



## Supporting Online Material for

### **Human-Induced Changes in the Hydrology of the Western United States**

Tim P. Barnett,\* David W. Pierce, Hugo G. Hidalgo, Celine Bonfils, Benjamin D. Santer,  
Tapash Das, Govindasamy Bala, Andrew W. Wood, Toru Nozawa, Arthur A. Mirin,  
Daniel R. Cayan, Michael D. Dettinger

\*To whom correspondence should be addressed. E-mail: [tbarnett-ul@ucsd.edu](mailto:tbarnett-ul@ucsd.edu)

Published 31 January 2008 on *Science Express*  
DOI: 10.1126/science.1152538

#### **This PDF file includes:**

SOM Text

Figs. S1 to S3

References

## 1. Data

The analysis uses observed snow water equivalent (SWE) at the beginning of April, precipitation (P), temperature, and naturalized river flow. The SWE data are from snow courses across the western U.S., obtained from the U.S. Department of Agriculture National Resources Conservation Service and from the California Department of Water Resources California Data Exchange Center. We selected a fixed subset of ~660 snow course sites that had at least 80% data coverage between 1950 and 1999. To exclude the possibility that our results would be biased by selecting stations that had all the missing values at the beginning of the period, we only included stations that had data in 1950 AND at least 80% coverage between 1950 and 1959 (inclusive), in addition to the overall requirement of 80% coverage between 1950 and 1999.

For temperature (JFM daily minimum, T<sub>min</sub>) and precipitation we used a gridded data set based on National Weather Service co-operative network observations (1). For naturalized river flow in the three major western drainages, we used data from the US geological Survey (2) and the U.S. Bureau of Reclamation. Details regarding all these data sets are given in (3), (4), and (5), respectively.

We regionally averaged the SWE/P and temperature data to nine major mountain chains in the western U.S. The main reason for doing this is to average random weather fluctuations across multiple locations. With ~660 snow courses and 9 regions, there are dozens of courses in each region, which helps minimize the noise. The exact regions chosen are arbitrary, but guided by traditional geographic features such as commonly recognized mountain ranges and state boundaries. The intent was to have the regions be identifiable to the inhabitants, and neither so large as to be irrelevant nor so small as to be of local concern only, since our desire was to illustrate how climate change might actually affect people.

## 2. Univariate D&A

We applied a univariate detection and attribution (D&A) analysis to each of the individual variates – SWE/P, JFM daily T<sub>min</sub>, and river flow center of timing (CT). The separate analyses were necessary to address the unique characteristics of each data type. The D&A work on air temperature is covered in (4), river flow by (5) and snow water equivalent by (3). These research efforts provide the foundation for the multivariable results presented here.

### 3. Climate model data

The anthropogenic runs in PCM and MIROC have significant differences in external forcings. Both PCM and MIROC use anthropogenic forcings of greenhouse gases, ozone, and the direct effect of sulfate aerosols. In addition to these, MIROC includes the indirect effect of sulfate aerosols, direct and indirect effect of carbonaceous aerosols, and land-use change as additional anthropogenic forcings. The indirect effect of aerosols, which give a cooling effect, is the most influential among them, although there is still large uncertainty in quantitative estimates of its radiative forcing. For example, see Fig.2.24 of the IPCC AR4 of WG1 ((6), available from IPCC web site). This results in the fact that the MIROC climate warming signal, while having a fingerprint quite similar to that of PCM and of the observations, is weaker. The physics resulting in this relatively cooler signal are highly uncertain, and observations in fact suggest that the aerosol forcing term is over estimated in the MIROC model.

Estimates of natural internal climate variability were obtained from two long pre-industrial control runs: a 750 year run of PCM and an 850 year run from the CCSM3-FV (7). CCSM3-FV was spun up for 240 years before use. These models were selected for their combination of exhibiting a realistic climate and level of natural variability in our area of interest, and the availability of many centuries of daily data. Additionally, the attraction of CCSM3-FV is a relatively high atmospheric resolution for a millennium-scale run ( $1.25^\circ$  longitude by  $1^\circ$  latitude, vs. T42 spectral resolution, approximately equivalent to  $2.8^\circ$ , for PCM), which we deemed beneficial for a regional D&A study. Both models are fully coupled ocean-atmosphere models run without any flux correction. The PCM was downscaled using the BCSD method while the CCSM3-FV global model was downscaled using the CA method; as noted in the main text, our comparison of the two methods found only modest differences between them. The downscaled physical fields from each global field were used as input to VIC model simulations, which yielded the required hydrological variables. An analysis of the two different simulations of natural variability showed them to be sufficiently similar to justify lumping them for further significance tests of signal to noise (S/N).

