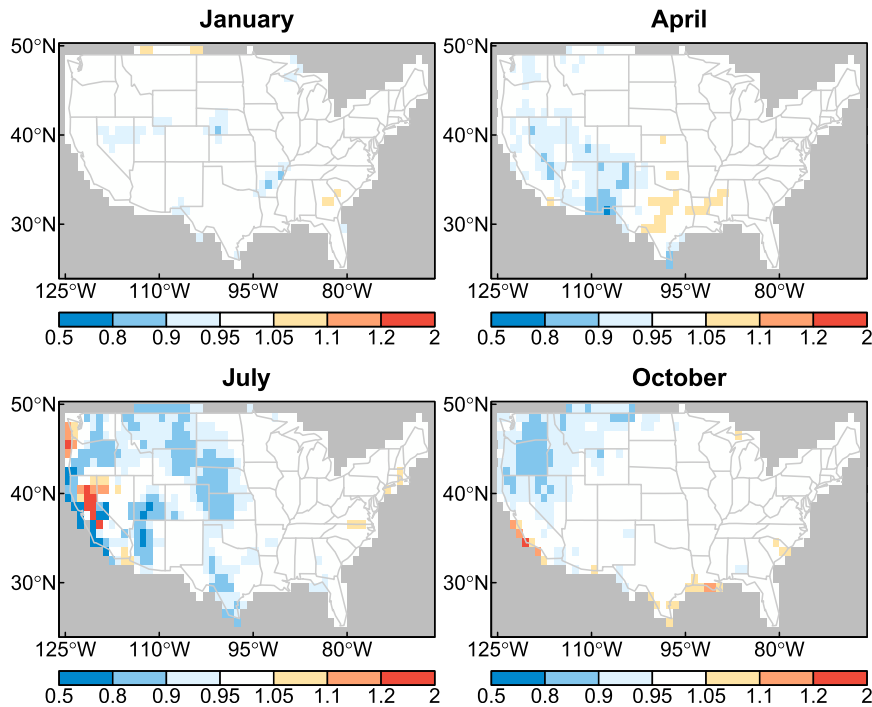
FIG. 4. Correction factors K for the PresRat scheme that are necessary to preserve model-predicted changes (2070–99 vs 1976–2005) in mean precipitation, illustrated for four months. Values are averaged across 21 GCMs. White areas are within 5% of unity.

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 (Fig. 3, right). Discrepancies only arise because of 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 IPCC (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 afterward].

4. Frequency-dependent bias correction

a. 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) time scales. We now address the more general question of model biases at different time scales and how to reduce them.

Details of our spectral approach are given in the appendix. 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 time scales of interest by this technique (e.g.,

bias correcting 2–10-day time-scale temperature biases in winter and summer separately).

Since we bias correct in 30-yr periods (section 2c), 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 2 days (the Nyquist frequency given daily model output) to 30 years. This interval is further refined to periods from 2 days to 11 years in light of our spectral analysis technique (see the appendix).

b. Frequency-dependent model errors

Figure 5 shows the observed (1976–2005) distribution of variance in daily maximum temperature across frequencies (labeled using equivalent periods; Fig. 5, left) and the multimodel mean errors in representing this distribution (Fig. 5, middle). Figure 5 (right) shows multimodel RMSE (i.e., at each point, the spread of values across the 21 models). The FDBC is based on normalized spectra (spectral values divided by the variance of the original time series) so that it leaves the overall variance unaltered. Therefore, at every location the values in Fig. 5 (left) summed across frequency bands totals 100%.

The annual cycle (9–15-month bands) dominates daily maximum temperature variability over almost all of the conterminous United States (CONUS), containing on average 62% of the variance. The main exceptions are along the California coast, Florida, and in a strip of the central United States downwind of the Rockies, where higher frequencies (<9 months) contribute more than elsewhere.

Models allocate less of the total variance to periods shorter than 9 months than observed. In the 10–30-day band, the mean error reaches −9% (not shown). The proportion of variance in the annual cycle is represented with little mean error and spread across models. Conversely, models allocate more of the total variance to periods longer than 30 months, with nearly ~40% more variance than observed, and the spread across models is large. However, the fraction of total variance in these long time scales is small (<1%).

Figure 6 shows the same analysis using daily precipitation. Periods between 2 and 10 days contain the majority of the variance (~62%). The exception is the west coast, where 10-day to 9-month variability is nearly as important, and the annual cycle contains >7% of the total variance. The models have a 5%–10% mean bias toward too much short-period (2–10 day) variability along the West Coast and upper Midwest, and too little variability in the southern Great Plains and the Gulf Coast. Model-simulated precipitation variability at 30 months or longer accounts for an anomalously large proportion of the total variance in the southeastern United States and 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 time scales longer than a year in the Pacific Northwest. Disagreements across the models are large at these longer periods.

c. 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 FDBC process is given in section S1 of the supplemental material.

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 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 CCSM4 are illustrated at a location in central Nevada (hot, dry) and a location in western Washington State (cool, wet) in Fig. 7a. The error in the model's representation of the spectrum of variability decreases substantially after FDBC is applied [i.e., green circles in Fig. 7 (right) are much closer to 1].

