#001045

REVIEWER #1

Climate Change, Hurst Phenomenon, and hydrologic statistics

Demetris Koutsoyiannis

Summary:

The author develops a new way (SSS) to estimating Hurst phenomenon and compares this to the more traditional approaches. The author then applies this to synthetic data sets and to three real life data sets and finds differences between the two approaches in estimating the usual statistics (mean, auto correlation, probability of exceedance etc..). The author argues that this is a better approach to modeling long-memory process.

The author has done a good job of providing a fairly comprehensive introduction/background. The motivation is clear and so are the development of relationships for the various statistics in the SSS context, but the results presented are only suggestive and requires quite a bit more to make them convincing and conclusive and useful.

Comments:

The application of Hurst phenomenon to long memory processes in hydrology (e.g. floods) and in climatology (e.g. global temperatures) has been tried out by several researchers in the past. However, the author proposes a different and somewhat better approach to estimating the various statistics. The author shows that the Hurst coefficient is greater than 0.5 in the three time series that is analyses - indicating a long memory process (largely due to trends in the data). The following questions come up naturally:

i. The question then is what can be done with this exponent? What if the process is not simple scaling? how does one test this? On a log-log plot one will always get a straight line and hence an estimate of H. Furthermore, how does the H vary over different sub-lengths of the time series considered in the paper?

ii. All along the underlying assumptions are that data is Gaussian - what if they are not? how is this treated?. Gaussian assumption needs to be tested on the data sets considered.

iii. The author repeatedly mentions that linear time series models (ARMA type models) are not suitable for long memory process, yet there is no attempt to fit an appropriate ARMA model to the time series (or at least one of them) and make synthetic simulations from this model and compare the reproduction of the statistics.

iv. Since the long memory aspect is more apparent in the spectrum of the data and also in other statistics like runs or statistics of crossing thresholds.

Recommendation:

The author needs to address the questions and issues raised in this review and re-submit the manuscript. As it is, the results of the paper are not substantial - but addressing the comments in this reviews by clear comparisons to linear models will strengthen the paper a lot and it will be a useful contribution.

Specific Comments:

review #1
#001045

1. In Figure 4 the author argues that the classical methods under simulate the aggregate variance. Given the small deviation between the author's proposed method and the classic it is difficult to argue for a bit difference. Furthermore, it has to be recognized that the estimates from SSS and the traditional method have variability. So if the variability (i.e. confidence intervals) were plotted for each of SSS estimates I am sure the other estimates will fall within these intervals, thus, making them indistinguishable. It needs to be shown that the differences are significant.

2. From Figure 5, the author suggests that the original method of estimating Hurst coefficient (using rescale range) results in higher variability and also has estimates of H below 0.5 and above 1.0 in some samples, while the author's proposed method does not. It is not clear from equation (22) that this will be case always or it was just that in the set of samples that the author had from Monte Carlo that this did not happen. Perhaps, a discussion on why one cannot expect H to be outside of 0.5 and 1 from equation (22) will be very helpful. Also, how sensitive is solving equation (22) to initial value of H?

3. From Figure 7, the author argues that SSS is better able to reproduce the quantiles as it has a wider confidence interval and it encompasses the true distribution. This is not fair, and perhaps not correct, as the figure is based on ONE sample. A correct way would be to compute the distribution from all the Monte-carlo samples and then get a 5% - 95% confidence from them for the traditional and SSS method and then see where the true distribution falls. With one sample (as is the case in this figure) it is hard to see if it is an artifact of sampling or the difference is indeed significant.

4. Figure 9 shows the autocorrelation for several lags from SSS and the classical approach along with the theoretical values (from Equation 6). The author argues, from this figure that the SSS method better follows the theoretical curve. The variability in the estimation of H is not taken into account in this curve. For Example, if I change the H to 0.75 the theoretical curve encompasses that from the classical method to a large extent. Confidence interval from the Monte Carlo samples have to be presented along with the empirical curves for SSS and then see how far outside the classical estimates fall. Clearly, the variability in the estimate of H is not accounted for in this comparison. Similar observation for Figures 11 and 14.

5. The Hurst estimate from the two methods in Figure 10 indicates that they are very similar. Here too putting this in the context of the variance of the estimate will be very helpful and will make it more clear.

6. The variability in the estimate of H need to be incorporated in estimating the trends and the jump (in Figure 15). Here too the conclusion is rather premature and comparison rather unfair to earlier works. The trend statistic should be computed for several simulations that incorporates the error estimate of H and the look at the 5 - 95% of the trend statistic and see how the observed compares with this ensemble. Then a firm conclusion can be made.

