**Hydrology and
Earth System
Sciences**

# The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California

E. P. Maurer[1], H. G. Hidalgo[2], T. Das[3], M. D. Dettinger[4,3], and D. R. Cayan[3,4]

[1]Civil Engineering Dept., Santa Clara University, Santa Clara, CA, USA
[2]Escuela de Física, Universidad de Costa Rica, San José, Costa Rica
[3]Scripps Institution of Oceanography, La Jolla, CA, USA
[4]US Geological Survey, Water Resources Division, La Jolla, CA, USA

**Abstract.** Three statistical downscaling methods were applied to NCEP/NCAR reanalysis (used as a surrogate for the best possible general circulation model), and the downscaled meteorology was used to drive a hydrologic model over California. The historic record was divided into an "observed" period of 1950–1976 to provide the basis for downscaling, and a "projected" period of 1977–1999 for assessing skill. The downscaling methods included a bias-correction/spatial downscaling method (BCSD), which relies solely on monthly large scale meteorology and resamples the historical record to obtain daily sequences, a constructed analogues approach (CA), which uses daily large-scale anomalies, and a hybrid method (BCCA) using a quantile-mapping bias correction on the large-scale data prior to the CA approach. At 11 sites we compared three simulated daily flow statistics: streamflow timing, 3-day peak flow, and 7-day low flow. While all downscaling methods produced reasonable streamflow statistics at most locations, the BCCA method consistently outperformed the other methods, capturing the daily large-scale skill and translating it to simulated streamflows that more skillfully reproduced observationally-driven streamflows.

## 1 Introduction

As climate change science matures and is better able to estimate the regional magnitudes of potential climate change, estimates of local and regional impacts to the resources at risk are of increasing interest (IPCC, 2007a). As for much of the globe, western United States (US) water resources, the focus of this study, are particularly at risk, which has inspired a plethora of recent studies aimed at estimating potential impacts to hydrology and water resources systems (Barnett et al., 2008; Cayan et al., 2008; Maurer, 2007; Vicuna et al., 2007).

One common issue facing all regional assessments of climate change impacts is that the scale of general circulation model (GCM) outputs are at too spatially coarse a scale for direct use in impact models. Regional studies, such as those examining hydrologic impacts of climate change, thus rely on spatial downscaling to translate the large-scale climatic shifts projected by GCMs to scales more representative of local areas of interest (Christensen et al., 2007).

The recent availability of large databases of raw GCM outputs in a consistent format (Meehl et al., 2007) has facilitated the use of multiple GCMs and greenhouse gas emissions scenarios in impact studies. The greatest value from studies of multiple GCM runs is that model-to-model, scenario-to-scenario, and even chaotic realization-to-realization uncertainties in the physical response of the climate system to changing greenhouse gas concentrations, the primary sources of uncertainty in climate impacts analysis (Fowler and Ekström, 2009), can be quantified to some degree (Hawkins and Sutton, 2009). Furthermore, the skill of a multimodel ensemble consistently outperforms any individual model for detection and attribution studies (Brekke et al., 2008; Gleckler et al., 2008; Pierce et al., 2009). To consider many future projections of climate in a regional impacts study requires a downscaling procedure that is computationally very efficient. This generally limits these studies to using statistical downscaling techniques, where some large-scale signal is related statistically to local climate, as opposed to regional climate simulations, where a dynamical model of regional climate is used to simulate local climate responses to the global

projected changes. Past studies generally have found that the differences between GCMs is much greater than the uncertainty in downscaling techniques (Fowler et al., 2007).

While projected changes in long-term mean water supply may have dire consequences for society (IPCC, 2007b; Oki and Kanae, 2006; Shen et al., 2008), changes in the frequency of extreme events are also of critical concern (Katz et al., 2002), especially for regional hydroclimate (Leung et al., 2004). Most climate projections suggest increases in the frequency of temperature and precipitation extremes, both at the monthly level (Benestad, 2006) and at the daily (and subdaily) level (Kharin et al., 2007). When downscaling from GCM-scale climate simulations to regional scales to study hydrologic impacts, the most desirable downscaling methods have the ability to translate the local changes in climatic extreme events simulated by GCMs to the local scale needed by hydrological models . While dynamic downscaling, using a regional climate model (RCM) driven at the boundary by a GCM, has been used in the western US to produce physically realistic projections of changes in hydrologic extremes (Kim et al., 2002; Snyder and Sloan, 2005), these types of models are still too computationally intensive to be applied to a large ensemble of GCM output to characterize uncertainties associated with inter-GCM variability and different emission scenarios. For this reason, more computationally efficient statistical downscaling approaches will continue to serve as the methodological workhorse for downscaling ensembles of long climate simulations.

Statistical methods, building statistical relationships between GCM-scale climate features and fine scale climate and applying those to future projections, have been more widely applied than dynamical model downscaling in studies of hydrologic impacts of climate change over the western United States (Christensen and Lettenmaier, 2007; Maurer et al., 2007; Payne et al., 2004; Wood et al., 2004). In most applications the focus has been on monthly, seasonal or annual hydrologic changes and generally only monthly GCM output was used. Some efforts have used daily GCM output to study extremes in this region (e.g., Dettinger et al., 2004), though this approach has generally been to downscale GCM output directly to specific weather stations. To characterize both projected seasonal and extreme changes for larger watersheds or over continental areas, a downscaling method should have the ability to generate gridded fields of downscaled daily climate, to capture the spatial structure of climate features. To achieve this using daily GCM output was a motivation for the development of the constructed analogues (CA) approach (Hidalgo et al., 2008).

In a prior effort (Maurer and Hidalgo, 2008) the CA approach was contrasted with the bias-correction/spatial disaggregation (BCSD) statistical downscaling approach, with each applied over the western US for downscaling large-scale observationally-derived reanalysis data as a surrogate GCM. The methods take different approaches to downscaling daily extreme precipitation and temperature. CA down-scales each day's output from the GCM simulation, capturing projected changes in daily weather events that sum together to reflect long-term climate changes, while BCSD works with GCM monthly output, then randomly selects a month from the historical record and rescales its daily precipitation and temperature to match the projected monthly values. Each has the ability to downscale to a gridded field over a wide region, maintaining spatial correlations of driving hydroclimatic conditions that drive hydrologic impacts. Wood et al. (2004) found the BCSD method performed well when compared to several statistical and dynamic downscaling methods in the context of assessing hydrologic impacts. The ability of the CA method to exhibit considerable skill of daily precipitation and temperature statistics has also been demonstrated (Hidalgo et al., 2008). Both methods have been widely used in regional studies in the United States and globally (Barnett et al., 2008; Cayan et al., 2009; Das et al., 2009; Girvetz et al., 2009; Hayhoe et al., 2007; Maurer et al., 2009).

Both methods are capable, to some degree, of capturing projected changes in extremes. They have been shown to produce similar downscaling skill for many measures of temperature and precipitation extremes (Maurer and Hidalgo, 2008). In that study, both CA and BCSD exhibited limited skill, attributed to substantial large scale precipitation biases, for both wet and dry daily precipitation extremes and the difference between the methods was not significant. Statistically significant differences were apparent, however, for some measures, notably that CA demonstrated better skill for downscaling cold-season low temperature extremes and warm season high temperature extremes. This illustrated the ability of CA to successfully translate large-scale daily skill to a fine scale, where by contrast, the BCSD method, using the assumption that climatological intra-monthly variability does not change, showed lower skill. There were several important questions raised by this prior study, which are the focus of this paper:

1. Is there any difference in the hydrology simulated by climate downscaled with these methods?

2. For extreme streamflow measures, do the downscaling approaches produce different results?

3. Are there opportunities to combine the best attributes of the methods to improve downscaling performance?

Ultimately, the goal of this study is to address question 3, by identifying, testing, and developing an improved statistical downscaling method capable of skillfully downscaling extreme hydroclimate, while being applied at regional to continental scales. To do this we refine the prior analysis in Maurer and Hidalgo (2008) to evaluate how differences in the downscaling approaches propagate through the hydrologic system, and to determine whether improvements in downscaling methods, especially in the context of simulating hydrologic extremes, may be possible.

