Hydrol. Earth Syst. Sci. Discuss., 12, C2786–C2788, 2015
www.hydrol-earth-syst-sci-discuss.net/12/C2786/2015/

Hydrology and Earth System Sciences Discussions

Open Access

HESSD

12, C2786–C2788, 2015

Interactive Comment

Interactive comment on "Hydrologic extremes – an intercomparison of multiple gridded statistical downscaling methods" *by* A. T. Werner and A. J. Cannon

Anonymous Referee #1

Received and published: 23 July 2015

Summary: the paper presents the results of an experiment in which a combination of observations, reanalyses, and downscaling methods are combined to simulate climate and hydrologic extremes. A few interesting and useful findings are that choice of reanalysis product, observational dataset, and calibration period for assessing downscaling performance can be important and affect results. There are a few shortcomings and corrections that should be made prior to publication, but generally these are minor.

Comments:

1) p. 6183, line 4, other perhaps better references for SDBC would be Hwang and

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

Graham (HESS, 2013) and Abatzoglou and Brown (J. Climate, 2012).

2) p. 6184, line 14, it might be noted that the presence of internal variability ensures that non-stationarity will always exist between two different time periods in any data set of observed meteorology. There is no way to "ensure against" it.

3) p. 6184, line 18, it is stated that "previous studies have included as many years as possible in the calibration" of downscaling. I do not believe that is the case – there is a balance between including enough years so internal variability does not dominate differences between different periods, but short enough where trends in the data do not introduce erroneous variability.

4) p. 6184, line 29, the text states that BCSD has not been tested with Tmax and Tmin, which is not correct. In its daily version it usually does explicitly include both Tmax and Tmin, by some method or another. See for example Thrasher et al. (HESS, 2012).

5) p. 6189-6190, a summary table (or maybe a graphical flowchart?) of the downscaling methods would be helpful, especially if it showed their relationships. The many acronyms were at times difficult to follow.

6) p. 6191, line 3 and Table 2, the use of different calibration periods for different reanalysis products is problematic, as noted in my comment 2 above and later in the manuscript (p. 6199, p. 6201). Was the motivation simply to include a long period for calibration? If so, that would also cause issues (as noted in my comment 3 above). Maybe adding one additional downscaling demonstration with NCEP1 using just 1979- might help show how important that decision was for the results.

7) p. 6196, line 16, elaborate a little on what the Walker field significance test is and how it is applied.

8) p. 6197, line 20, BCCA is shown to perform better with one observed data set. It also seems that for all other downscaling methods the two observed data sets are fairly consistent.

9) p. 6197, line 27 and elsewhere, performance of different methods is summarized by counting the number of statistical tests that passed some metric. In Table 7, it is obvious that these different metrics are not independent, so adding up the number of passed tests, especially among all metrics, does not seem meaningful. And is the total number of tests the product of the experiments in Figure 3?

10) p. 6200, line 3-5, this sentence is difficult to understand.

11) p. 6201, line 29, ERA products are said to be 'superior' to NCEP1, but in the context of this study it is not clear why this would be important. Since the reanalyses are being used as surrogate GCMs, and even NCEP1 is far superior to any GCM (Reichler and Kim, BAMS, 2008), for testing downscaling performance it would seem the choice of reanalysis might not result in better results (but perhaps different, as this study shows).

12) p. 6202, line 8, the statement that 'We should be able to assume that although the biases in GCMs will be greater than those found in reanalyses they are consistent over time.' is not entirely correct. The cited Maurer et al. 2013 paper shows that these biases are not stationary, though perhaps they are close enough where bias correction still works. But how does that non-stationarity in GCM output compare in magnitude to the difference among reanalyses? the claim in the paper should be more nuanced to reflect this uncertainty.

Minor comments:

p. 6182, line 5, "is greater than" might be better phrased as "includes" – is the point that both uncertainties in future GHG levels and uncertain earth system responses are included? p. 6188, line 1, typo for 'artifacts' p. 6197, line 8, add 'to' between 'due differences' p. 6228, Figure 3, work proceeds from left to right?

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 6179, 2015.