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

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

where the angle brackets indicate the mean over the logarithmically spaced frequency values. This expression treats equal ratios of error equally (i.e., the model having twice the observed variance produces the same
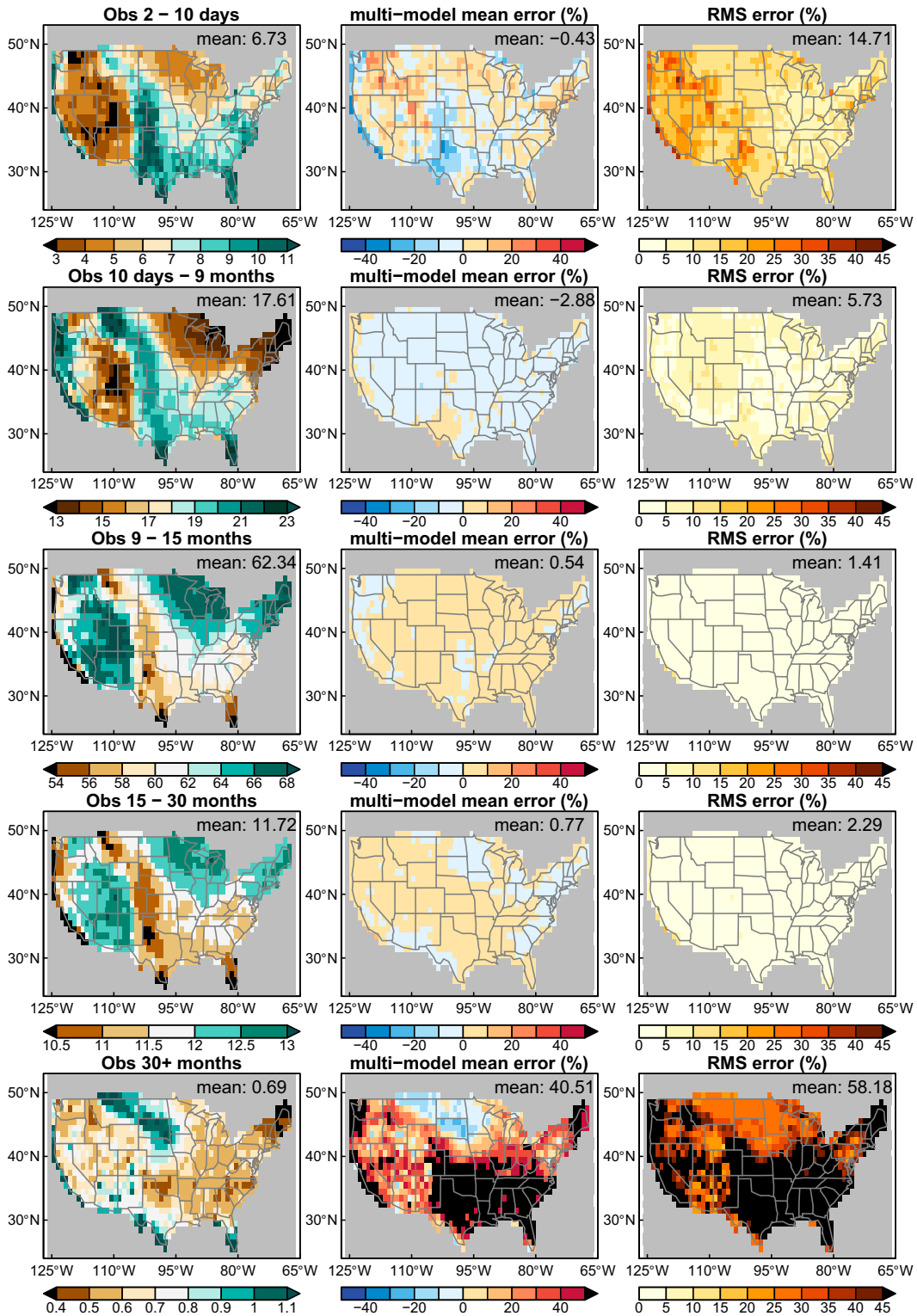
FIG. 5. (left) Proportion (%) of total variance of daily max temperature that