## 2 Methods and data

The approach for this study follows that of Maurer and Hidalgo (2008), in which National Center of Environmental Prediction and the National Center of Atmospheric Research (NCEP/NCAR) reanalysis (Kalnay et al., 1996) was used as a surrogate for a General Circulation Model (GCM) output, as others have done (e.g., Widmann et al., 2003). The benefit of using reanalysis rather than GCM output is that biases will be expected to be lower, since atmospheric observations are assimilated in the reanalysis framework. In addition, the data assimilation process produces year-to-year and day-to-day correspondence to observed climate and weather that an unconstrained GCM would not, making it more defensible to compare downscaling performance against observations. We downscale the reanalysis daily and monthly precipitation and temperature using different techniques, and use the downscaled data to drive a hydrologic model. The hydrologic model skill is evaluated by comparing these simulations to the hydrologic model output produced by driving the hydrologic model directly with the gridded observed precipitation and temperature of Maurer et al. (2002). Since one of the downscaling methods, BCSD, uses random daily sequences that are not connected to observed daily meteorology, statistics typically used for hydrologic model performance, such as root mean square error or correlation, which assume observed and simulated sequences to correspond to the same events, are not used. Rather, we opt for statistical tests that assess whether the flows produced with downscaled data have the same statistical characteristics as those driven by observed meteorology. Results are compared primarily using a 2-sample Kolmogorov-Smirnov (KS) test (Wilks, 2006) at a 0.05 significance level, with other tests applied selectively as discussed in Sect. 3.2 below.

### 2.1 Reanalysis as a surrogate GCM

NCEP/NCAR reanalysis (Kalnay et al., 1996) data include daily and monthly precipitation and temperature on a T62 Gaussian grid (approximately 1.9° square), a resolution comparable to recent GCMs. Reanalysis is often held up as an example of the best possible historical GCM output (Reichler and Kim, 2008), which makes it appropriate for use in this study, as the focus is on how downscaling approaches distinguish themselves in the presence of large-scale skill. As noted by Maurer and Hidalgo (2008), because reanalysis temperature is strongly connected to observations, the comparisons of temperature skill will reflect differences almost exclusively in the downscaling techniques. However, because precipitation observations are not assimilated into reanalysis estimates, the intercomparison will reflect differences between the downscaling methods, plus influences of the reanalysis precipitation biases and errors. The precipitation and temperature daily variability in reanalysis has been shown to be realistic in many locations in the western US

(Widmann and Bretherton, 2000), and so the existence of skill in daily statistics of large-scale climate model output (in this case, reanalysis) will be a major factor potentially distinguishing the downscaling methods compared in this study.

Following Maurer and Hidalgo (2008), we divide the second half of the 20th century into two periods, with 1950–1976 representing "observations" used as the sample catalog from which model estimates are derived, and 1977–1999 "projections" for which the model estimates are derived and verified upon. The later period exhibits small but statistically significant differences in both temperature and precipitation compared to the early period, with 1977–1999 being generally wetter and warmer over the study domain. While there are documented climatic drivers that could explain this difference, such as a 1976/77 shift in the Pacific Decadal Oscillation phase (Mantua and Hare, 2002), there are also changes in the sources of observations assimilated in the NCEP/NCAR reanalysis beginning in 1979 (Kistler et al., 2001). These differences provide the opportunity to assess the performance of the downscaling techniques under a climate that, while not dramatically different, is statistically significantly different.

A more robust assessment of the comparative skill of the downscaling methods studied here could be designed by randomly assembling different sets of years for training the methods and for validation, and employing a cross-validation or bootstrapping method to assess skill (e.g., Feddersen, 2003; Feddersen et al., 1999; Li et al., 2010). For this study, the two periods were intentionally selected to be observed sequences of years that differ significantly in both temperature and precipitation, with the intention of creating a validation condition that could highlight differences between the methods and to explore the potential for improving the methods.

### 2.2 Downscaling techniques

The two primary downscaling techniques used in this study are the constructed analogues (Hidalgo et al., 2008; van den Dool, 1994) and bias correction and spatial downscaling (BCSD, Wood et al., 2004). These are described and contrasted in detail by Maurer and Hidalgo (2008). In general, BCSD corrects for large scale biases using coarse-resolution model output (from reanalysis or a GCM) and observations, and then interpolates the bias-corrected anomalies onto a fine-scale surface of observations. The CA technique begins with a library of observed daily coarse-resolution and corresponding high resolution climate anomaly patterns of the variable to be downscaled, with each day's library compiled from observations within ±45 days of the day to be downscaled (Hidalgo et al., 2008). To downscale each day, a subset of the 30 patterns (predictors) with the closest similarity to the simulated anomalies are found from the coarse-resolution library. A linear combination of the coarse-resolution version of the predictors is used to produce a coarse-resolution analogue, and the downscaled anomaly

is produced by applying the same linear combination to the 30 corresponding fine-scale library patterns. The most important distinction between the two methods is that by using daily reanalysis (or GCM) output CA retains the daily sequencing of weather events from the coarse resolution, while in BCSD only monthly reanalysis averages are used, with daily patterns reconstructed by randomly resampling a historic month and scaling its daily precipitation and temperature values to match the monthly projected values. Where a climate model exhibits skill in simulating daily variability, CA would in theory be capable of capturing that skill, while BCSD would reflect historical intra-month variability. Thus, for daily statistics, the two methods will be expected to distinguish themselves only inasmuch as the large-scale climate model exhibits skill at the daily time scale. Another distinction between BCSD and CA has been observed in areas near coasts and other areas with sharp climate gradients at a scale much finer than the large-scale climate model output being downscaled. While BCSD reproduces climatological patterns of precipitation and temperature, projected changes tend to be smooth spatially. CA by contrast captures changes in day-to-day variability, which can evolve differently than the large-scale forcing, and thus CA can produce sharper spatial gradients of precipitation and temperature changes than BCSD.

A second distinction between CA and BCSD that bears on the analysis that follows is that CA builds relationships between large-scale climate anomalies and fine-scale anomalies based on observations, and then applies those relationships to large-scale reanalysis (or GCM) anomalies. BCSD first bias corrects the large scale monthly reanalysis data, using a quantile-mapping approach (Panofsky and Brier, 1968), so that for each month there is a statistical match (for the observed period) for all statistical moments to those of large-scale observations, and the bias-corrected monthly data are then spatially downscaled. The implication of this is that while CA accounts for potential biases in the mean by using anomalies, higher order biases in reanalysis spatial or temporal variability feed directly into the CA downscaled results in ways that BCSD explicitly corrects and avoids.

## 2.3 Hydrologic modeling

To assess the ability to downscale to the watershed scale, daily downscaled meteorology is used to drive the variable infiltration capacity (VIC) hydrologic model (Cherkauer et al., 2003; Liang et al., 1994). VIC is a spatially distributed hydrologic model that solves the energy and water budgets at the land surface. It has been widely applied in forecasting and climate change analyses on spatial scales ranging from watershed to continental areas (Abdulla et al., 1996; Maurer, 2007; Maurer and Lettenmaier, 2003; Nijssen et al., 1997; Wood et al., 2002). In this study, we apply the VIC model at the same resolution (1/8 degree, approximately 12 km) and with the same parameterization as was used in
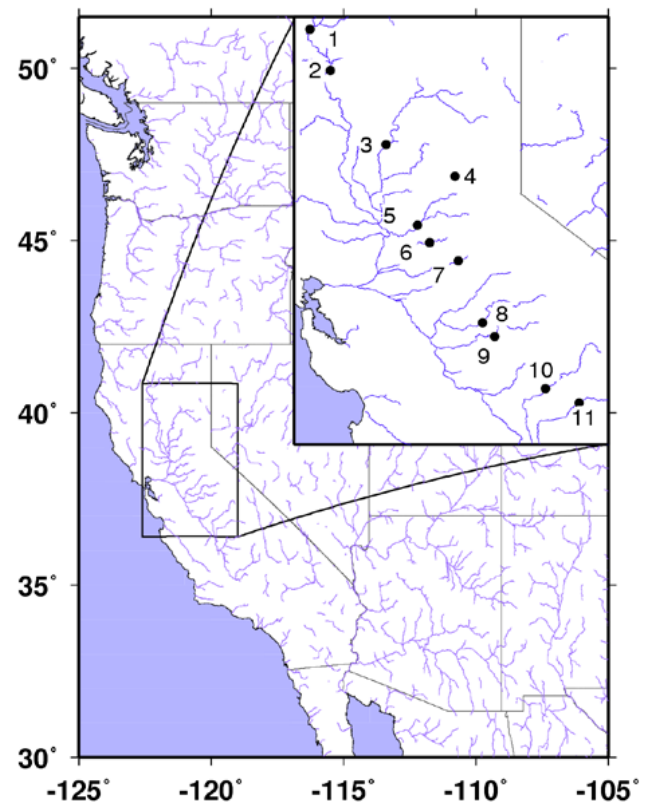


