Statistical downscaling of the Community Climate System Model (CCSM) monthly temperature and precipitation projections

White Paper by
Tim Hoar and Doug Nychka (IMAGe/NCAR)
April 2008

Statistical Downscaling

The outputs from the Community Climate System Model (CCSM) may be relatively coarse for applications on regional and local scales. This results from an attempt to balance the computational cost of producing many long simulations and the cost of running at higher resolutions. While useful for their intended purpose, it is also desirable to use this information at a local scale. The spatial resolution of climate change datasets generated from the CCSM runs is approximately 1.4 x 1.4 degrees (Figure 1).

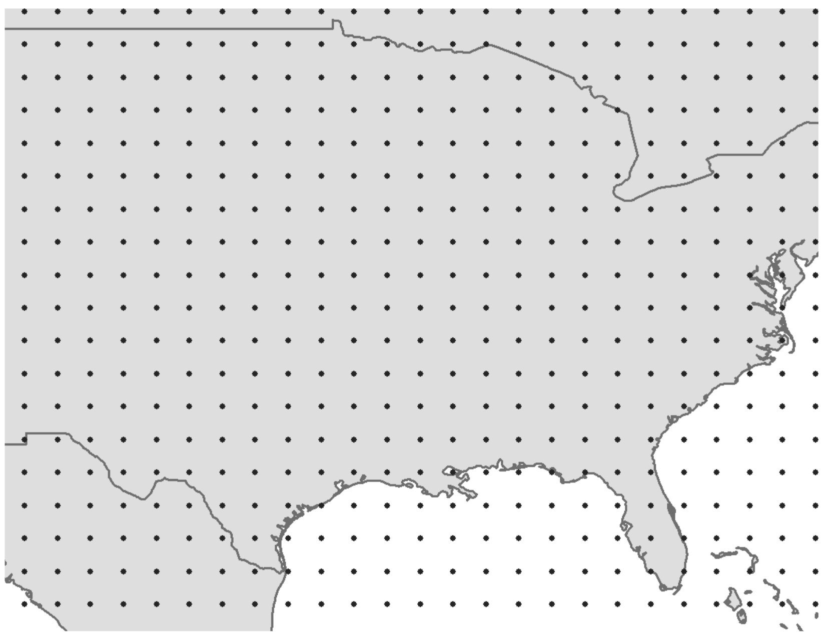

Figure 1: Centroids of CCSM grid cells at approximately 1.4 degree spatial resolution

Downscaling is the general name for a procedure to take information known at large scales to make predictions at local scales.

Statistical downscaling is a two-step process consisting of i) the development of statistical relationships between local climate variables (e.g., surface air temperature and precipitation) and large-scale predictors (e.g., pressure fields), and ii) the application of such relationships to the output of Global Climate Model (GCM) experiments to simulate local climate characteristics in the future.

Statistical downscaling may be used in climate impacts assessment at regional and local scales and when suitable observed data are available to derive the statistical relationships.

A variety of statistical downscaling methods have been developed, ranging from seasonal and monthly to daily and hourly climate and weather simulations on a local scale. The majority of methods have been developed for the US, European and Japanese locations, where long-term observed data are available for model calibration and verification. A description of various statistical downscaling methods and their over-arching assumptions are presented in Wilby et al. (1998). Additional information about downscaling methods can be found in Haylock *et al.* (2006), Fowler *et al*. (2007) and other articles published in the 2007 special issue of *International Journal and Climatology*.

Statistical downscaling of the Community Climate System Model (CCSM) monthly temperature and precipitation projections

The statistical downscaling method used here is developed by Tim Hoar and Doug Nychka at IMAGe/NCAR and involves three steps: 1) determining a simple linear model for every location in the prediction domain, 2) using an off-the-shelf method to provide an initial estimate at every prediction location from the CCSM data, and then 3) applying the linear model to the initial estimate to produce the final downscaled estimate. The following sections discuss steps 2 and 3 as they apply to the CCSM data. The derivation of the linear model (step 1) will be reserved for the last discussion.

The process for downscaling temperature and precipitation are largely similar, so temperature will be used for illustration.