falls in the frequency band whose period is indicated in the header, from observations over the period 1976–2005. Note that the color range varies substantially by frequency band. (middle) The multimodel mean error (%) for the same quantity in the GCMs, relative to the observations. (right) The multimodel RMSE (%).
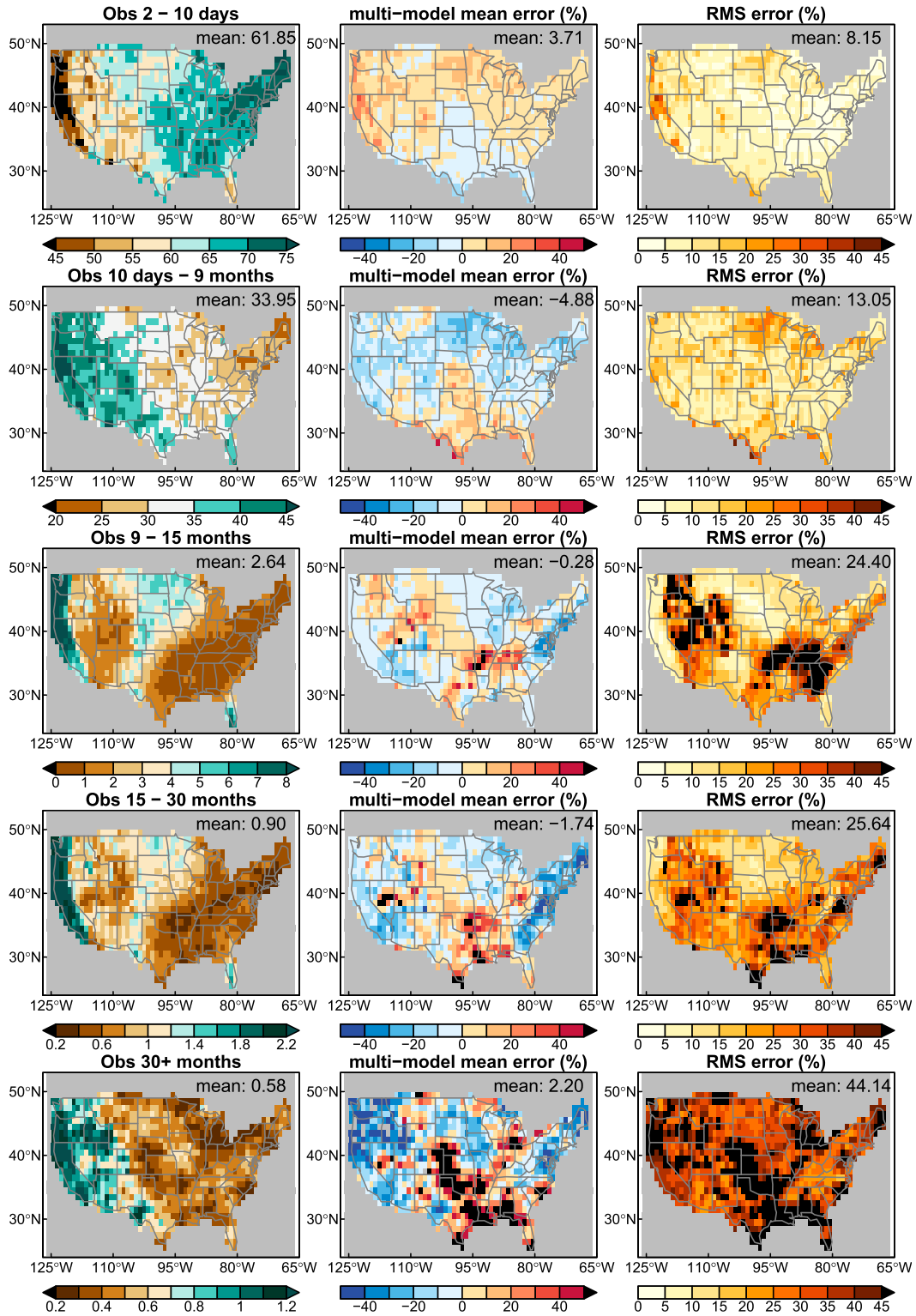
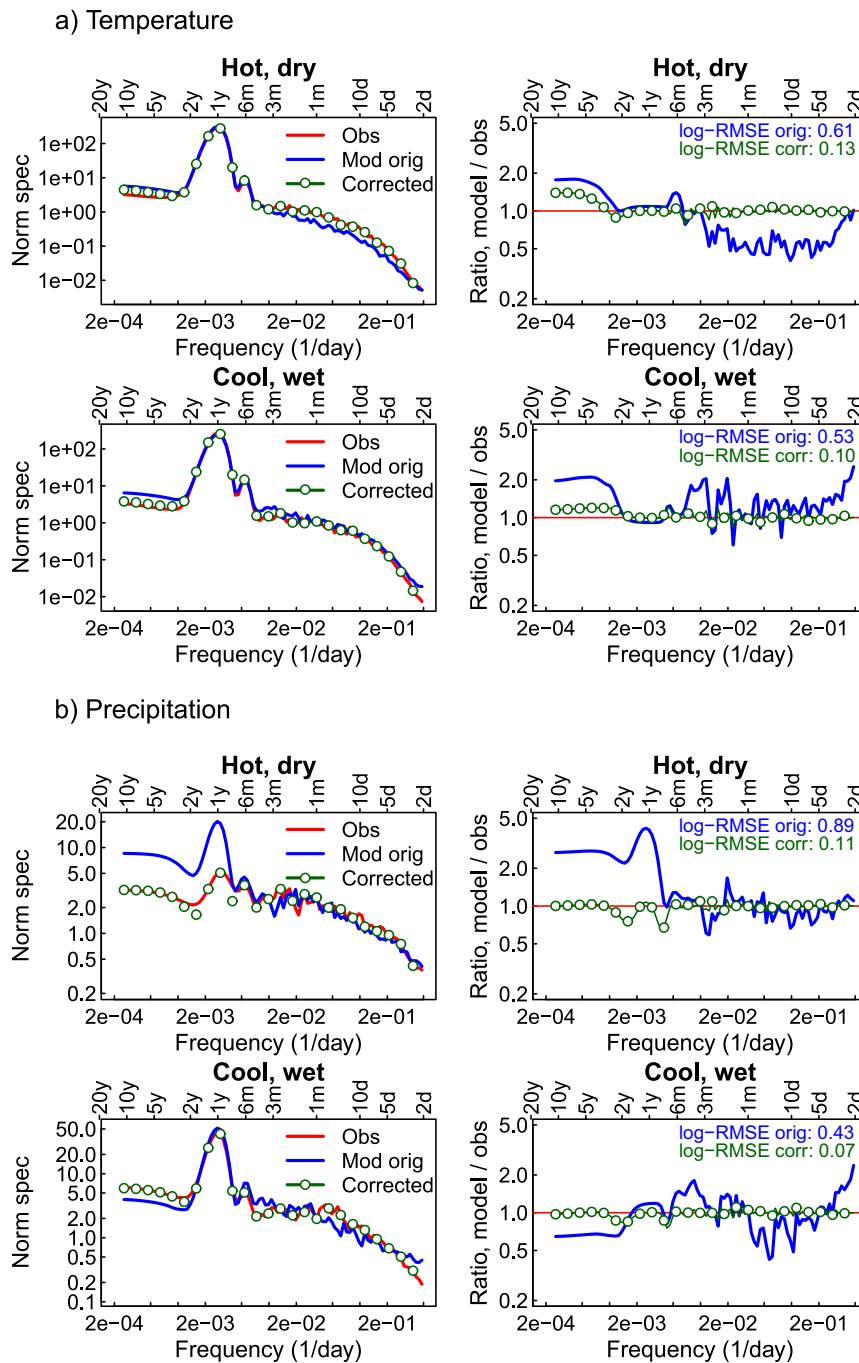FIG. 6. As in Fig. 5, but for daily precipitation.