**Fig. 1.** Location map of the 11 streamflow gauges listed in Table 1.

several prior studies of the area (Barnett et al., 2008; Cayan et al., 2008). Prior studies have assessed the VIC model performance (with the same parameterization as in this study), comparing observed flows with those simulated by VIC being driven by the same gridded observed meteorology as used in this study. Maurer et al. (2007) used four of the sites used in the current study, and found observed flows were well simulated, with biases below 10%. The VIC model output is processed through a stream routing network following Lohmann et al. (1996), which is used to generate simulated flow at the stream gauge locations listed in Table 1 and shown on Fig. 1. These stream gauges are chosen to represent watersheds having much of their elevations above 1200 m, and thus being dominated by snowmelt (with the exception of the Consumnes, which has a smaller fraction of high elevation area).

## 3 Results and discussion

Large scale skill in reanalysis temperature data is well established, since observations of temperature are assimilated. This skill has been demonstrated for monthly data as well as for daily statistics. While precipitation is less well simulated in reanalysis, being model output rather than assimilated data, some skill is evident. We summarize below the

**Table 1.** Streamflow gauges included in this study.

| No. | Gauge | Name | Average Elevation, m |
|---|---|---|---|
| 1 | SHAST | Sacramento River at Shasta Dam | 1480 |
| 2 | SAC_B | Sacramento River at Bend Bridge | 1310 |
| 3 | OROVI | Feather River at Oroville | 1560 |
| 4 | NF_AM | North Fork American River at North Fork Dam | 1530 |
| 5 | FOL_I | American River at Folsom Dam | 1340 |
| 6 | CONSU | Cosumnes River at Michigan Bar | 640 |
| 7 | PRD_C | Mokelumne River at Pardee | 1580 |
| 8 | DPR_I | Tuolumne River at New Don Pedro | 1760 |
| 9 | LK_MC | Merced River at Lake McClure | 1810 |
| 10 | MILLE | San Joaquin River at Millerton Lake | 2120 |
| 11 | KINGS | Kings River – Pine Flat Dam | 2200 |

ability to recover fine scale precipitation and temperature statistics from the large-scale reanalysis, assess how the differences in downscaling skill affect hydrology, and develop a method for combining positive attributes of the two methods to improve downscaling skill.

### 3.1 Downscaling meteorology for assessing hydrologic impacts

Monthly and daily skill for downscaling precipitation and temperature using the two downscaling methods were analyzed in a prior study (Maurer and Hidalgo, 2008), which forms the basis for the current study. Monthly downscaling skills for CA and BCSD were found in that study to be comparable, as were their skill levels for daily extreme precipitation amounts (which was generally low for both methods, reflecting the lack of skill in precipitation simulation at the large native reanalysis scale). However, CA demonstrated better skill at some locations in downscaling some of the daily statistics, such as sequences of wet and dry days, and high and low temperature extremes, where the large-scale reanalysis data contain greater skill.

While correlations were higher for CA than for BCSD for some variables, correlation analysis is unable to pick up systematic biases in the large-scale data. For example, while Maurer and Hidalgo (2008) show comparable correlations with observations for both CA and BCSD downscaled reanalysis for monthly, daily, and extreme wet and dry precipitation amounts, Fig. 2 reveals that both methods produce bias in the downscaled precipitation intensity (the average rainfall rate on rainy days, defined as days with non-zero precipitation). Focusing on regions with high observed precipitation intensity, especially January in the Pacific Northwest (PNW) and the Sierra Nevada in California, two features emerge. Most notably, CA shows a large negative bias in precipitation intensity in California, and a positive bias in the PNW.

This bias in downscaled CA precipitation intensity in regions with relatively low precipitation is similar to the well-documented "drizzle" bias typical in GCMs (Iorio et al., 2004; Mearns et al., 1995), where weak precipitation events are overly common. Figure 3 illustrates that while reanalysis produces average precipitation intensities (for a grid point over central California) that appear reasonable, the frequency of occurrence of events at the lowest intensities is oversimulated. Approximately 40% of the daily January observations (from Maurer et al. (2002) aggregated to the Reanalysis grid resolution) show zero precipitation (Fig. 3, center panel, where the "OBS" line intersects the ordinate at a value of 0.4), while all days in the reanalysis have some precipitation (same panel, the dashed line never intersects the ordinate). At higher precipitation intensities there is a similar bias, with observed data indicating approximately 1% of daily values above 9 mm, while Reanalysis shows 4% of daily precipitation above this level and 1% of daily precipitation above 16 mm.

By working with anomalies, CA effectively removes the biases in Reanalysis mean precipitation and mean temperatures. However, it is evident from the biases in precipitation intensity that, especially in light of our interest in hydrologic extremes, that accounting for mean biases at the large scale is inadequate. We introduce here a third downscaling approach, by combining the initial large-scale bias correction step of BCSD prior to applying the CA method. We refer to this approach as BCCA.

The bias correction employed in BCCA is conceptually identical to that in BCSD, using the same quantile mapping approach. However, rather than applying this to monthly precipitation and temperature, the quantile mapping is used for all daily (precipitation, maximum and minimum temperature) values within each month. For example, all daily precipitation observations (aggregated to the reanalysis spatial scale) for all Januarys in the 27-year "observed" period are lumped into one pool to create a distribution of daily precipitation observations for January of $n = 27 \times 31 = 837$ days. The pool of n daily observed precipitation values is then
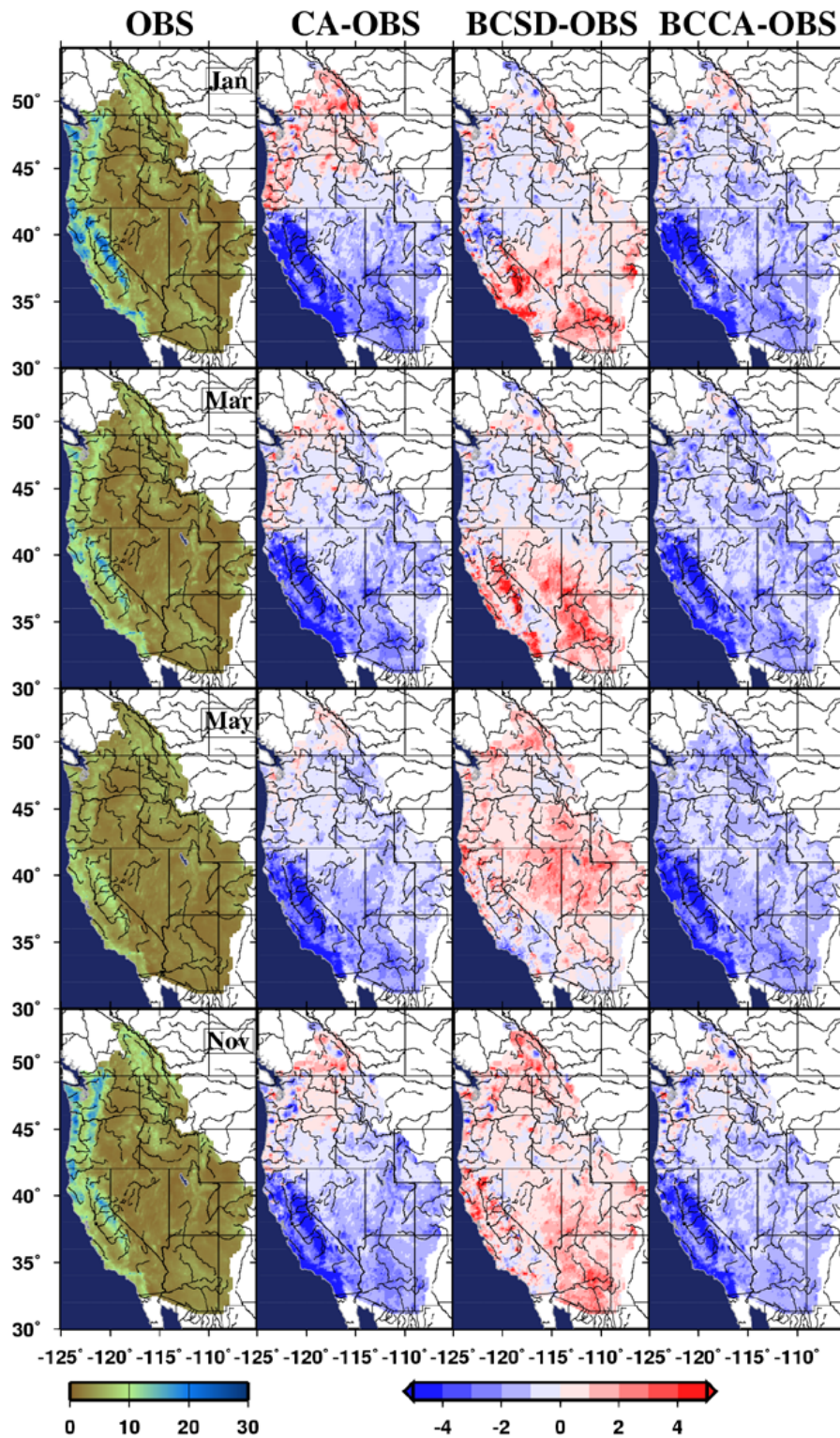
**Fig. 2.** Precipitation intensity in mm/d for four selected months for gridded observations (OBS, left panels), and the difference between downscaled CA and OBS (second column), between BCSD and OBS (third column), and BCCA and OBS (right panels).
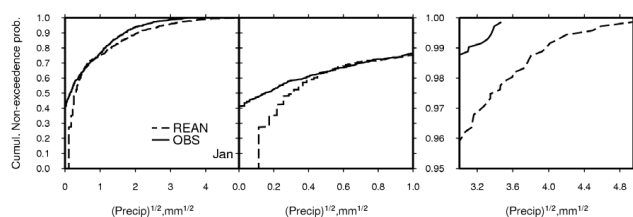
**Fig. 3.** Cumulative distribution function (CDF) for daily precipitation for January at Reanalysis grid point located at 37.1422° N, 110.625° W. Reanalysis is compared to gridded observations aggregated to reanalysis scale. Left panel shows entire CDF, while center panel shows only daily precipitation values below 1 mm, and right panel shows precipitation over 9 mm/d.