The downscaling process presented here critically depends upon the PRISM (Parameter-elevation Regressions on Independent Slopes Model) climate mapping system, developed by Dr. Christopher Daly at Oregon State University (http://www.ocs.oregonstate.edu/prism). PRISM is a "unique knowledge-based system that uses point measurements of precipitation, temperature, and other climatic factors to produce continuous, digital grid estimates of monthly, yearly, and event-based climatic parameters" (Daly and Neilson 1992, Daly et al. 1994, 1997). Example of PRISM data is shown in Figure 2.

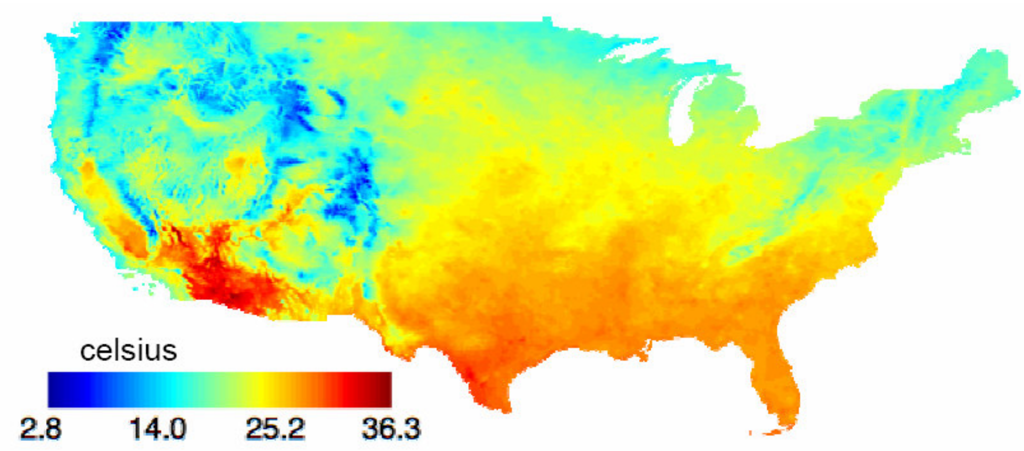

Figure 2: PRISM estimate of average surface air temperature (TAVE) for July 1895

Creating the initial estimates.

The surface of the contiguous US is entirely represented by less than 800 CCSM grid cells as it was configured for the different climate scenario experiments. We have chosen the same prediction grid as the PRISM dataset for reasons that will become clear. The prediction grid has almost 900,000 (1405*621 = 872,505) locations over this same domain. The prediction grid has a spatial resolution of 0.041666 degrees -- approximately 4.5 km, compared to the 1.4 degree resolution of CCSM. The statistical software package "R" (http://www.R-project.org/) is used to fit an interpolating thin plate spline to the CCSM data. The spline model is then used to predict at each of the prediction grid locations. This results in a smoothly-varying reproduction of the CCSM data that does not utilize any additional information about local topography or local climate. It's the same surface, sampled much more often. See Figures 3 and 4.

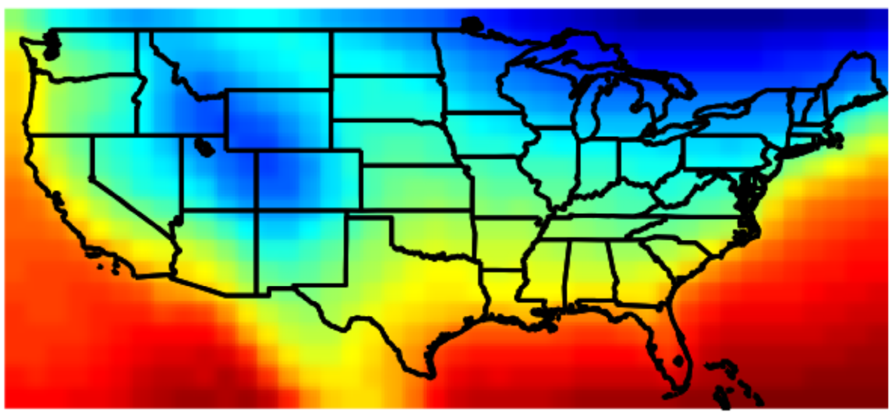

Figure 3: Monthly mean temperature for January, 1896 from the '20C3M' scenario Realization 1. The native grid is a 1.4 degree grid. There are 41x19 grid cells in this domain. This is the data before downscaling.