In summary, all the global model output, without exception, was downscaled. All the temperature, precipitation, and snow depth model output, without exception, was masked to the locations of the snow courses (as interpolated to the nearest grid box of the downscaled, fine resolution grid:  $1/8 \times 1/8$  degree lat/lon). The river flow model output was taken from the river routing model at the locations corresponding to the actual river gauges we used (The Dalles, Lees Ferry, and so on).

### 4. Fingerprinting and Signal Strength

The anthropogenic fingerprint was defined as the leading EOF of the joint (concatenated) data set consisting of SWE/P, JFM Tmin and river CT. SWE/P and Tmin were each averaged across each of 9 regions as described in the main text. Thus, there were nine temperature, nine SWE/P, and three river CT time series, for a total of 21 time series, each 50 consecutive years long. We variously tried forming

the fingerprint from the ensemble average of the PCM anthropogenic runs only, the MIROC anthropogenic runs only, and the combination of both, but found the results to be insensitive to these choices. Results in the main text were obtained with the fingerprint using the ensemble average of PCM and MIROC combined. The fingerprint was normalized such that the associated principal component was standardized (in contrast to the more common normalization that the leading EOF have length 1); this aids in physical interpretation of the results.

Due to the different units for each variable, we normalized each variate by its own standard deviation before the EOF was computed (8). In order to equally weight the 3 river time series with the 9 SWE/P and Tmin series, we weighted the river series by  $3^{1/2}$ . Each individual variate time series was also normalized by the fraction of area it represented relative to the entire area represented by that variate, so that time series representing larger areas would contribute proportionally more. As a sensitivity test, we also tried setting all the areal weights to 1, and found this did not affect the results (not shown). Note we did not have to use an optimal detection approach (e.g., 9) since the signal was strong enough to be detected without additional filtering.

The signal strength  $S$  was calculated as

$$S = \text{trend}(F(x) \bullet D_w(x, t)) \quad (\text{S1})$$

where  $F(x)$  is the fingerprint and ‘trend’ indicates the slope of the least-squares best fit line.  $D_w(x, t)$  are the 21 regional/variante time series (area weighted as described above) taken from the observations, any individual model run, or in the case of a model’s ensemble averaged statistics, the model ensemble-averaged time series. The time mean of  $D_w(x, t)$  was removed before the projection.

## 5. Significance testing

The significance tests in the lower panel of Figure 4 were conducted using an ensemble approach. Instead of comparing each run against a single distribution describing natural variability, we had sufficient data and realizations to test entire ensembles. The ensemble approach reduces the random “weather noise”, helping to define the signal common to the model ensemble members. For example, MIROC provided 10 realizations of the 21 multivariable time series, from which we first calculated the ensemble-averaged signal strength for the 10 realizations, as described above. We then randomly sampled the control run to obtain ten independent groups of the 21 variates, each 50 consecutive years long, and calculated their ensemble average signal strength. This process was repeated 10,000 times in order to construct the empirical probability distribution function (pdf) of having 10 member ensembles of the control run (i.e., natural variability) having a signal strength equal to or exceeding that of the anthropogenic runs. Ensemble sampling considerably increases the signal to noise ratio of the anthropogenically forced runs, by reducing the effects of natural variability uncorrelated between

different ensemble members. We also projected the single realization of the observations onto the model fingerprint.

The results reported in the main paper were obtained by combining the estimates of natural variability from PCM and CCSM3-FV. We also repeated the significance tests for various permutations of signal and noise (Fig S1). The results show that: 1) the pdf of either of the natural variability projections is well fit by a Gaussian distribution. 2) it makes little difference which combination of signal and noise we use; and 3) combining noise estimates makes little difference to the significance levels. Various combinations of model-estimated natural variability and anthropogenic signal all indicate that natural variability cannot account for the changes seen over the last 50 years in the hydrological cycle of the West.