**Fig. 4.** CDFs for the same grid cell as in Fig. 3, but based on monthly average precipitation rate data for January for the "observed" period 1950–1976.

sorted, and each day is ranked, as in BCSD, with a quantile of rank/$(n + 1)$, and assembled into a cumulative distribution as in Fig. 3. Daily January precipitation values for the large-scale model output (reanalysis) is similarly arranged into a cumulative distribution. Similar pairs of distributions are prepared for all 12 months. The bias step is completed by using these relationships for each day in the reanalysis time series, where the precipitation value is converted to a quantile using the cumulative distribution for reanalysis, and that quantile is then drawn from the observed cumulative distribution to obtain a new, bias-corrected precipitation value for that day. For example, if reanalysis simulates a very small precipitation amount that is exceeded 90% of the time (for that month), and observations show 30% of the days with no precipitation, the small amount of reanalysis precipitation will be re-mapped to a value of zero. In this way, the bias-corrected daily data will match the observations for the number of rainy days and the average rainfall intensity (for the observed period). Finally, in BCCA since all biases are explicitly corrected, the constructed analogues are then developed on absolute values rather than anomalies, which contrasts with the use of anomalies in the original CA. Since the biases in reanalysis temperatures are much smaller, in a relative sense, than precipitation biases, the discussion below focuses on bias correction of precipitation.

While the bias correction included with BCCA forces the cumulative distribution function to match observations for the historical (observed) period, some biases due to the downscaling methods remain. Figure 2 shows the comparison of BCCA to observations, after downscaling to the 1/8 degree spatial resolution. The high bias in precipitation intensity in the PNW was successfully reduced by the bias correction process in BCCA, showing an improvement over CA. This indicates that large scale bias may have been the primary factor for bias in downscaled precipitation intensity in this area. The bias toward underestimation of precipitation intensity with CA over California is not removed by the bias correction, suggesting that some bias is introduced in the CA method in this region, which was also noted by Hidalgo et
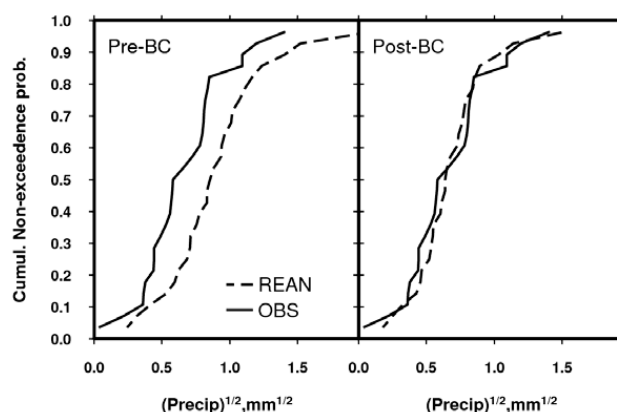
al. (2008). Although the underestimation by CA and BCCA in California is largest during the rainy season, January and March in Fig. 2, the bias, while appearing large, is small relative to the mean observed intensity in these months (in the leftmost column of Fig. 2).

While the daily bias correction ensures that the cumulative distribution of daily precipitation (or maximum and minimum temperature) values will exactly match the observed distribution for all the daily values for any month, it does not explicitly force the monthly distributions to match. In other words, by assembling all January daily values for 1950–1976 for a reanalysis grid cell into a single cumulative distribution function (as in Fig. 3), the bias correction only guarantees that the entire set of daily values (for this example, 837 days) will match the statistics for the set of 837 days for the observations. There is no guarantee that the distribution of monthly precipitation values (for example, 27 January average precipitation values) is also improved. However, as illustrated in Fig. 4, the monthly values are also largely corrected for their biases at all quantiles when the daily values are bias corrected. This indicates that the modeled precipitation variability in reanalysis at the daily scale within a month is consistent with observations, inasmuch as the monthly bias is largely addressed by the daily bias correction.

## 3.2 Impact of downscaling approaches on daily hydrology

Prior to analyzing daily metrics, we assessed the ability of each downscaling method to reproduce annual flow volumes for the projected 1977–1999 period at each gauge site listed in Table 1. Three separate statistical tests were performed: The Mann-Whitney U test (for central tendency, the non-parametric alternative to mean) (Wilks, 2006); the Siegel-Tukey test (for scale, the non-parametric alternative to variance) (McCuen, 2003), and the KS test (noted above

**Table 2.** Statistical test results for BCSD, CA, and BCCA for annual flow volume simulations. A gauge name in bold face indicates that the downscaled streamflow using the indicated technique differs from the observations. Two tests are used: the Mann-Whitney U (also referred to as the Wilcoxon-Mann-Whitney), and the Kolmogorov-Smirnov 2-sample test (both tests performed at $p = 0.05$).

| Mann-Whitney Test | | | Kolmogorov-Smirnov Test | | |
|---|---|---|---|---|---|
| BCSD | CA | BCCA | BCSD | CA | BCCA |
| SHAST | SHAST | SHAST | SHAST | SHAST | SHAST |
| SAC_B | SAC_B | SAC_B | SAC_B | SAC_B | SAC_B |
| OROVI | OROVI | OROVI | OROVI | OROVI | OROVI |
| **NF_AM** | NF_AM | NF_AM | **NF_AM** | NF_AM | NF_AM |
| **FOL_I** | **FOL_I** | FOL_I | **FOL_I** | FOL_I | FOL_I |
| CONSU | **CONSU** | CONSU | CONSU | **CONSU** | CONSU |
| PRD_C | PRD_C | PRD_C | PRD_C | PRD_C | PRD_C |
| DPR_I | **DPR_I** | DPR_I | DPR_I | **DPR_I** | DPR_I |
| **LK_MC** | **LK_MC** | LK_MC | **LK_MC** | LK_MC | LK_MC |
| **MILLE** | **MILLE** | MILLE | MILLE | **MILLE** | MILLE |
| KINGS | **KINGS** | KINGS | KINGS | **KINGS** | KINGS |