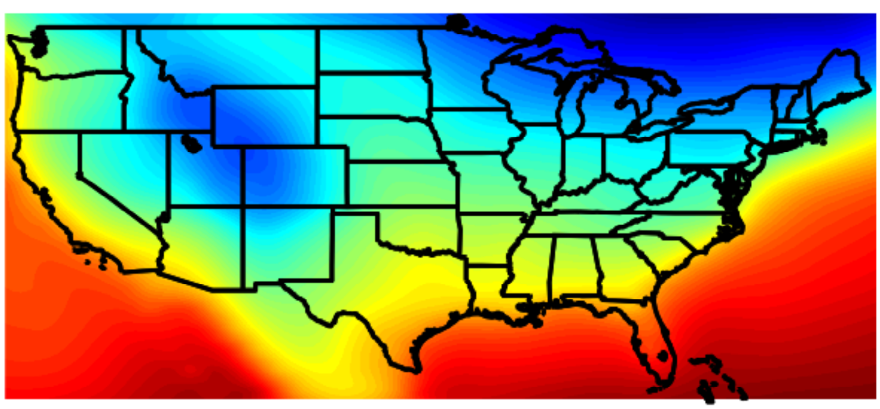

Figure 4: The result of the interpolation by a thin-plate spline to the data depicted in Figure 3. The image varies much more smoothly. This is the initial estimate of the downscaled data at each of the prediction locations.

Local Effects.

In the next step, these predictions are adjusted by applying a linear regression model (slope and intercept) derived from the PRISM data. Each prediction location has a unique model for each month of the year. In other words, there are 872505*12 linear models used to predict mean monthly temperature. Multiply the value of the smooth spline estimate (i.e. Figure 4) by the slope (shown in Figure 5), add an intercept (not shown) and the downscaling is complete (Figure 6).

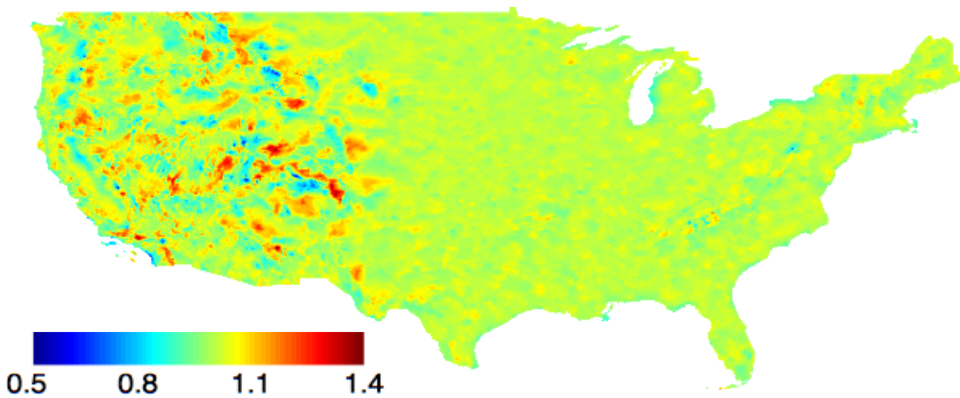

Figure 5: The slope of the linear model for (any) January.

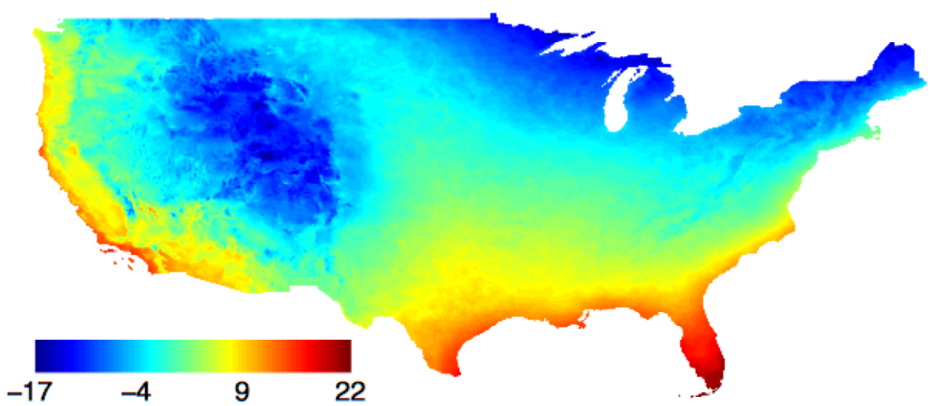

Figure 6: The final downscaled product. This happens to be "Jan. 1896" from the "20C3M" scenario – Realization 1. This cannot and should not compare directly to the real Jan. 1896 – that is not what the CCSM experiment is designed for.