## 6. Precipitation

In the main paper, we described the negative D&A results that were obtained when we used total precipitation, separately, to explain the observed hydrological changes. We also included precipitation in the multivariable metric and computed the fingerprint with that variable included. The relative contributions to the energy in the multivariate fingerprint were SWE/P at 45.3%, Tmin at 27.1%, river flow CT at 24.3% and total precipitation at 3.3%. The total precipitation alone clearly did not contribute appreciably to the fingerprint, and hence was omitted from the multivariate analysis described in the main text.

## 7. Sensitivity tests

Figure 5 shows that the results of our study are insensitive to the two different estimates of natural variability. Figure 4 shows that the results are also insensitive to choice of downscaling method, although the CA approach gave a somewhat weaker signal than the BCSD approach. Previous work has found only minor differences between these downscaling mechanisms (10). Comparing our PCM/BCSD and PCM/CA results, we also find only minor differences between the two methods. So the overall set of runs is a compromise that allowed us to complete the project in a timely way and yet not be dependent on a single downscaling method. Figure 3 shows that the fingerprints derived from the two models are similar. Figure S2 presents these findings together in a common framework which illustrates that the D&A results are robust to perturbations of the estimated background noise, the model fingerprint used, and method of downscaling.

A critical question has to do with the levels of natural variability used in the significance tests. Are they realistic? We investigate this question by verifying (a) the power spectra of our modeled Colorado River flow and (b) the low frequency variance in downscaled temperature minima and maxima. The power spectra of Colorado River flow are from the downscaled, VIC-modeled 850 year CCSM-FV and 750 year PCM control run, and involve both downscaling methods. We choose the Colorado River because its flow is an integrating mechanism over a large region, and in particular, a region where current climate models predict reductions in runoff as global warming progresses. River flow depends on both precipitation and

temperature, so it is a good proxy for the other metrics we use in this article. Future extensions of this comparison to include other major western US river basins – e.g., the Columbia River and California rivers, as well as to the temperature and snow pack characteristics are beyond the scope of the present effort, but would provide perspective on the robustness of the verification in a broader range of hydro climatic regimes.

The modeled flow power spectra compare well with that obtained from the paleohydrologic tree ring reconstructions of (11). Figure S3 shows the model power spectra are similar to and largely encompassed by the 5-95% confidence limits of the reconstructed river flow. Over the low frequency range of interest here (e.g., less than 0.1 cycles per year), there is little statistically significant difference between the modeled variables and reconstructed observations, although PCM tends to have too much variability at the very lowest frequencies considered ( $<0.004$  cycles/year). The same is true for the low frequency variance in observed Tmin and Tmax in the mountain regions shown in Figure 1. We find that both variables are well matched with their downscaled CCSM3-FV and PCM-based counterparts, further demonstrating that the levels of natural variability used in this study are realistic. We therefore conclude that the estimates of natural variability that we have used in the significance test are realistic.

## References

1. A. F. Hamlet, P. W. Mote, M. P. Clark, D. P. Lettenmaier, *J. Climate* **18**, 4545 (2005).
2. J. R. Slack, J. M. Landwehr, *U.S. Geological Survey Rep.* 92, 193 pp. (1992).
3. D. W. Pierce *et al.*, *J. Climate*, submitted.
4. C. Bonfils *et al.*, *J. Climate*, submitted.
5. H. G. Hidalgo *et al.*, *J. Climate*, submitted.
6. IPCC, Report of Working Group 1, 2007.
7. G. Bala *et al.*, *J. Climate*, in press.
8. T. P. Barnett, M. Schlesinger, *J. Geophys. Res.* **92** (1987).
9. G. Hegerl *et al.*, *J. Climate* **9**, 2281 (1996).
10. E. P. Maurer, H. G. Hidalgo, *Hydrology and Earth System Sciences* **4**, 3413 (2007).
11. D. M. Meko *et al.*, *Geophys. Res. Lett.* **34**, L10705 doi:10.1029/2007GL029988 (2007).