FIG. 7. For daily max (a) temperature and (b) precipitation, normalized spectra from observations (red line), CCSM4 (blue line), and CCSM4 after FDBC (green dots and line) (left) and ratio of CCSM4 spectral power to observations before (blue line) and after FDBC (green dots and line) (right). Values are shown at a hot, dry location in central Nevada (39.5°N, 116.5°W) and a cool, wet location between Seattle and Portland (46.5°N, 122.5°W), as indicated in the header.

error as the observations having twice the model's variance), and the final −1 makes a perfect result (model variance equal observed, so $\sigma = 1$) give a log-RMSE of 0. In general, if the model values are incorrect (on average across log-spaced frequencies) by a factor of σ, then the log-RMSE is $\sigma - 1$. These log-RMSE values are indicated in Fig. 7 (right). When we refer to log-RMSE below, we specifically mean the model's error in
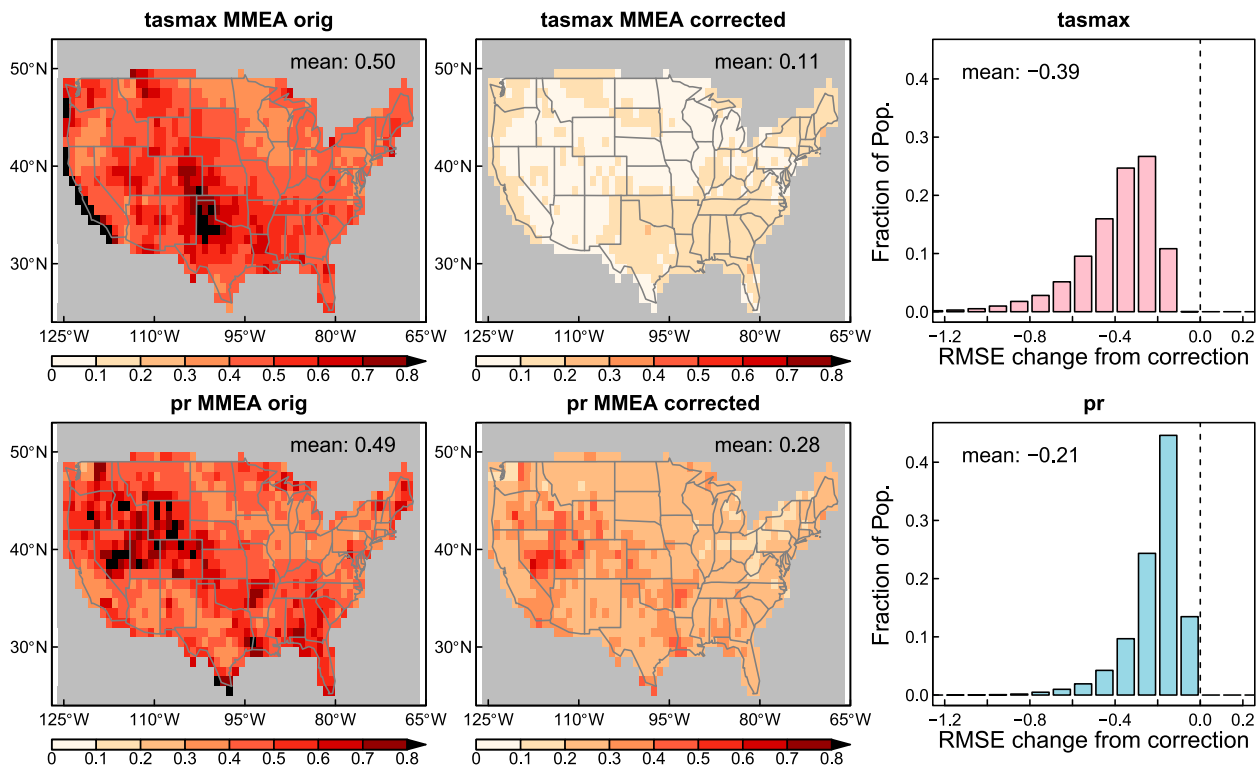
FIG. 8. For daily max (top) temperature and (bottom) precipitation, the multimodel ensemble average log-RMSE in simulating the observed distribution of variance across frequency, both (left) before and (middle) after the FDBC. (right) Histograms of how the FDBC changes the log-RMSE, taken over all models and all locations.

reproducing the distribution of variance across frequencies, as illustrated in Fig. 7.

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

Precipitation results at the two example locations are shown in Fig. 7b. CCSM4 shows a much stronger than observed annual cycle at the hot dry location, 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 multimodel ensemble average log-RMSE for daily maximum temperature is shown in Fig. 8 (top) both before (Fig. 8, top left) and after (Fig. 8, top middle) 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 Fig. 8 (bottom). The models do worse in the Rocky Mountains and the 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 Fig. 8 (right) show the difference between each grid cells' 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 Fig. 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
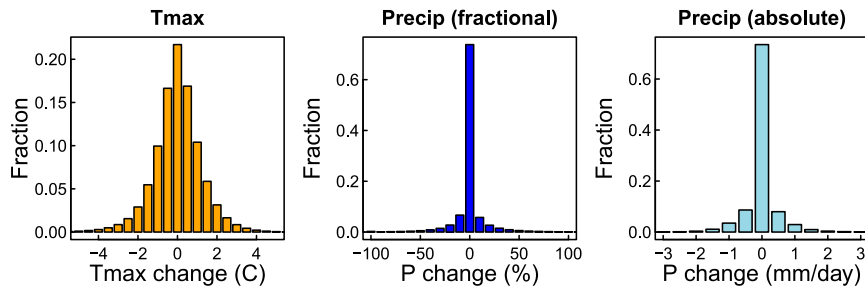
FIG. 9. Histograms of how much the FDBC alters the daily (left) temperature (°C) and (middle),(right) precipitation. The precipitation results are given both as the fraction change (%) and absolute change (mm day^{-1}). Results are shown for all the models across all points in the CONUS.

distribution of variance across frequencies without altering the overall variance, the mean changes are approximately zero.

COMBINED EFFECTS OF STANDARD AND FREQUENCY-DEPENDENT BIAS CORRECTION

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

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

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

To evaluate the effect of FDBC on runoff in a hydrological simulation, we used the VIC (Liang et al. 1994), configured for the western United States and forced over the period 1950–99 with four sources of daily

temperatures and precipitation: 1) observations (Livneh et al. 2013), 2) CCSM4, 3) CCSM4 fields bias corrected using QM (since this is a historical simulation), and 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 10 times too much spectral power in a given frequency band, while −1 means 10 times 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 (Fig. 10b) improves the simulation markedly, while FDBC (Fig. 10c) improves it somewhat more. Averaged across points in the domain, the mean error after bias correction is greatest at highest frequencies (Fig. 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. Preconditioning 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., 31 January is corrected using information from 1 January, which is 30 days away, but no information from 1 February, which is only 1 day away). To reduce these discontinuities Thrasher et al. (2012) 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) GCM power err @ 30 yrs