Creating the linear model in the first place.

The PRISM data exist as monthly means for 1895-2005 (at the onset of our downscaling effort) on a 0.041666 degree grid over the Contiguous United States. Ultimately, we are going to use the PRISM data to derive the linear relationship between the smooth estimate and the fully downscaled data.

The smooth estimate is created in the following manner. Each month of the year will have a unique model, so only consider the data for all of the Januaries, for example. The PRISM data are aggregated to the CCSM grid using a simple average. The averaged data is then interpolated using the same thin-plate spline technique to the prediction grid. We now have a consistent set of 'data' (the original PRISM data) and 'observations' (the aggregated-then-splined PRISM data) for 111 Januaries (1895-2005) at every spatial location and can estimate the linear relationship between the smooth estimate and the original. Figure 7 indicates the locations of 12 prediction grid points that will be explored further.

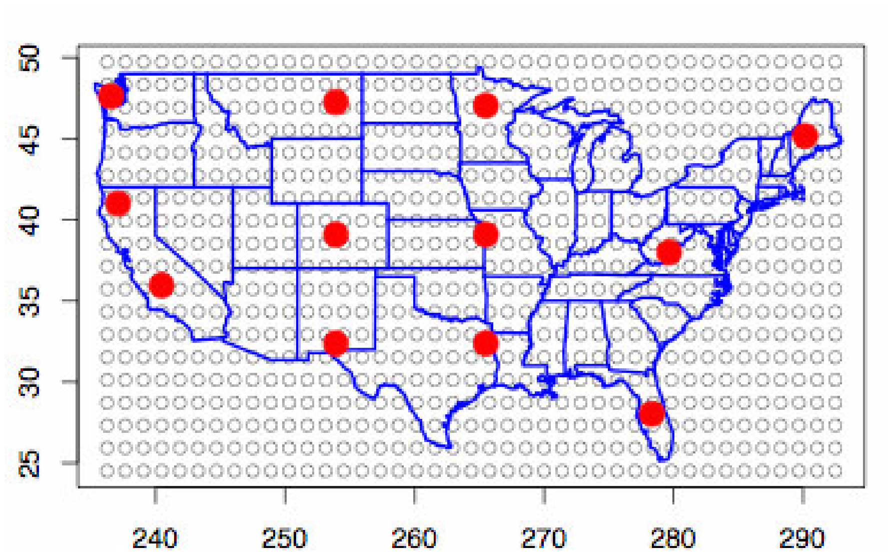

Figure 7: The CCSM grid and 12 select PRISM locations (red dots).

At each of the prediction grid points, all the data that goes into the regression can be plotted on a scatterplot to illustrate the technique. Figure 8 is 12 of the ~900,000 scatterplots arranged in approximate geographical position, each scatterplot representing one location. Florida is in the lower right, Maine is the upper right, Washington State is in the upper left, etc. The locations were chosen solely for their space-filling design. The data are believed to be representative. The strong linear trend in the scatterplots is the rationale behind using a linear model in addition to the smooth estimate. The process is repeated for each of the prediction locations.