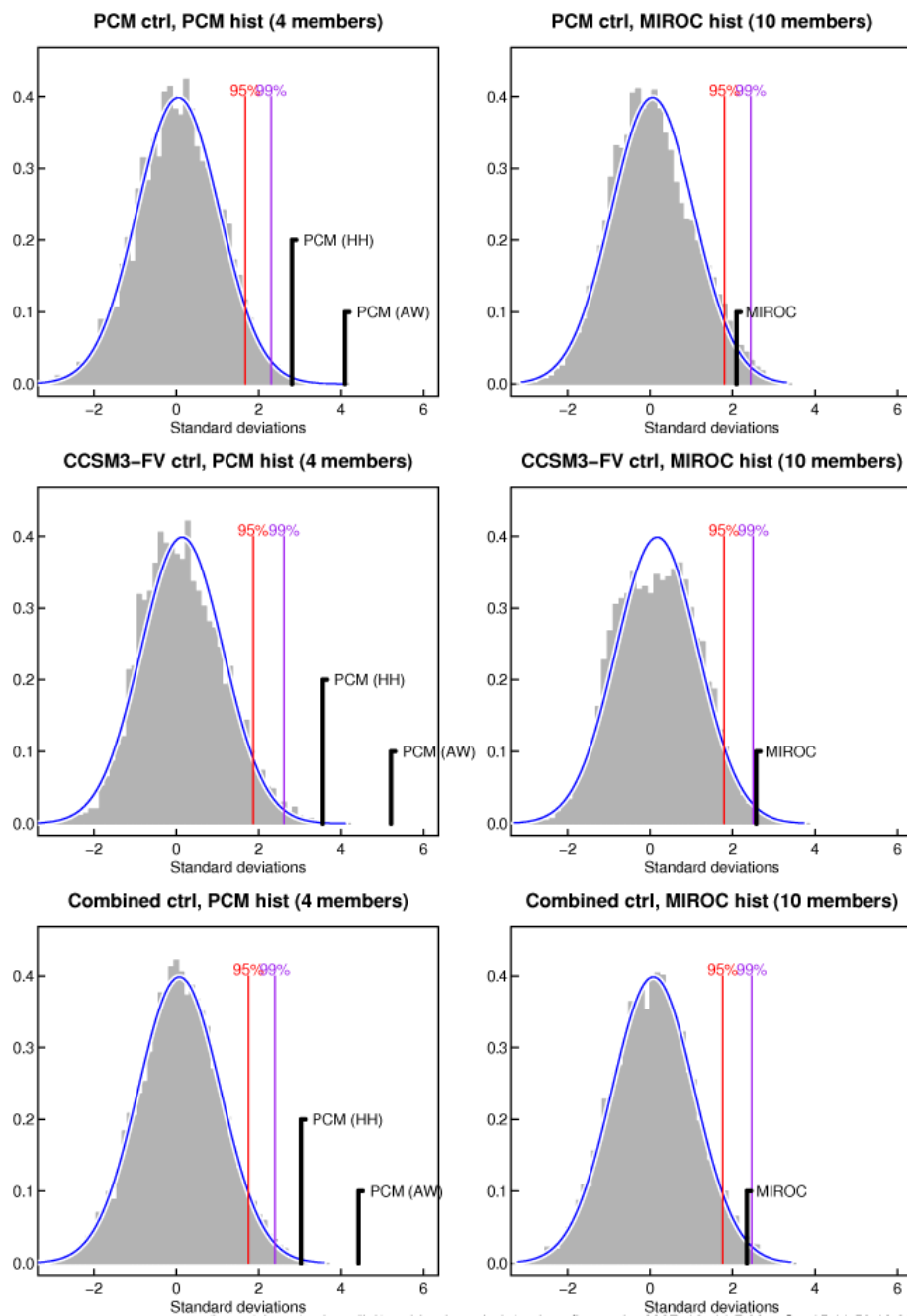


Figure S1. The statistical significance of the signal strength in various anthropogenically forced model runs, taken in combination with the two estimates of natural internal variability from the control runs, as indicated by each panel's title.