**Fig. 5.** Center timing for each streamflow site using each downscaling method, and the observationally-derived streamflows. Day is day of the water year, so 1 corresponds to 1 October.

in Sect. 2, for distribution characteristics, including central tendency, scale, and shape). Results are summarized in Table 2 for the Mann-Whitney U test and the KS test; The Siegel-Tukey test detected no sites with varying scale characteristics at the $p = 0.05$ level, and thus is not shown. For the annual flow volume simulations, the Mann-Whitney and KS test results in Table 2 show similar results at each site, suggesting that differences in the central tendency (as detected by the Mann-Whitney test) are the primary difference in the simulation of downscaled hydrology, as opposed to changes in inter-annual variability (based on the Siegel-Tukey test finding no significant differences between observed and simulated scale/variability at any site). Based on the KS test results, for BCSD, three sites had distributions of annual flow volumes that differed from the annual flow volumes produced by the hydrologic model simulation driven by observations. Similarly, CA differed at four of the stream gauge sites. BCCA, by contrast, produced a distribution of annual flow volumes that were indistinguishable from the observation-driven hydrologic model run, showing substantially improved downscaling skill even for annual measures of performance.

Three daily-scale streamflow metrics are evaluated in this study: center timing (CT), 3-day peak flow, and 7-day low flow. Center timing is defined as in Stewart et al. (2005) as the day on which half of the annual (water year, 1 October–30 September) flow volume has passed a particular point on a stream. For each water year in the verification (or projection) period of 1977–1999 the metrics are calculated and then the results are assembled into distributions for each metric. These distributions are compared among downscaling
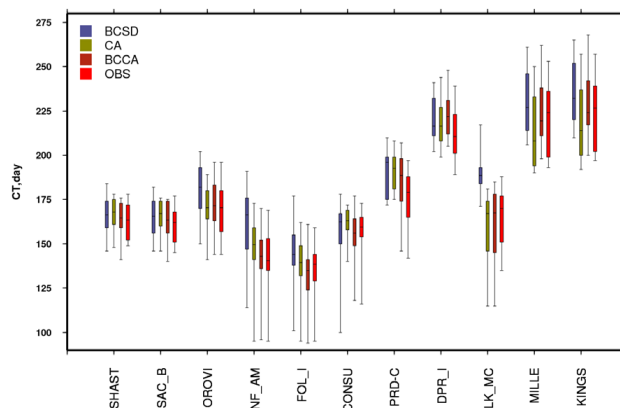
methods and with the simulation using gridded 1/8 degree observations to drive the VIC model. As was found with the annual flow volumes (Table 2), the results of the Mann-Whitney test very closely resemble the results of the KS test, indicating that differences in the distribution of these daily hydrologic metrics are due principally to differences in central tendency rather than interannual variability, and thus KS test results are relied upon as the primary measure of downscaling skill.

Figure 5 shows the performance of the three downscaling techniques along with the observations-based streamflow simulation for the CT statistic. Since CT in snowmelt-dominated basins tends to be driven more by temperature than precipitation, the distribution of CTs simulated by CA are able to capture the skill in daily temperature present in the reanalysis (since temperature observations are assimilated in the reanalysis product, bias is relatively low). CA appears to perform better than BCSD at several locations, for example OROVI, NF_AM, and LK_MC. What this demonstrates is that there is skill in simulating CT at many sites with BCSD, which assumes the distribution of daily values within any month are statistically the same for the observed (or training) period of 1950–1976 as for the later projected period of 1977–1999. The CA method, by contrast, recognizes changes in the occurrence of large-scale climate patterns at the daily scale, and produces downscaled daily values that reflect them, allowing the specific variations within months in each given year to change in the projected period, which results in improved skill at some locations. BCCA does not appear to differ greatly from CA at most locations, suggesting that the relatively low bias in reanalysis temperature causes the bias correction step to have a relatively small effect on this temperature-driven statistic. It should be noted that the small ($0.2\,^\circ$C), but significant domain average temperature difference between the 1950–1976 and 1977–1999 periods is dwarfed by the large projections for later in the 21st century
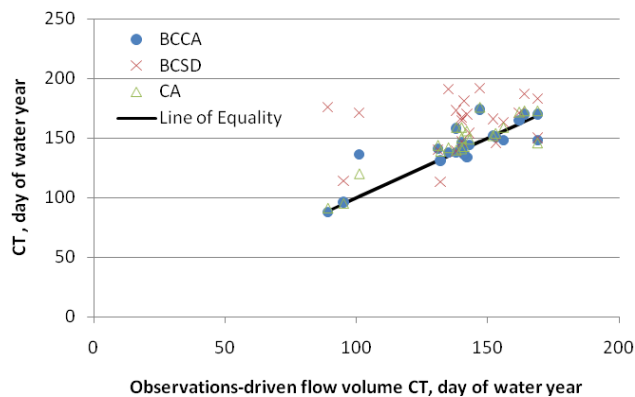
**Fig. 6.** Center timing for the NF_AM site for 1977–1999 using each downscaling method, compared to the CT values for the observationally-derived streamflows. Units are day of the water year, as in Fig. 5.



**Fig. 7.** 3-day peak flow for each streamflow site. Note the vertical axes are different for each of the panels.

for this region (Cayan et al., 2008) of up to 4.5 °C. Thus, as the climate diverges from the historical record to a greater degree, it would be expected that the difference in skill between BCSD and the analogue-based methods (CA and BCCA) could become more stark.

Figure 6 shows the CT values for each water year for the "projected" period of 1977–1999 at the NF_AM site for BCSD, CA, and BCCA relative to the observations-driven simulated CT values. This supports the observation in the prior paragraph, where the temperature-driven daily CT statistic benefits from the large scale daily skill used by the CA and BCCA downscaling methods, while the BCSD method shows considerably less correlation, and suppressed interannual variability, relative to the observations-driven CT values. This suggests that skill in projections of how daily temperature sequences may evolve under changed climatic conditions can be captured by ingesting daily large-scale data into a downscaling technique.

The first three columns of Table 3 summarize the KS test performed to determine whether the 22 simulated CT values using the three downscaling methods can be assumed to be drawn from the same distribution with 95% confidence. This verifies that CA outperforms BCSD, providing a statistically significant improvement at two locations (NF_AM and LK_MC). BCCA is generally as good or better than CA, producing CT values with a distribution statistically indistinguishable from the CT values from the observationally driven hydrologic simulation at all sites.

Figure 7 shows the results for the 3-day peak flow for distribution of values for each of the 22 water years from 1978–1999 at each site. The statistical test results for peak flows are in columns 4–6 in Table 3. In contrast to the CT measure, 3-day peak flow is much more highly driven by precipitation, which is less well represented in reanalysis, and thus would be expected to benefit from the bias correction
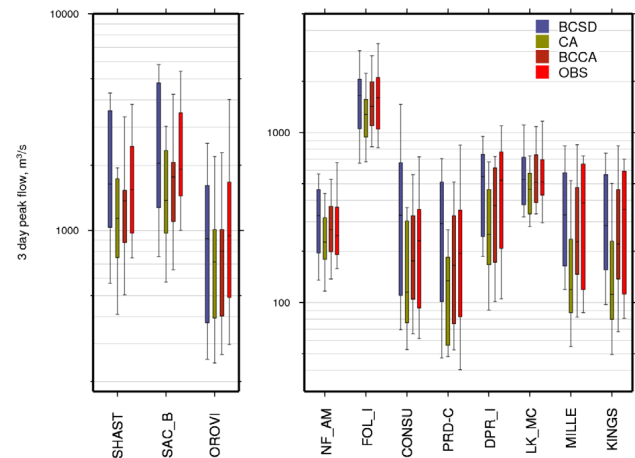
to a greater degree than temperature driven phenomena. Figure 7 shows that CA has a tendency to somewhat underpredict peak flow at most locations. BCCA produces visibly better simulated peak flow values than CA at many sites (e.g., NF_AM, FOL_I, DPR_I, MILLE, KINGS), with BCCA distributions showing a closer match to observations than CA. Surprisingly, Table 3 shows peak flows derived using downscaled meteorology from all three techniques are statistically indistinguishable from those driven by observations at all sites at 95% confidence, so while BCCA appears to be an improvement over CA, the difference is small relative to natural variability. This shows that, for precipitation-driven impacts, the bias-correction step used in BCSD (and BCCA) effectively accommodates the precipitation bias in the large-scale raw forcing data. Also, the use of anomalies in CA, which accounts for biases in the mean at the large scale, appears to work adequately, if not as well as possible, for supporting hydrologic skill of this peak flow statistic.