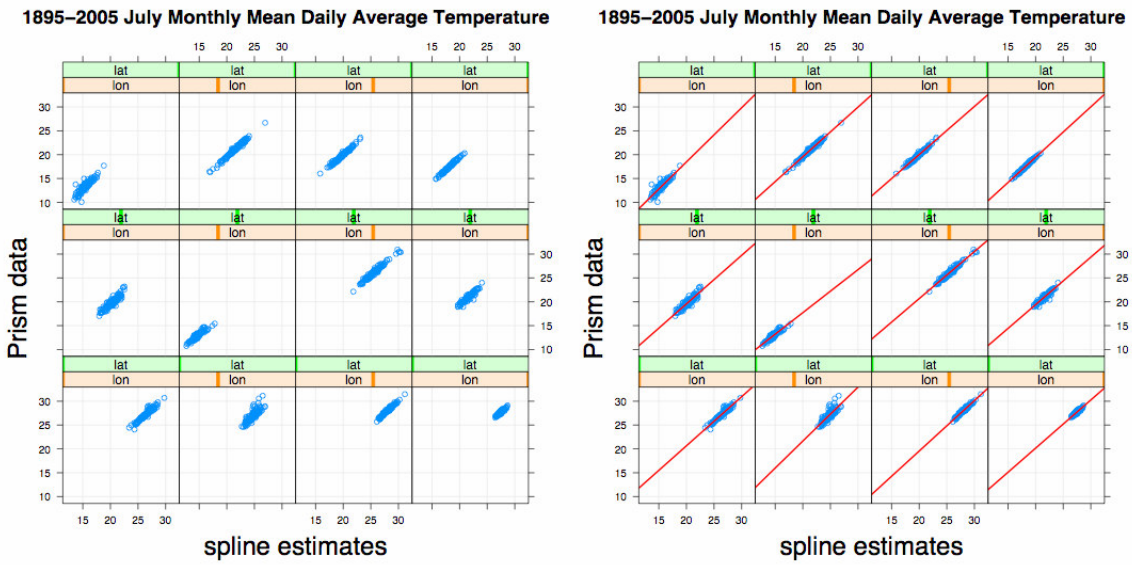

Figure 8: Two panels of scatterplots depicting the linear model phase. The 'aggregated-then-splined' estimates for each July temperature are plotted vs. the actual PRISM value. Note the panels are arranged in approximately the same orientation as the locations on the map. For example, Florida is in the lower right, the Pacific Northwest is in the upper right. The panel on the right has the individual linear models superimposed.

The application of the linear model is illustrated in the following figure (Figure 9) using the July data from the Montana location. The linear model is derived from the PRISM data. The CCSM generates estimates of temperature for the US under some scenario. That data is splined to the prediction grid and, for example, results in a preliminary temperature of exactly 22 degrees Celsius. The linear model adjusts that temperature to the final downscaled estimate of 21.644384 degrees.

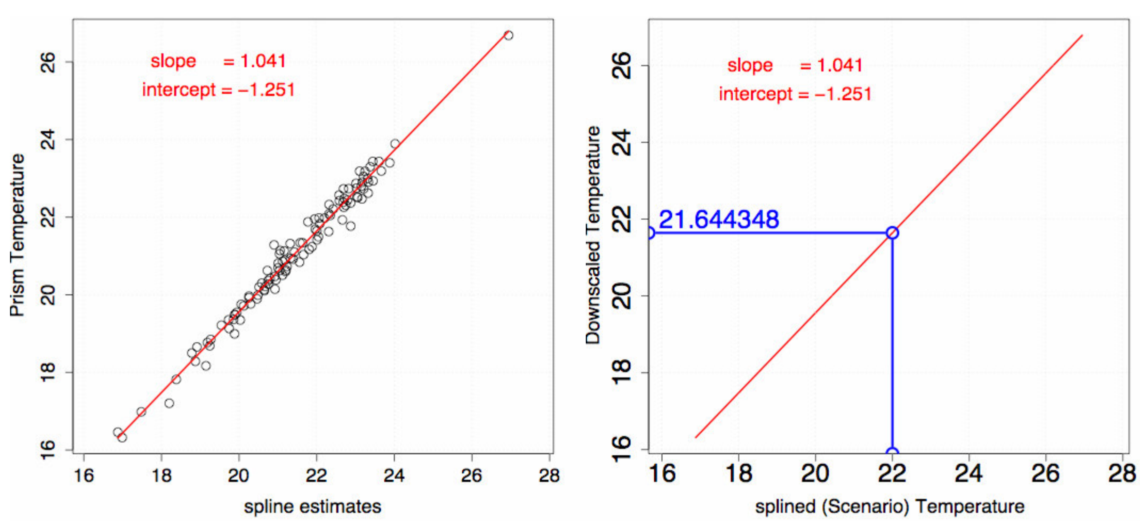

Figure 9: Scatterplot of the July regression data for the Montana location (left) and the application of the model to a hypothetical temperature of 22 degrees.

The left panel of Figure 10 illustrates the slope component of the model that relates a smooth background field to the PRISM grid for (any) January. The right panel of Figure 10 shows the sum of the squared error (sse) – an estimate of the reliability of the model fitting process. Locations whose scatterplots are perfectly diagonal would have an sse of zero.