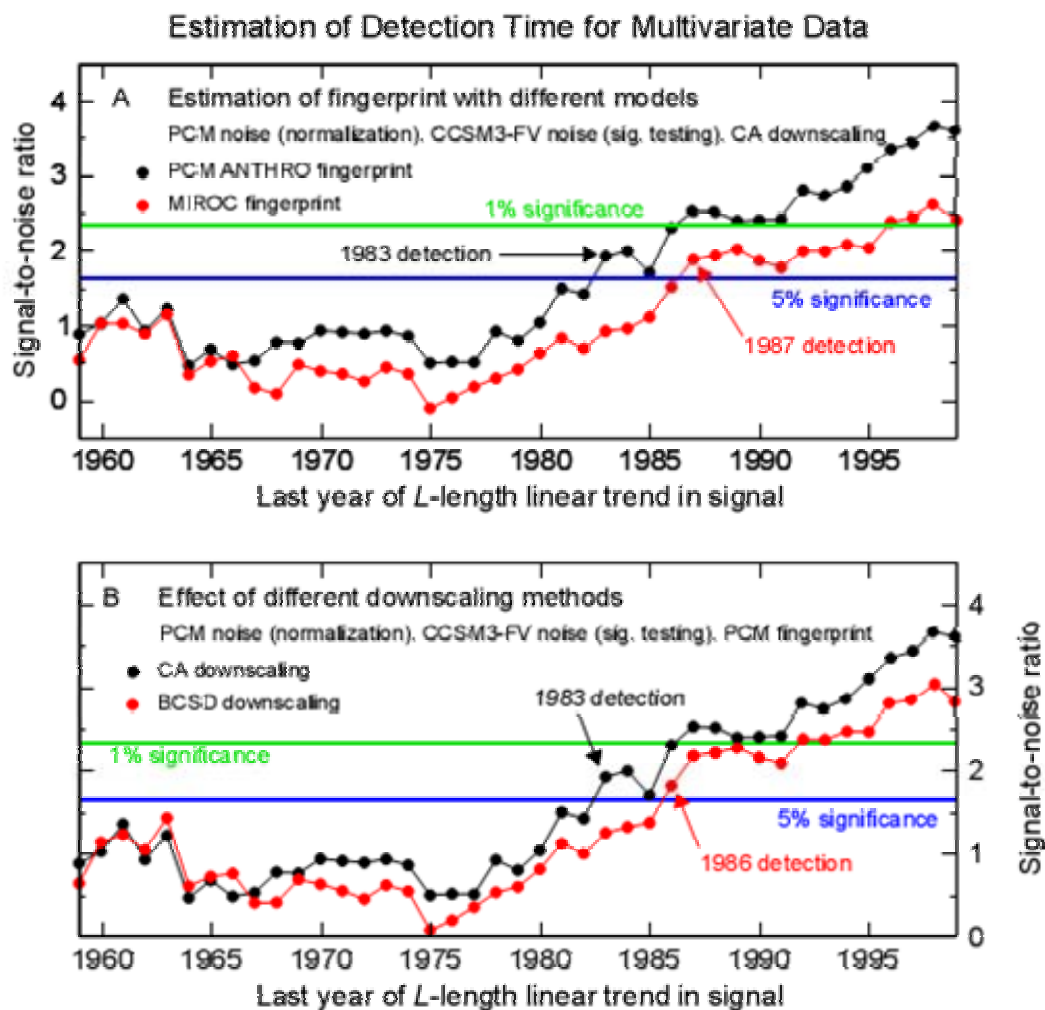


Figure S2. Effect of using different model fingerprints (top) and downscaling methods (bottom) on the detection time.



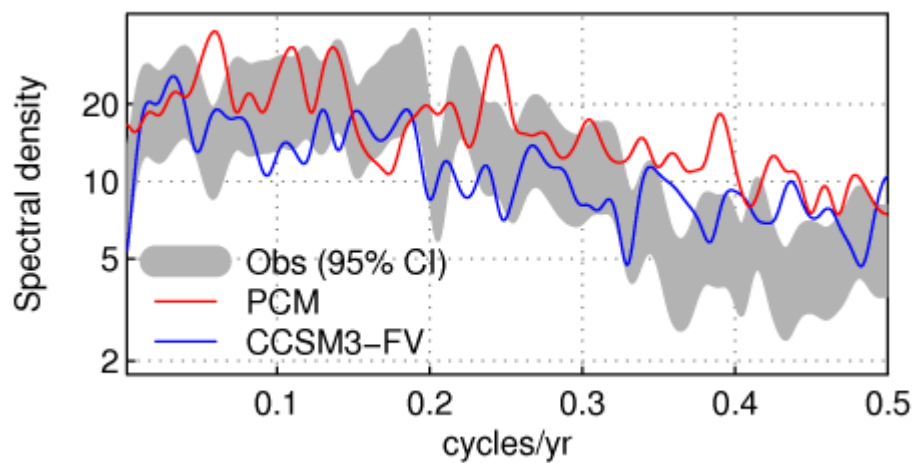


Figure S3. Power spectra of Colorado River flow at Lee's Ferry. Grey: 95% confidence interval on flow reconstructed from proxy indicators over a 1200 yr period (11). Red: spectrum from PCM control run. Blue: spectrum from CCSM3-FV control run.