Subject: Megadrought realizations
From: Ben <bc9z@ldeo.columbia.edu>
Date: 3/28/18, 8:25 AM
To: Elizabeth Wolkovich <e.wolkovich@ubc.ca>
CC: Megan Donahue <donahuem@hawaii.edu>

Hi Lizzie and Megan,

I've finished generating the drought scenarios for the megadrought project. Data and figures are in the linked zip archive:
https://www.dropbox.com/s/q0qhrr22g9ccos1/output.zip?dl=0

Briefly, I'll just take you through my process, but of course I'm happy to answer any other questions you may have.

1) I'm using the North American Drought Atlas (NADA), which is a gridded tree ring reconstruction of PDSI (an index of soil moisture). PDSI is a normalized indicator, so that values of zero indicate "normal" conditions, negative values indicate drought, and positive values indicate wetter than normal conditions. The reference for the particular version of the dataset I am using is here:
https://pdfs.semanticscholar.org/70f2/482df0bcfb49ee386a114caed19a3b355be8.pdf

2) To generate the PDSI time series, I am averaging over California and Nevada (see **fig01_pdsi_map.eps** for what an example year looks like). I then trim this reconstruction to 800-2000 (the most reliable years in the reconstruction) and recenter it to a mean of zero over this time period (**fig02_pdsi_time.svg**). Then, I rescale this time series to a min of 0 and max of 4.5 (**fig03_rescale_time.svg**). I do this by mean shifting everything up by the minimum PDSI value, dividing by the maximum PDSI value (so values are between zero and 1) and then multiplying by 4.5.

3) In those figures, I have highlighted the two intervals (denoted by the horizontal bars) that I will be resampling from. The first is from 801-1200, and this is a period of substantial extended megadrought activity. The alternative is 1601-2000, a much wetter period. The resampling is based on phase randomization. Basically I take the rescaled time series, subtract the mean, conduct a fast Fourier transform, randomize the phase information, and then convert it back from frequency space into a time series and add the original mean back on. Effectively, for each 400 year period, we get a new time series with its own unique temporal evolution, but with frequency characteristics consistent with the original time series. I do this 10,000 times. Because there are no max/min constraints in the resampling, I had to artificially force values below 0 to 0 and above 4.5 to 4.5 on the very rare occasion when things got outside of these bounds.

4) The last figure (**fig04_resample_compare.svg**) compares the resampled distributions against the original time series: shaded brown=all years from all 10000 dry samples, brown line=original dry time series; shaded green=all years from 10000 wet samples, green line=original wet time series. I ran some Kolmogorov-Smirnov goodness of fit tests comparing the distributions, and the resampled vs original are not significantly different. Dry, however, is significantly drier than wet (these are both things we want).

5) Finally, the data files! **data_orig.csv** contains the original time series: year in first column, PDSI in second column, rescaled PDSI to R in 3rd column. **data_dryresamp.csv** contains the dry time series resampling (10,000 rows, each one a unique resampled time series that is 400 years long). **data_wetresamp.csv** is the same, but for the wet time series resampling.

Anyway, hopefully this is enough for you all to get started! Let me know if you have any more questions or if something looks weird.

Cheers,
Ben