Mean: 0.91 RMS: 1.50

b) BC power err @ 30 yrs

Mean: −0.12 RMS: 0.23

c) BC+FDBC power err @ 30 yrs

Mean: −0.08 RMS: 0.19

d) Domain avg mean error

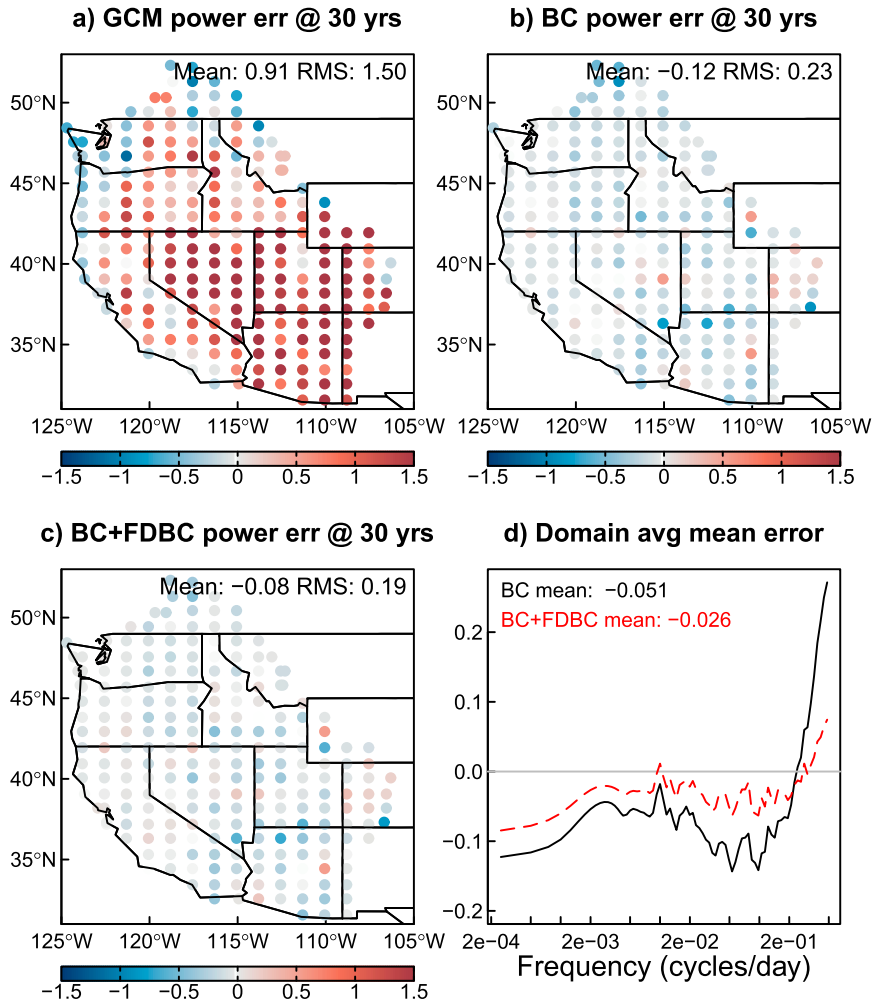BC mean: −0.051
BC+FDBC mean: −0.026

Frequency (cycles/day)

FIG. 10. Analysis of runoff simulated by the VIC with various meteorological forcing fields. (a) Error in simulated spectral power of runoff (at a period of 30 years) when VIC is forced with temperature and precipitation fields from CCSM4, where error is defined as $\log_{10}$ (power using GCM forcing/power using observed forcing). (b) As in (a), but for bias-corrected GCM forcing fields. (c) As in (a), but for bias-corrected and FDBC fields. (d) Domain-averaged mean error as a function of frequency; black line is for bias-corrected forcing fields and dashed red line is for bias-corrected and FDBC forcing fields.

A drawback to using a time window of a month is that many weather extremes can occur anytime over a multimonth season. For example, the 20 highest values of California-averaged daily precipitation over the period 1930–2002 have occurred as early as November and as late as February, while extreme hot days have occurred as early as June and as late as September. Ideally, the largest model value would be bias corrected to the largest observed value even if the maximum fell at the beginning of the season in the observations and the end of the season in the model. This argues for using a time window that is no narrower than a multimonth season if the extremes are distributed over a season. (Of course, if the variable being bias corrected truly does have all its

extreme values fall in a single month of the year, then a single-month time window is appropriate.) A more complete illustration of the problems obtained when using a 31-day sliding time window is given in section S2 of the supplemental material.

In this work we apply bias correction over a 91-day window, chosen to be wide enough to encompass seasonal weather phenomena. To address the issue of discontinuities at the edges of time windows, we iteratively apply the bias correction two additional times, with windows of 181 and 365 days, respectively. This ensures that every day is bias corrected with at least some information from adjoining days no matter where it falls in the initial 91-day window. A similar approach, dubbed

nested bias correction, was adopted by Johnson and Sharma (2012), although they used it for a different purpose than is done here. We use fixed, nonoverlapping time windows rather than moving ones to avoid the complications of matching quantiles in datasets with greatly different sizes. For example, consider the case described above of bias correcting a single central day-of-year using training data from the surrounding 31-day window, and the whole process is moved through the year. In a 50-yr record the training data will consist of $50 \times 31 = 1550$ days while the data to be corrected will consist of only 50 days. It is not straightforward to match the most extreme event in a 50-event record to the most extreme event 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 the 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-yr 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 CCSM4 in Fig. 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 Fig. S1 of the supplemental material.) 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 (Fig. 11a), which are reduced with a simple monthly bias correction (Fig. 11b). Correcting a day at a time based on statistics of a surrounding 31-day window yields the least error (Fig. 11c). 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 (Figs. 11d,e). However, preconditioning helps substantially (Fig. 11f), giving a result with less error than monthly bias correction although somewhat more than with the sliding central day in a 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 CCSM4 and observations at a point in the central Sierra Nevada (37.5°N, 119.5°W; 1976–2005). In a perfect model, values would fall along the diagonal (gray). Before bias correction (Fig. 12a), the model under-represents the strongest events by a factor of 2. Simple monthly bias correction (Fig. 12b) and using the central day in a 31-day sliding window (Fig. 12c) improve the representation considerably, but still with errors. Using a wide bias correction window gives good agreement between the observed and model-simulated extrema (Figs. 12d,e). Preconditioning, which addresses the annual cycle rather than the extremes, has little effect on this measure (Fig. 12f).