Figure 8 illustrates the performance of BCCA and BCSD at one site relative to the hydrologic simulation using gridded observations, showing one wet year and one dry year. For the wet year both peak flows and low flows are captured relatively well, compared to the observations-driven simulation, for both BCSD and BCCA. BCCA shows the temporal correspondence to the simulation driven by observations, demonstrating that, even though the large scale reanalysis precipitation is numerical model output rather than assimilated observations and has well-known biases, the bias correction procedure employed here recovers the daily signal present in the observations. BCSD, by design, has no correspondence to the sequencing in the daily observations-driven simulation. However, even with its random generation of daily sequences within any month, BCSD does produce numbers and magnitudes of peak flows that resemble the observations-driven peak flows.

**Table 3.** Statistical test results for BCSD, CA, and BCCA for daily flow measures. A gauge name in bold face indicates that the distribution of 22 values for downscaled streamflow differs from the observed distribution, based on a Kolmogorov-Smirnov 2-sample test (at $p = 0.05$). Hence, non-bold face indicates the downscaling method produces values statistically indistinguishable from observations.

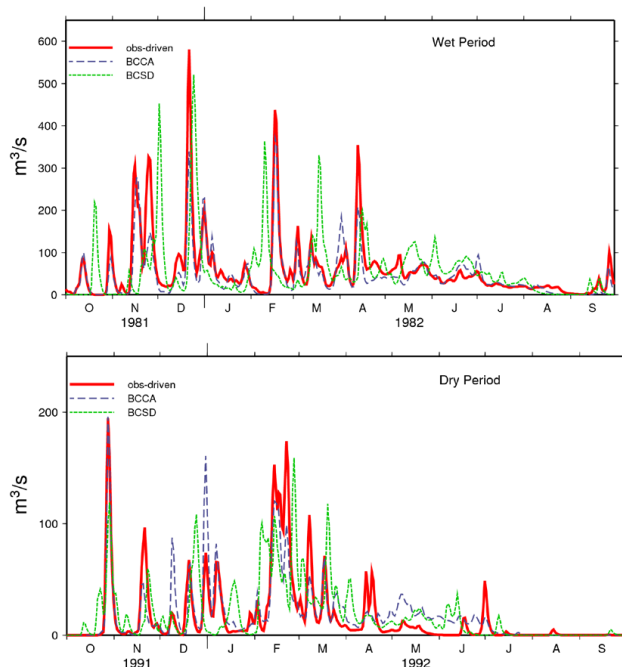| Center Timing | | | 3- Day Peak | | | 7-Day Low Flow | | |
|---|---|---|---|---|---|---|---|---|
| BCSD | CA | BCCA | BCSD | CA | BCCA | BCSD | CA | BCCA |
| SHAST | SHAST | SHAST | SHAST | SHAST | SHAST | SHAST | SHAST | SHAST |
| SAC_B | SAC_B | SAC_B | SAC_B | SAC_B | SAC_B | SAC_B | SAC_B | SAC_B |
| OROVI | OROVI | OROVI | OROVI | OROVI | OROVI | OROVI | OROVI | OROVI |
| **NF_AM** | NF_AM | NF_AM | NF_AM | NF_AM | NF_AM | **NF_AM** | NF_AM | NF_AM |
| FOL_I | FOL_I | FOL_I | FOL_I | FOL_I | FOL_I | **FOL_I** | **FOL_I** | FOL_I |
| CONSU | CONSU | CONSU | CONSU | CONSU | CONSU | CONSU | CONSU | CONSU |
| **PRD_C** | **PRD_C** | PRD_C | PRD_C | PRD_C | PRD_C | PRD_C | PRD_C | PRD_C |
| DPR_I | DPR_I | DPR_I | DPR_I | DPR_I | DPR_I | DPR_I | **DPR_I** | DPR_I |
| **LK_MC** | LK_MC | LK_MC | LK_MC | LK_MC | LK_MC | LK_MC | LK_MC | LK_MC |
| MILLE | MILLE | MILLE | MILLE | MILLE | MILLE | MILLE | **MILLE** | MILLE |
| KINGS | KINGS | KINGS | KINGS | KINGS | KINGS | KINGS | **KINGS** | KINGS |



**Fig. 8.** Simulated streamflow for the NF_AM site (listed in Table 1) using driving meteorology from BCCA and BCSD downscaling methods, and from the hydrologic model simulation driven by gridded observations. A wet water year (top panel) and dry water year (bottom panel) are shown.

The flows during the dry year in Fig. 8 show similar patterns to the wet year. However, one example of the shortcoming of selecting random daily sequences in BCSD is seen in October–November where BCSD shows too many smaller peak flows, whereas BCCA concentrates the flow on one larger peak event, better matching the observations-driven peak flow. The difficulty in matching the very low flows during May–June, at the end of the snow melt season, in the dry year by both downscaling procedures reinforces observations by others that small variations in precipitation can result in larger differences in late season low flows (Vidal and Wade, 2008). The large scale reanalysis signal of precipitation and temperature has been shown to be the most important determinant of uncertainty in simulations of low flows, with downscaling technique secondary (Wilby and Harris, 2006). Specifically related to the current study, Maurer and Hidalgo (2008) found generally lower skill in reproducing observed precipitation statistics with either the BCSD or the CA downscaling technique (applied to the same reanalysis data used in the current study), reflecting the limited daily skill in the large-scale reanalysis precipitation fields. While the bias correction included in each downscaling method can accommodate systematic biases in the large-scale predictor, it cannot produce skill where little exists in the large scale signal. This demonstrates that since the different bias correction and downscaling procedures employed in this study will inevitably still contain some biases at the fine scale, their effect on simulated flows may be especially evident during low flow periods. However, it should be noted that Fig. 8 depicts only one year; Fig. 5 shows that in general the seasonal cycle of accumulation and runoff at the NF_AM site, as expressed by the streamflow timing, is well represented by the downscaled hydrology, both in mean and interannual variability, especially by BCCA.

Simulating 7-day low flows with downscaled meteorology is more problematic, as shown in Fig. 9 and columns 7–9 of Table 3. Several sites exhibit a distribution of low flows that are statistically different for both BCSD and CA downscaling approaches from low flows simulated using gridded observed meteorology. As with peak flows, CA appears to
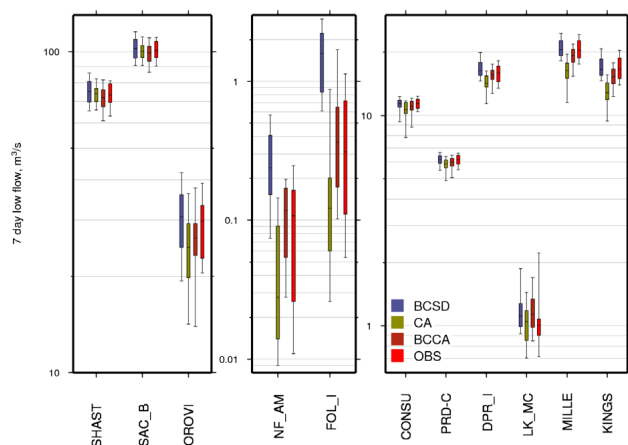
**Fig. 9.** 7-day low flow simulated for each site. Note the y-axes have different scales for each panel.

have a tendency to produce low flows that are lower than observed at many sites. While BCSD produces reasonable values at some sites, low flows are overpredicted in some locations, especially apparent at NF_AM and FOL_I. BCCA, by contrast, appears to produce low flows values that are closer to those produced by the observationally-driven simulation. Table 3 bears these observations out, showing that at two sites BCSD produces low flows different from observations, and at four sites CA produces different values from observations, with high statistical confidence. For the low flow distribution, BCCA is again statistically indistinguishable from observationally-driven hydrology at all sites. It is evident that the choice of downscaling method may influence results more for low flows than for other measures of streamflow. A factor contributing to this may be the relatively greater reanalysis skill (lower biases compared to reanalysis precipitation) for daily temperature, allowing the bias correction to have a greater effect. Since low flows would be affected by evapotranspiration more so than peak flows, a better representation of daily temperatures, more closely resembling observations, would improve skill for the BCCA method.