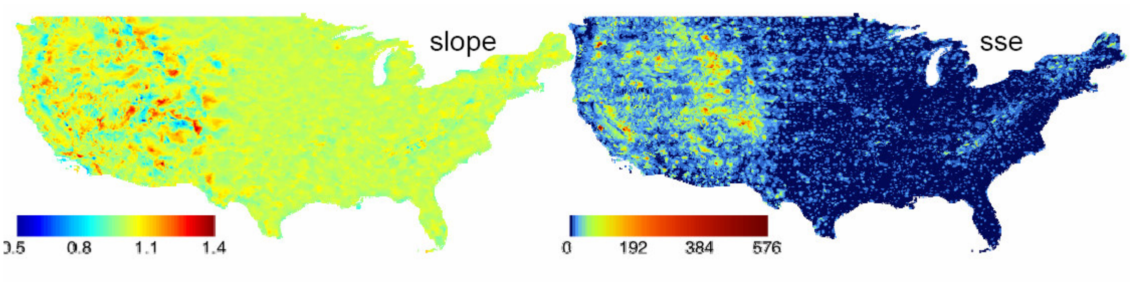

Figure 10: Attributes of the linear model for January. The slope for every prediction location is on the left. The image on the right is the sum of the squared error (sse) in the fitting of the linear model.

Downscaling Monthly Temperature

Temperature is a nicely-behaved monthly mean quantity. All that was necessary to aggregate the data to a CCSM-like product was to average the data from all the PRISM locations within one CCSM grid cell and put that value at the center of mass of the grid cell. The rest of the process is straightforward as outlined above.

Downscaling Monthly Total Precipitation

Precipitation is another matter. It is spatially discontinuous and has a probability distribution function that is non-Gaussian. It is well known that precipitation amounts can be transformed into something more Normally distributed by using a cube-root transformation, which results in more accurate models. Aggregating the precipitation amounts by averaging all the data from all the PRISM locations within one CCSM grid cell resulted in spurious patterns in the regression model coefficients and, ultimately, the predicted fields. The simplest solution was to double the domain of the aggregation step in each direction. The full effect of this choice (on both the temperature and on precipitation downscaling) remains to be explored. One can envision a procedure to find the optimal elliptical-shaped region such that the moments of the aggregated data match the moments of the CCSM data to be predicted.

Nonetheless, the precipitation data was transformed by the cube-root, aggregated to the CCSM-like grid, interpolated using a thin-plate spline (truncating negative values to zero) and the linear model was determined. The CCSM data was converted from precipitation fluxes to precipitation amounts (which matched the PRISM units), similarly splined (while truncating negative values to zero) and the linear model was applied. The cube-root transform was back-transformed, which entailed applying a small bias correction (worked out by Dr. Bo Li) and the final downscaled product was finished.

References:

Daly, C., and R.P. Neilson. 1992. A digital topographic approach to modeling the distribution of precipitation in mountainous terrain. In: Jones, M.E., and Laenen, A. (eds.) *Interdisciplinary Approaches in Hydrology and Hydrogeology*. American Institute of Hydrology. Pp. 437-454

Daly, C., R.P. Neilson and D.L. Phillips. 1994 A statistical-topographich model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology* **33**: 140-158

Daly, C. G.H. Taylor, and W.P. Gibson. 1997 The PRISM approach to mapping precipitation and temperature. In: *Proc., 10th AMS Conf. On Applied Climatology*, Amer. Meteorological Soc., Reno, NV, Oct 20-23, 10-12.

Fowler, H. J., S. Blenkinsop and C. Tebaldi. 2007. Linking climate change modeling to impacts studies: recent advances in downscaling techniques for hydrological modeling. *International Journal and Climatology*, **27**: 1547–1578.

Haylock, M.R., G. C. Cawley, C. Harpham, R. L. Wilby, C. M. Goodess**.** 2006. Downscaling heavy precipitation over the United Kingdom: a comparison of dynamical and statistical methods and their future scenarios, *International Journal of Climatology*, **26**(10): 1397-1415.

Wilby, R.L., Wigley, T.M.L., Conway, D., Jones, P.D., Hewitson, B.C., Main, J. and Wilks, D.S., 1998. “Statistical downscaling of general circulation model output: A comparison of methods.” *Water Resources Research*, **34**: 2995-3008