REVIEWER #2

Review of "Climate change, Hurst phenomenon, and hydrologic statistics" (Koutsoyiannis) WR001045

The principal contribution of this manuscript is a series of statistical formulas for characterizing means, variances, etc. from time series that exhibit self-similar scaling (as characterized by the Hurst phenomenon). Classical estimators of variances and confidence intervals will greatly under-estimate the variability of a time series exhibiting self-similar scaling, because successive residuals in such a time series are not independent of one another. Here, the author derives a series of alternative formulas for self-similar time series, and shows that these yield much better agreement with self-similar Monte Carlo data sets.

The fact that self-similar hydrologic time series exhibit serially correlated residuals, and thus that classical estimators are inappropriate, is something that has been generally recognized for quite some time. Thus the original contribution here is in the formulas that provide unbiased estimators for SSS time series. I am not sure how much overlap there is between the results presented here and those in the book by Beran (which is continually missing from our library, so I have been unable to check). From the manuscript it appears that the results up through at least equation (12) are also found in the Beran book, but I have been unable to make a detailed comparison. However, giving the author the benefit of the doubt, I would support publication of the technical core of the manuscript -- the author's estimators and their comparison with the classical estimators.

My major concerns with the manuscript are as follows:

1. The SSS estimators are based on the assumption that the time series is known to exhibit SSS scaling, and that the Hurst exponent H is known. The analysis presented here shows that the estimators are generally unbiased if this is the case. However, in real-world cases the Hurst exponent must be estimated from the data, and the manuscript's own simulations (for example, figure 6) demonstrate that it is difficult to do this reliably. For example, Monte Carlo data sets with a known Hurst exponent of H=0.5 yield estimates of H that range from 0.2 to 0.8 or more, and Monte Carlo data sets with H=0.8 yield estimates of H that range from 0.6 to nearly 1.0. Across this range of H values, the SSS estimators vary enormously, calling their practical utility into question.

2. A more fundamental problem arises if the time series exhibit some other form of scaling than SSS (such as drift, or a step change, or long-term cyclicity, or some kind of ARMA dependence... the possibilities are endless). In many of these cases, the standard deviation will appear be a power-law function of timescale, leading one to infer that the time series exhibits SSS even if it does not. What are the consequences of applying this approach to time series that are not strictly SSS? On page 14, the manuscript says that SSS "is much more effective in representing hydrometerological series than, for instance, the ARMA process." This assertion needs to be substantiated. On p. 11, the manuscript says, "conclusively, a stochastic representation of hydrometerological time series that is consistent with the varying climate hypothesis must be also consistent with the Hurst phenomenon." This is only true if the climate variations obey SSS, which may not be the case. The key question is whether _non-SSS_ fluctuations will nevertheless exhibit power-law scaling of standard deviations, over ranges similar to those examined here. If so, then the tests used here would lead to non-SSS time series being interpreted as SSS.

review#2
#001045

3. I am deeply concerned about the line of argument developed in section 4 of the manuscript. The general approach is encapsulated on p. 9: "The apparent falling and rising trends in all our examples and other time series can be considered as climate changes or variations. In all cases these changes are irregular and, in the absence of an accurate deterministic model that could explain and predict them, are better models as stochastic fluctuations on many timescales." In other words, any fluctuations are _assumed_ to be stochastic (i.e. noise) rather than a real climate signal. By assuming that any self-similar fluctuation is random noise, rather than a physically meaningful climatological signal, this approach leads to excessively conservative hypothesis tests, and thus potentially to large Type II errors. A statistical treatment of a time series is only a description, not an explanation. In the approach pursued here, it is assumed that the explanation is stochastic rather than causal, but there is no rigorous basis for that assumption. On p. 31, the manuscript argues, "... several patterns within these time series would be regarded as evident trends or shifts if classic statistical tests were used, but using modified tests, based on the scaling hypothesis, it turns out that they are regular behavior of the time series, provided that these time series are consistent with the scaling hypothesis." Why does this make those trends any less real, or any less worthy of mechanistic explanation?

4. The introduction through p. 6 does not strengthen the paper, but instead leads the reader off on tangents.

In summary, the issues identified in this review are similar to those outlined by the associate editor concerning the previous version of the manuscript. The manuscript would be improved by:

1) taking a more direct approach to motivate the paper's technical contributions

2) being more careful in interpreting the implications of the analysis, keeping in mind the necessary distinction between a statistical description and a physical explanation

3) being more careful to consider other possible sources of long-term fluctuations, including nonlinear trends, mixtures of fluctuations on different scales, and ARMA processes.