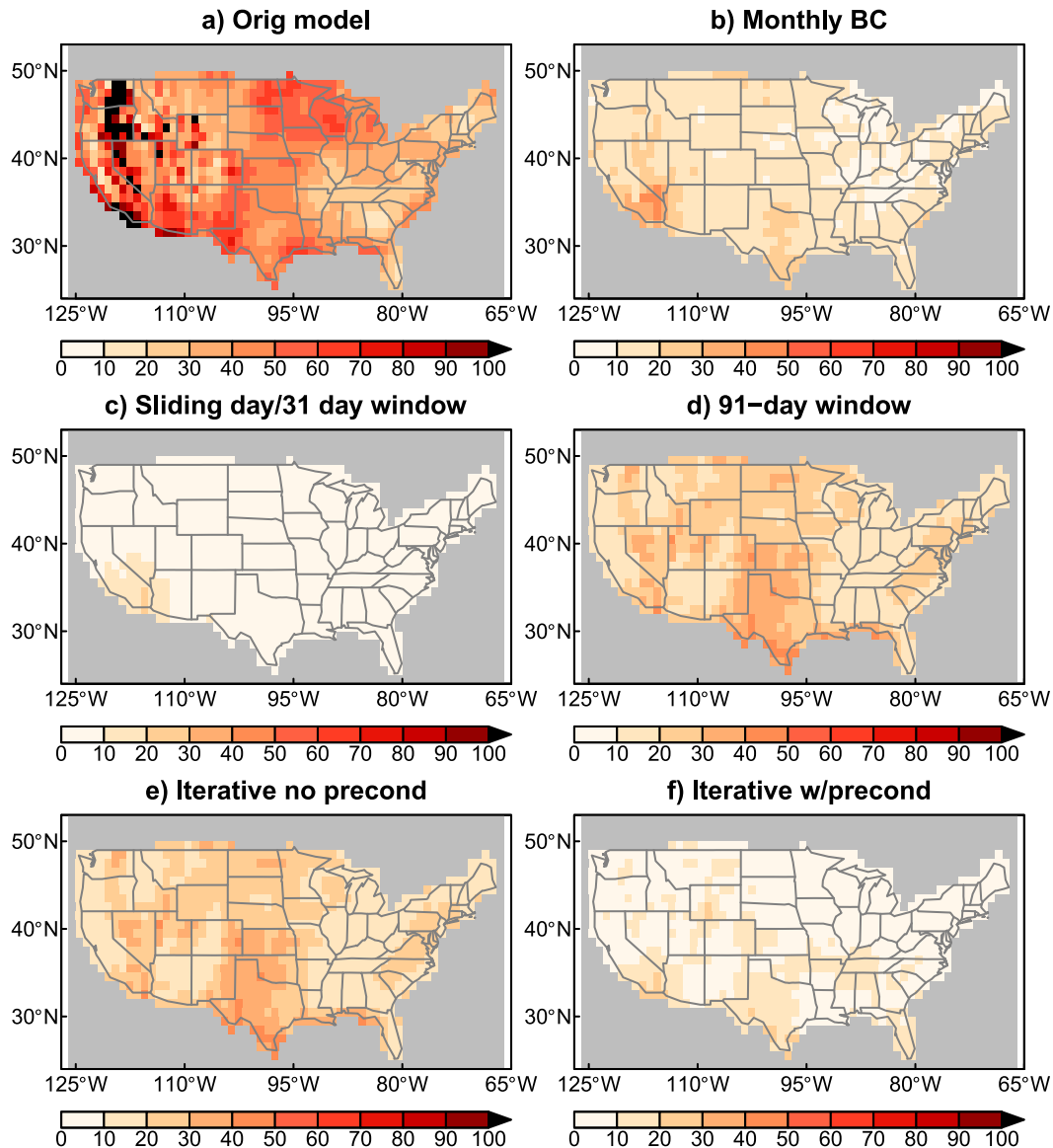Summary statistics of the modeled representation of extremes at every grid cell can be obtained by fitting a

FIG. 11. RMSE (percent of climatological annual mean value) in the annual cycle of precipitation (smoothed with a 31-day boxcar filter) as simulated by CCSM4, using various bias correction approaches. (a) Original model (no bias correction); (b) simple monthly bias correction; (c) single central day corrected based on statistics of a sliding 31-day window; (d) 91-day window; (e) iterative bias correction with 91-, 181-, and 365-day windows, but no preconditioning; and (f) iterative bias correction with preconditioning.

line between the top five observed and model extremes (red dashed lines in Fig. 12). The slopes and intercepts of the lines at all locations can then be mapped (Fig. 13). A perfect model representation of extremes would give a slope of 1 and intercept of 0. By this measure, the original model (Figs. 13a,b) has appreciable errors in its representation of daily extremes, as does the model after bias correction using either simple monthly bias correction (Figs. 13c,d) or bias correction using a central day in a sliding 31-day window (Figs. 13e,f). Using a wider, 91-day window improves

the representation considerably (Figs. 13g,h), and iterating over the 91-, 181-, and 365-day windows gives excellent agreement between the model and observations (Figs. 13i,j).

In summary, bias correction techniques that map one distribution to another are not optimally suited for correcting the annual cycle. The traditional solution of applying the correction in time windows of about a month is not necessarily a good fit with weather extremes, which in many locations can occur anytime in a multimonth season. To get around this problem, we
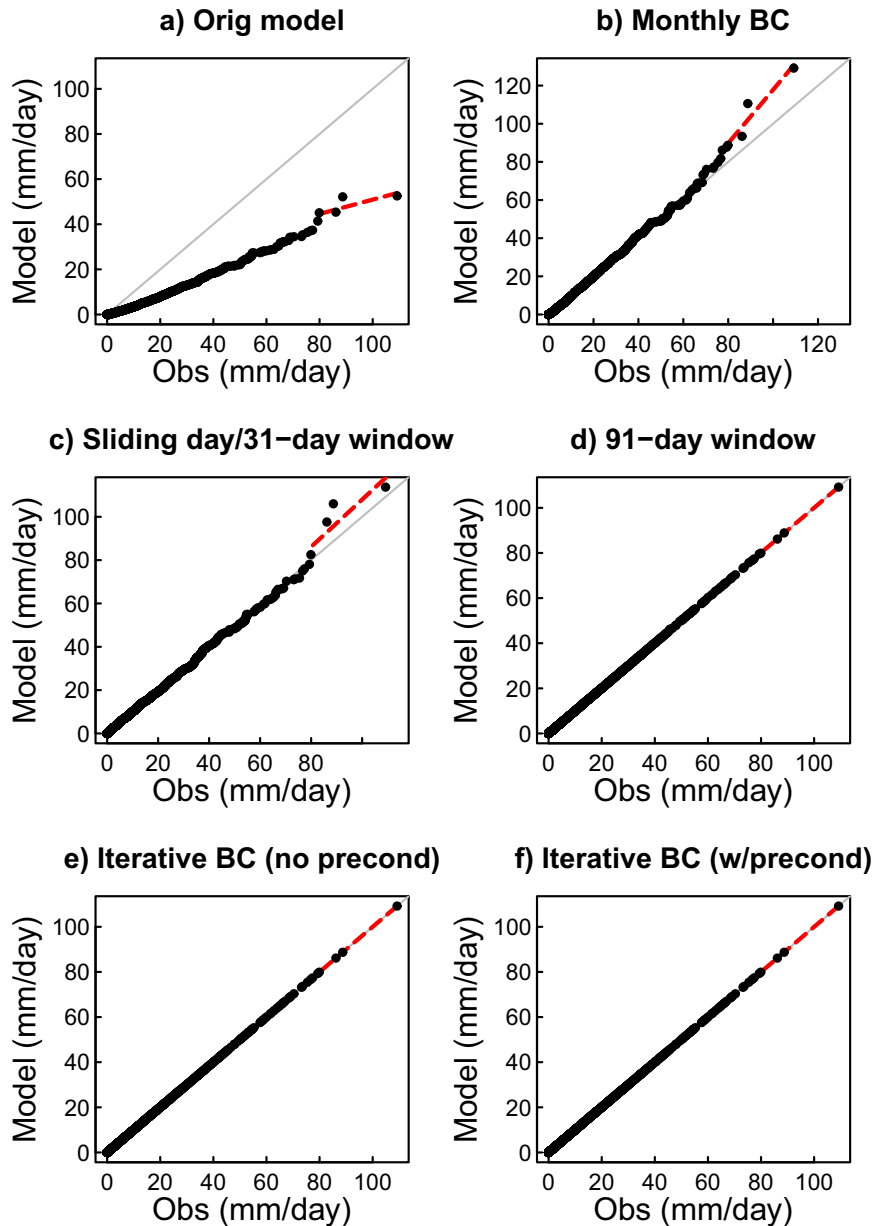
FIG. 12. Scatterplot of sorted daily precipitation values, observed vs model with the following bias correction applied: (a) original model (no bias correction); (b) simple monthly bias correction; (c) single central day corrected based on statistics of a sliding 31-day window; (d) 91-day window; (e) iterative bias correction with 91-, 181-, and 365-day windows, but no preconditioning; and (f) iterative bias correction with preconditioning. The dashed red line shows the best fit least squares line based on the five largest values. Model data are from CCSM4 at a point in the Sierra Nevada (37.5°N, 119.5°W), over the period 1976–2005.