As a postscript, the improvement seen in applying the bias correction to large-scale daily forcing data begged the question of whether a post-downscaling bias correction, applied using the same quantile mapping approach at the 1/8 degree spatial scale, could provide additional improvement in simulated hydrology. We conducted this experiment using both the BCCA and the BCSD downscaled meteorology, performing quantile mapping bias correction of daily precipitation, and maximum and minimum temperatures, again using 1950–1976 as the "observed" period and 1977–1999 as "projections." We found no consistent improvement in the simulated hydrologic measures used in this paper. This suggested that, since the systematic, large-scale biases had already been removed in both BCSD and BCCA, the remaining fine scale biases during 1950–1976 were not generally

the same as 1977–1999, and the assumptions embedded in the quantile mapping at fine scales were not substantiated in this study.

## 4 Summary and conclusions

We statistically downscaled NCEP/NCAR reanalysis precipitation and temperature over the western US using three different methods and drove a hydrologic model with the resulting sets of downscaled meteorology. The historic record was divided into an "observed" period of 1950–1976 and "projections" from 1977–1999. Streamflow was estimated at 11 sites across California, and these were analyzed to determine the ability to estimate three streamflow statistics important to hydrology: seasonal timing, peak flow, and low flow. One method, BCSD, uses monthly large-scale output, and rescales a historic month to estimate daily variability within each month. A second method, CA, uses daily large-scale output to downscale daily precipitation and temperature to a 1/8 degree grid. A new hybrid, the third method, BCCA, combined the bias correction step of BCSD and the daily downscaling of CA.

We found that daily large scale skill can be effectively downscaled from the large scale to the regional scale to simulate these streamflow statistics. Reanalysis assimilates daily temperature observations, and thus has some large-scale skill for temperature, though reanalysis precipitation is solely model output and is prone to substantial biases. The timing of the annual hydrograph was captured by all downscaling methods at most locations, though the hybrid BCCA method was the only one to perform well at all sites. For downscaling meteorology to generate extreme peak flows (3-day annual peaks), all methods performed well at all sites. The annual flow volume was reproduced with better skill by the hybrid BCCA method than either the BCSD or CA methods, showing that the improvement with the BCCA method is also evident at temporal scales longer than daily.

Low flows were more difficult to capture with the downscaled data. While most of the streamflow sites included in our study had low flows simulated with downscaled data that were statistically indistinguishable from those derived when driving the hydrologic model with observations, BCSD and CA had shortcomings. As with the seasonal flow timing statistic the BCCA method outperformed both the BCSD and the CA methods, statistically matching observationally-driven low flows at all sites.

In summary, to downscale large-scale climate data to generate estimates of extreme hydrologic events, downscaling daily large-scale output can provide measurable improvements in regional hydrologic skill, exceeding that of simply assuming that variability within a month will be similar to historical variability. However, without a bias correction step to correct large-scale biases (which can only be expected to be worse in free-running GCMs than in the data-assimilation

constrained reanalysis model outputs), the skillful signal in the daily data was less likely to be exhibited in the downscaled data and the resulting hydrology. The bias correction step, applied to daily large-scale meteorology prior to downscaling, produced some significant improvements in skill in simulating hydrologic extremes. The biases exhibited at the large scale are in both mean and variability, thus working with anomalies (as in the CA method) is not adequate to compensate for large scale biases, but the quantile mapping approach used in BCCA appears more promising.

# References

Abdulla, F. A., Lettenmaier, D. P., Wood, E. F., and Smith, J. A.: Application of a macroscale hydrologic model to estimate the water balance of the Arkansas-Red River basin, J. Geophys Res., 101, 7449–7459, 1996.

Barnett, T. P., Pierce, D. W., Hidalgo, H. G., Bonfils, C., Santer, B. D., Das, T., Bala, G., Wood, A. W., Nozawa, T., Mirin, A. A., Cayan, D. R., and Dettinger, M. D.: Human-Induced Changes in the Hydrology of the Western United States, Science, 319, 1080–1083, doi:1010.1126/science.1152538, 2008.

Benestad, R. E.: Can We Expect More Extreme Precipitation on the Monthly Time Scale?, J. Climate, 19, 630–637, 2006.

Brekke, L. D., Dettinger, M. D., Maurer, E. P., and Anderson, M.: Significance of model credibility in estimating climate projection distributions for regional hydroclimatological risk assessments, Climatic Change, 89, 371–394, doi:310.1007/s10584-10007-19388-10583, 2008.

Cayan, D., Tyree, M., Dettinger, M., Hidalgo, H., Das, T., Maurer, E., Bromirski, P., Graham, N., and Flick, R.: Climate change scenarios and sea level rise estimates for the 2008 California climate change scenarios assessment, California Energy Commission, California Climate Change Center, Sacramento, CA, 64, 2009.

Cayan, D. R., Maurer, E. P., Dettinger, M. D., Tyree, M., and Hayhoe, K.: Climate change scenarios for the California region, Climatic Change, 87, Suppl. 1, 21–42, doi:10.1007/s10584-10007-19377-10586, 2008.

Cherkauer, K. A., Bowling, L. C., and Lettenmaier, D. P.: Variable infiltration capacity cold land process model updates, Global Plan. Change, 38, 151–159, 2003.

Christensen, J. H., Hewitson, B., Busuioc, A., Chen, A., Gao, X., Held, I., Jones, R., Kolli, R. K., Kwon, W.-T., Laprise, R., Magaña Rueda, V., Mearns, L., Menéndez, C. G., Räisänen, J.,

Rinke, A., Sarr, A., and Whetton, P.: Regional climate projections, in: Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.

Christensen, N. S. and Lettenmaier, D. P.: A multimodel ensemble approach to assessment of climate change impacts on the hydrology and water resources of the Colorado River Basin, Hydrol. Earth Syst. Sci., 11, 1417–1434, doi:10.5194/hess-11-1417-2007, 2007.

Das, T., Hidalgo, H. G., Dettinger, M. D., Cayan, D. R., Pierce, D. W., Bonfils, C., Barnett, T. P., Bala, G., and Mirin, A.: Structure and Detectability of Trends in Hydrological Measures over the Western United States, J. Hydrometeorology, 10, 871–892, doi:10.1175/2009jhm1095.1, 2009.

Dettinger, M. D., Cayan, D. R., Meyer, M., and Jeton, A. E.: Simulated hydrologic responses to climate variations and change in the Merced, Carson, and American River basins, Sierra Nevada, California, 1900–2099, Climatic Change, 62, 283–317, 2004.

Feddersen, H., Navarra, A., and Ward, M. N.: Reduction of model systematic error by statistical correction for dynamical seasonal predictions, J. Climate, 12, 1974–1989, 1999.

Feddersen, H.: Impact of tropical SST variations on the linear predictability of the atmospheric circulation in the Atlantic/European region, Ann. Geophys., 46, 109–124, 2003, http://www.ann-geophys.net/46/109/2003/.

Fowler, H. J., Blenkinsop, S., and Tebaldi, C.: Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, Int. J. Climatol., 27, 1547–1578, doi:1510.1002/joc.1556, 2007.

Fowler, H. J. and Ekström, M.: Multi-model ensemble estimates of climate change impacts on UK seasonal precipitation extremes, Int. J. Climatol., 29, 385–416, 2009.

Girvetz, E. H., Zganjar, C., Raber, G. T., Maurer, E. P., Kareiva, P., and Lawler, J. J.: Applied Climate-Change Analysis: The Climate Wizard Tool, PLoS ONE, 4, e8320, 2009.

Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, J. Geophys. Res., 113, D06104, doi:06110.01029/02007JD008972, 2008.

Hawkins, E. and Sutton, R.: The Potential to Narrow Uncertainty in Regional Climate Predictions, B. Am. Meteorol. Soc., 90, 1095–1107, 2009.

Hayhoe, K., Wake, C. P., Huntington, T. G., Luo, L., Schwartz, M. D., Sheffield, J., Wood, E., Anderson, B., Bradbury, J., Degaetano, A., Troy, T. J., and Wolfe, D.: Past and future changes in climate and hydrological indicators in the US Northeast, Clim. Dyn., 28, 381–407, 2007.

Hidalgo, H. G., Dettinger, M. D., and Cayan, D. R.: Downscaling with constructed analogues: daily precipitation and temperature fields over the United States, California Energy Commission, Public Interest Energy Research Program, Sacramento, CA, 62, 2008.

Iorio, J. P., Duffy, P. B., Govindasamy, B., Thompson, S. L., Khairoutdinov, M., and Randall, D.: Effects of model resolution and subgrid-scale physics on the simulation of precipitation in the continental United States, Clim. Dyn., 23, 243–258,

doi:210.1007/s00382-00004-00440-y, 2004.

IPCC: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007a.

IPCC: Climate change 2007: Impacts, adaptation and vulnerability, working group II contribution to the Intergovernmental Panel on Climate Change Fourth Assessment Report, Intergovernmental Panel on Climate Change, Geneva, Switzerland, 2007b.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., and Reynolds, B.: The NCEP/NCAR 40-year reanalysis project, B. Am. Meteorol. Soc., 77, 437–472, 1996.

Katz, R. W., Parlange, M. B., and Naveau, P.: Statistics of extremes in hydrology, Adv. Water Resour., 25, 1287–1304, 2002.

Kharin, V. V., Zwiers, F. W., Zhang, X., and Hegerl, G. C.: Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations, J. Climate, 20, 1419–1444, 2007.

Kim, J., Kim, T. K., Arritt, R. W., and Miller, N. L.: Impacts of increased atmospheric $CO_2$ on the hydroclimate of the western United States, J. Climate, 15, 1926–1942, 2002.

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., and Fiorino, M.: The NCEP-NCAR 50-year reanalysis: monthly means CD-ROM and documentation, B. Am. Meteorol. Soc., 82, 247–268, doi:210.1175/1520-0477(2001)1082, 2001.

Leung, L. R., Qian, Y., Bian, X. D., Washington, W. M., Han, J. G., and Roads, J. O.: Mid-century ensemble regional climate change scenarios for the western United States, Climatic Change, 62, 75–113, 2004.

Li, H., Sheffield, J., and Wood, E. F.: 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, 2010.

Liang, X., Lettenmaier, D. P., Wood, E., and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, J. Geophys Res., 99, 14415–14428, 1994.

Lohmann, D., Nolte-Holube, R., and Raschke, E.: A large-scale horizontal routing model to be coupled to land surface parameterization schemes, Tellus, 48A, 708–721, 1996.

Mantua, N. J. and Hare, S. R.: The Pacific Decadal Oscillation, J. Oceanogr., 58, 35–44, 2002.

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

Maurer, E. P. and Lettenmaier, D. P.: Predictability of seasonal runoff in the Mississippi River basin, J. Geophys Res., 108, 8607, doi:8610.1029/2002JD002555, 2003.

Maurer, E. P.: Uncertainty in hydrologic impacts of climate change in the Sierra Nevada, California under two emissions scenarios, Climatic Change, 82, 309–325, doi:310.1007/s10584-10006-19180-10589, 2007.

Maurer, E. P., Stewart, I. T., Bonfils, C., Duffy, P. B., and Cayan, D.: Detection, attribution, and sensitivity of trends toward earlier streamflow in the Sierra Nevada, J. Geophys Res., 112, D11118, doi:11110.11029/12006JD008088, 2007.

Maurer, E. P. and Hidalgo, H. G.: Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods, Hydrol. Earth Syst. Sci., 12, 551–563, doi:10.5194/hess-12-551-2008, 2008.

Maurer, E. P., Adam, J. C., and Wood, A. W.: Climate model based consensus on the hydrologic impacts of climate change to the Rio Lempa basin of Central America, Hydrol. Earth Syst. Sci., 13, 183–194, doi:10.5194/hess-13-183-2009, 2009.

McCuen, R. H.: Modeling Hydrologic change: Statistical Methods, Lewis/CRC, Washington, D.C., 433 pp., 2003.

Mearns, L. O., Giorgi, F., McDaniel, L., and Shield, C.: Analysis of daily variability or precipitation in a nested regional climate model: comparison with observations and doubled CO2 results, Global Plan. Change, 10, 55–78, 1995.

Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J. F. B., Stouffer, R. J., and Taylor, K. E.: The WCRP CMIP3 multimodel dataset: A new era in climate change research, B. Am. Meteorol. Soc., 88, 1383–1394, 2007.

Nijssen, B., Lettenmaier, D. P., Liang, X., Wetzel, S. W., and Wood, E.: Streamflow simulation for continental-scale basins, Water Resour. Res., 33, 711–724, 1997.

Oki, T. and Kanae, S.: Global Hydrological Cycles and World Water Resources, Science, 313, 1068–1072, doi:10.1126/science.1128845, 2006.

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

Payne, J. T., Wood, A. W., Hamlet, A. F., Palmer, R. N., and Lettenmaier, D. P.: Mitigating the effects of climate change on the water resources of the Columbia River Basin, Climatic Change, 62, 233–256, 2004.

Pierce, D. W., Barnett, T. P., Santer, B. D., and Gleckler, P. J.: Selecting global climate models for regional climate change studies, Proc. National Academy Sci., 106, 8441–8446, 2009.

Reichler, T. and Kim, J.: How Well Do Coupled Models Simulate Today's Climate?, B. Am. Meteorol. Soc., 89, 303–311, doi:310.1175/BAMS-1189-1173-1303, 2008.

Shen, Y., Oki, T., Utsumi, N., Kanae, S., and Hanasaki, N.: Projection of future world water resources under SRES scenarios: water withdrawal, Hydrolog. Sci. J., 53, 11–33, 2008.

Snyder, M. A. and Sloan, L. C.: Transient future climate over the western United States using a regional climate model, Earth Interactions, 9, Paper No. 11, 11–21, 2005.

Stewart, I. T., Cayan, D. R., and Dettinger, M. D.: Changes toward earlier streamflow timing across western North America, J. Climate, 18, 1136–1155, 2005.

van den Dool, H. M.: Searching for analogues, how long must one wait?, Tellus, 2A, 314–324, 1994.

Vicuna, S., Maurer, E. P., Joyce, B., Dracup, J. A., and Purkey, D.: The sensitivity of California water resources to climate change scenarios, J. Amer. Water Resources Assoc., 43, 482–498, doi:410.1111/j.1752-1688.2007.00038.x, 2007.

Vidal, J.-P. and Wade, S. D.: Multimodel projections of catchment-scale precipitation regime, J. Hydrol., 353, 143–158, 2008.

Widmann, M. and Bretherton, C. S.: Validation of mesoscale pre-

cipitation in the NCEP reanalysis using a new grid-cell precipitation dataset for the Northwestern United States, J. Climate, 13, 1936–1950, 2000.

Widmann, M., Bretherton, C. S., and Salathe, E. P.: Statistical precipitation downscaling over the Northwestern United States using numerically simulated precipitation as a predictor, J. Climate, 16, 799–816, 2003.

Wilby, R. L. and Harris, I.: A framework for assessing uncertainties in climate change impacts: low-flow scenarios for the River Thames, UK, Water Resour. Res., 42, W02419, doi:02410.01029/02005WR004065, 2006.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, 2 Edn., Academic Press, New York, NY, USA, 627 pp., 2006.

Wood, A. W., Maurer, E. P., Kumar, A., and Lettenmaier, D. P.: Long-range experimental hydrologic forecasting for the eastern United States, J. Geophys Res., 107, 4429, doi:4410.1029/2001JD000659, 2002.

Wood, A. W., Leung, L. R., Sridhar, V., and Lettenmaier, D. P.: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, Climatic Change, 62, 189–216, 2004.