Earth Syst. Dynam. Discuss., 3, C372–C376, 2012
www.earth-syst-dynam-discuss.net/3/C372/2012/

Earth System Dynamics Discussions

Interactive comment on "Power-law behavior in millennium climate simulations" *by* S. V. Henriksson et al.

S. V. Henriksson et al.

svante.henriksson@fmi.fi

Received and published: 10 September 2012

We thank Anonymous Referee 3 for his/her review that includes positive feedback and through criticism has allowed for an improvement of the the article, particularly the statistical analyses. Below are our replies to the comments, with the original comments in italics.

Henriksson et al. investigate the power-law behavior of atmospheric and sea-surface temperature variability, focusing on interannual to multidecadal time scales. More specifically, the authors evaluate the power-law behavior of multiple GCM simulations, which incorporate "internal variability" and a range of different external forcings. The results from the different simulations are then compared with power-law estimates derived from instrumental data. In principle, this approach provides the opportunity to bet-

ter constrain the key factors responsible for observed stochastic weather/climate variability and their regional heterogeneity, and it has the potential to be a powerful tool. The spatial ("gridpointwise") evaluation is a particularly interesting application. However, there are a number of substantial problems in the presentation and documentation that must be addressed before this work can be further evaluated for publication.

This review specifically focuses on issues related to the statistical analyses, central to the interpretation of the presented results.

1. In the power spectrum plots, please include confidence intervals for the power estimates. This information will aid in evaluating the statistical significance of the spectral results. Related to this issue, when averaging over all possible windows for the final power spectrum estimate (as indicated on Pg. 395, line 15), it is important to recognize that one is averaging over many statistically dependent results (see Harris, 1978 for details on "overlap correlation" for specific types of windows). The authors should indicate what corrections (e.g., for confidence intervals) are used to account for the statistical interdependence of the calculated spectra. Also, please show the spectral bandwidth resolution on the power spectrum plots.

Confidence intervals have been produced and shown in Figures 2 and 3 and discussed in Section 3. They are produced by creating alternative temperature time series with added white noise (with a constant probability distribution) and choosing the 2.5th and 97.5th percentiles of the spectral estimates as limits for the confidence interval. In this method, the confidence interval is obtained directly from the ensemble without having to consider the overlap correlation. We have at this moment not added confidence intervals in all the figures, as we feel it would compromise visual clarity. The resolution is now mentioned in the captions of the figures.

2. Please use consistent (and conventional; e.g., cycles/year) units for the frequency axis. The current presentation makes evaluation of the results unduly difficult.

The units on the frequency axis have been changed to cycles/year.

3. Please indicate the methodology used to generate the 95% confidence intervals for the "best fit" beta results.

They were generated assuming uncorrelated, Gaussian residuals in the linear regressions, as has been done in many previous studies. A sentence stating this has been added to the first paragraph of Section 4. Figure 6 was plotted to motivate this approximation as it illustrates apparently non-autocorrelated and Gaussian residuals.

4. At present it is not possible to evaluate the statistical significance of the "gridpointwise" best fit beta results presented in Figures 8 and 9. Please incorporate a statistical significance evaluation in the figures.

To obtain significance intervals, the residual would have to be "whitened" first by using a suitable model for removing the autocorrelation, which for a spectrum that does not allow as good a linear fit as that those in Figs. 3-5 is challenging, let alone for 48 times 96 spectra. However, we produced indicative error bars by assuming the global mean temperature spectrum (that has been analysed in response to one of the comments below) to have an AR(1) shaped residual in logarithmic space, which might be used as a rough indicator for the gridpointwise distribution's pointwise error bars too.

Finally, note that the spectral methodology used in this study was introduced by Welch (1967), and it is known as "Welch's Overlapping Segment Analysis" or "Welch overlapped section averaging" (WOSA). At present, no attribution is made to Welch for the method, and the language sometimes gives the impression that it is new (see reference to Henriksson et al., 2012 on line 11; also see line 17, "we also compare our method to an earlier result").

We were not aware of Welch's work using the method in another field of science, but as such a straightforward application of the discrete Fourier transform, we expected that the same method had been used in some context also before. An attribution to Welch has now been made, and the reference is added to the list of references.

Pg. 395 line 4: Here and at other locations in the manuscript, the term "amplitude" is incorrectly used in the place of "power".

Thank you for pointing this out. Now corrected.

Pg. 396 line 26 – pg. 397 line 6: As presented, this passage is too vague. Please be more specific about the analyses that have been done, or remove this passage altogether. Also, it is relevant to note that Pelletier (1998, 2002) has presented a theoretical explanation for the power-law shape of the atmospheric temperature spectrum.

The paragraph has been heavily revised. Pelletier's analytical work and work by Kantelhardt et al. is now cited and discussed and our analyses are only suggested as future work, as the results are still unclear.

Pg. 397 line 22: The "error term" is more commonly referred to as "residuals".

"Error term" has been changed to "residuals".

Pg. 399 lines 6-7: The "gridpointwise" analysis deviates from the frequency range approach used in the global mean temperature assessment. To give a sense of the impact of the new approach, please discuss the results obtained when analyzing mean temperature in the same manner.

This kind of calculations have now been added, not only for global mean temperature but also for continents and oceans separately for further illustration. The results are in discussed in what is now the 4th paragraph of Section 5.

Pg. 400 lines 17-18: Here the authors note, "This raises confidence in our method as compared to the multitaper method". Theoretical studies have conclusively demonstrated the overall advantages of the multitaper method for power spectrum estimation (e.g., see Percival and Walden, 1993). The agreement of the WOSA and multitaper method results for one empirical example does not guarantee good agreement for all other analyses conducted in this study. I encourage the authors to check their other key results using the multitaper method.

The results presented in Section 3 were checked using the multitaper method and are now briefly quoted in Section 4. Using a 64-year time window, the multitaper method gave beta estimates of 0.33$\pm$0.05 , 8.34$\pm$0.02 for the unforced simulation, and 0.77$\pm$0.05 , 8.22$\pm$0.03 for the forced simulation and the two considered frequency ranges, respectively. Using a 300-year time window, the corresponding betas were 0.39$\pm$0.03 and 7.84$\pm$0.01 for the unforced simulation and 0.70$\pm$0.03 and 8.66$\pm$0.01 for the forced simulation. Thus, indeed the power spectral density estimates converge to be nearer the power-law fit, but on the other hand the best-fit betas for time windows of different lengths are no longer inside each other's 95% significance intervals when considering the multidecadal to El Nino frequency range.

Figures 8-9: Please use a consistent color scale for all plots. Also, why not present the results for the unforced simulations as well?

Color scales have been made consistent, wherever figures are similar. We have tried to limit the amount of results presented to allow for efficient communication of our main points without confusion from a too large volume of material. Unfortunately some potentially interesting results are therefore left out, but we have tried to balance the fact that we show more results from the forced simulation for instance by showing the results for the monthly mean time series in Figure 4.

References Cited: Harris, F.J., 1978, Proc. IEEE, v. 66, p. 51-83.; Pelletier, J.D., 1998, EPSL, v. 158, p. 157-164.; Pelletier, J.D., 2002, PNAS, v. 99, p. 2546-2553.; Percival,D.B., and Walden, A.T., 1993, Cambridge University Press, 583 pp.; Welch, P.D., 1967, IEEE Trans. on Audio Electroacoustics, v. AU-15, p. 70–73.

Interactive comment on Earth Syst. Dynam. Discuss., 3, 391, 2012.