use a simple preconditioning step that improves the representation of the annual cycle along 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
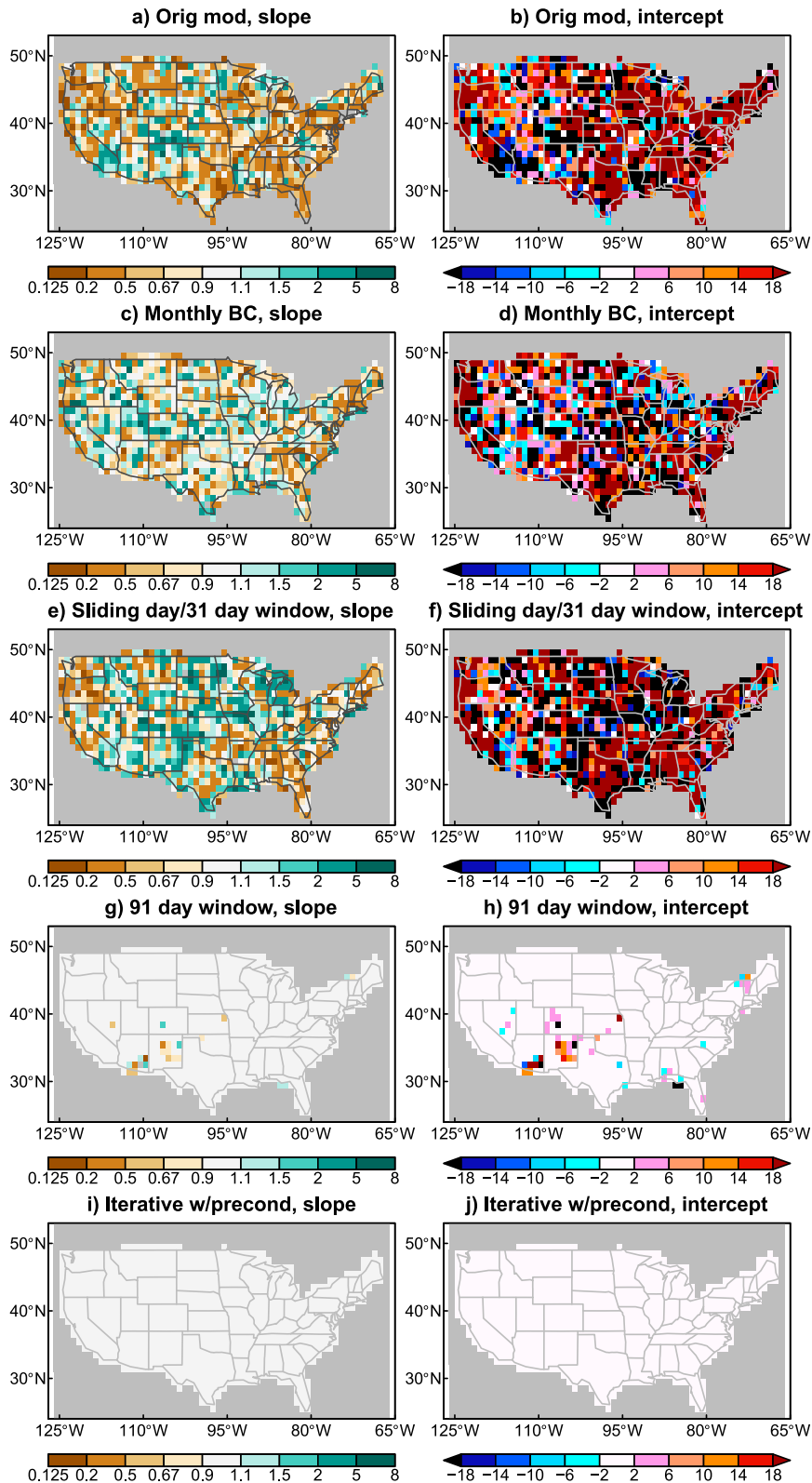
FIG. 13. The (left) slope and (right) intercept of the best-fit least squares line between the top five observed and modeled (CCSM4) extreme events for different bias correction approaches as indicated in the header.

to remove these biases before using the model-simulated fields in applications that have nonlinear 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% 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 model-predicted mean precipitation change in locations that are so dry there are insufficient precipitation days to bias correct.

We also examined the more general issue of the models' representation of variance across a range of time scales and introduced an FDBC method that reduces inaccuracies in the GCMs' spectra. As a group, the 21 GCMs apportion too little variability of daily maximum temperature to time scales between 10 and 90 days and too much to time scales longer than 30 months. The models' simulation of daily precipitation variability was more mixed, but at long time scales (>30 months) they show more variability than observed in the Gulf Coast region and less than observed in the Pacific Northwest. These problems can be reduced by a frequency-dependent bias correction implemented as digital filter in the frequency domain. This is one step toward addressing time-dependent model biases, an important subject that has many implications for impacts such as droughts and heat waves. We implement the FDBC as a separate step following the EDCDFm or PresRat bias correction, which means this step could be combined with any other existing bias correction method (such as quantile mapping or CDF-t) as well. However, the current implementation operates on the entire time series of daily values, so frequency-dependent errors on the seasonal or monthly time scale can persist under some circumstances.

Traditional bias correction is done in a time window, often of about a month, to reduce errors in the annual cycle. However, in many locations weather extremes can occur sometime during a multimonth season, which argues for using a time window on the order of a season in such places. A simple preconditioning technique has been shown to yield a good simulation of the seasonal cycle even when using a seasonwide time window. The end result captures both the 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 preselected 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.

APPENDIX

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 nonzero frequency corresponds to a period of ~11 years. Longer periods are unresolved, and the FDBC 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 overfitting. 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 five 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) across constant widths of the logarithmic frequency axis. However, these approaches change the angle of a plotted spectrum (e.g., a white spectrum is then no longer flat), which can be confusing. To avoid this potentially misleading situation, values shown here are simply averaged in frequency so that the spectra appear similar to what is typically found in the literature (i.e., a white spectrum is flat).

REFERENCES

Dai, A., 2006: Precipitation characteristics in eighteen coupled climate models. *J. Climate*, **19**, 4605–4630, doi:10.1175/JCLI3884.1.

Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical–topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor.*, **33**, 140–158, doi:10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2.

Deser, C., R. Knutti, S. Solomon, and A. S. Phillips, 2012: Communication of the role of natural variability in future North American climate. *Nat. Climate Change*, **2**, 775–779, doi:10.1038/nclimate1562.

Ghil, M., and Coauthors, 2002: Advanced spectral methods for climatic time series. *Rev. Geophys.*, **40**, 1003, doi:10.1029/2000RG000092.

Hagemann, S., C. Chen, J. O. Haerter, J. Heinke, D. Gerten, and C. Piani, 2011: Impact of a statistical bias correction on the projected hydrological changes obtained from three GCMs and two hydrology models. *J. Hydrometeor.*, **12**, 556–578, doi:10.1175/2011JHM1336.1.

IPCC, 2007: *Climate Change 2007: Impacts, Adaptation, and Vulnerability.* Cambridge University Press, 976 pp.

——, 2013: *Climate Change 2013: The Physical Science Basis.* Cambridge University Press, 1535 pp.

Jenkins, G. M., and D. G. Watts, 1969: *Spectral Analysis and Its Applications.* Holden-Day, 525 pp.

Johnson, F., and A. Sharma, 2012: A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations. *Water Resources Res.*, **48**, W01504, doi:10.1029/2011WR010464.

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

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land surface water and energy fluxes for GSMs. *J. Geophys. Res.*, **99**, 14 415–14 428, doi:10.1029/94JD00483.

Livneh, B., E. A. Rosenberg, C. Lin, V. Mishra, K. M. Andreadis, E. P. Maurer, and D. P. Lettenmaier, 2013: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *J. Climate*, **26**, 9384–9392, doi:10.1175/JCLI-D-12-00508.1.

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

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

Maurer, E. P., and D. W. Pierce, 2014: Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean. *Hydrol. Earth Syst. Sci.*, **18**, 915–925, doi:10.5194/hess-18-915-2014.

——, 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, doi:10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2.

——, L. Brekke, T. Pruitt, and P. B. Duffy, 2007: Fine-resolution climate projections enhance regional climate change impact studies. *Eos, Trans. Amer. Geophys. Union*, **88**, 504, doi:10.1029/2007EO470006.

——, H. G. Hidalgo, T. Das, M. D. Dettinger, and D. R. Cayan, 2010: The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California. *Hydrol. Earth Syst. Sci.*, **14**, 1125–1138, doi:10.5194/hess-14-1125-2010.

——, and Coauthors, 2014: An enhanced archive facilitating climate impacts and adaptation analysis. *Bull. Amer. Meteor. Soc.*, **95**, 1011–1019, doi:10.1175/BAMS-D-13-00126.1.

Michelangeli, P.-A., M. Vrac, and H. Loukos, 2009: Probabilistic downscaling approaches: Application to wind cumulative

distribution functions. *Geophys. Res. Lett.*, **36**, L11708, doi:10.1029/2009GL038401.

Panofsky, H. A., and G. W. Brier, 1968: *Some Applications of Statistics to Meteorology*. The Pennsylvania State University, 224 pp.

Pierce, D. W., and Coauthors, 2013: Probabilistic estimates of future changes in California temperature and precipitation using statistical and dynamical downscaling. *Climate Dyn.*, **40**, 839–856, doi:10.1007/s00382-012-1337-9.

Polade, S. D., D. W. Pierce, D. R. Cayan, A. Gershunov, and M. D. Dettinger, 2014: The key role of dry days in changing regional climate and precipitation regimes. *Sci. Rep.*, **4**, 4364, doi:10.1038/srep04364.

Rupp, D. E., J. T. Abatzoglou, K. C. Hegewisch, and P. W. Mote, 2013: Evaluation of CMIP5 20th century climate simulations for the Pacific Northwest USA. *J. Geophys. Res. Atmos.*, **118**, 10 884–10 906, doi:10.1002/jgrd.50843.

Sheffield, J., and Coauthors, 2013: North American climate in CMIP5 experiments. Part II: Evaluation of historical simulations of intraseasonal to decadal variability. *J. Climate*, **26**, 9247–9290, doi:10.1175/JCLI-D-12-00593.1.

Sun, Y., S. Solomon, A. Dai, and R. W. Portmann, 2006: How often does it rain? *J. Climate*, **19**, 916–934, doi:10.1175/JCLI3672.1.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.

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

Thrasher, B., E. P. Maurer, C. McKellar, and P. B. Duffy, 2012: Technical note: Bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrol. Earth Syst. Sci.*, **16**, 3309–3314, doi:10.5194/hess-16-3309-2012.

Trewin, B. C., 2007: The role of climatological normals in a changing climate. WMO Tech. Directive WCDMP-61 and WMO-TD 1377, 40 pp. [Available online at http://www.wmo.int/datastat/documents/WCDMPNo61_1.pdf.]

von Storch, H., and F. W. Zwiers, 2001: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.

Wang, L., and W. Chen, 2014: A CMIP5 multimodel projection of future temperature, precipitation, and climatological drought in China. *Int. J. Climatol.*, **34**, 2059–2078, doi:10.1002/joc.3822.

Wigley, T. M. L., and P. D. Jones, 1985: Influences of precipitation changes and direct CO_2 effects on streamflow. *Nature*, **314**, 149–152, doi:10.1038/314149a0.

Wood, A. W., E. P. Maurer, A. Kumar, and D. P. Lettenamier, 2002: Long-range experimental hydrologic forecasting for the eastern United States. *J. Geophys. Res.*, **107**, 4429, doi:10.1029/2001JD000659.