Earth Syst. Dynam. Discuss., 6, C596–C609, 2015
www.earth-syst-dynam-discuss.net/6/C596/2015/

Interactive comment on "Are there multiple scaling regimes in Holocene temperature records?" *by* T. Nilsen et al.

T. Nilsen et al.

tine.nilsen@uit.no

Received and published: 29 September 2015

1 General comments

Reviewer: The manuscript is well written, though with short passages that are technical.

Response: The most technical parts, about the Lomb-Scargle periodogram (LSP), will be moved to the Supplementary Material (SM). This is not only because they were technical, but because the revised manuscript does not rely extensively on the use of this spectral estimator. On the other hand, in the revision we will define some

other estimators in the Methods section. This involves a few more equations, but they are simple and are included as an aid for readers who are not familiar with these techniques.

Reviewer: It is not overly exciting, as the relevance of the scaling models for future prediction is not made very clear.

Response: We encourage to read our comprehensive response to reviewer Lovejoy. Our paper deals with the most fundamental issues in statistical modelling of climatic records. Has scaling analysis any meaning without invoking statistical models, and how shall we select the model to use? The relevance for prediction is discussed there. In the revision we will discuss this issue more explicitly and in greater depth.

Reviewer: As a manuscript with paleoclimate focus this manuscript may be more suitable for a journal like "Climate of the Past".

Response: In the revision it will be even more clear that our purpose is to establish the correct statistical scaling model for Holocene climate as a tool for prediction of future climate on centennial time scales. This paper should be suitable for Earth System Dynamics.

Reviewer: At present several data (independence of the temperature reconstructions, calibration of the ice core data, effects of noise) and method aspects (bias in the estimation procedure around $\beta = 1$, use of Lomb-Scargle periodograms) are not sufficiently discussed and justified in the manuscript to allow for firm conclusions.

Response: All these issues are discussed in our responses to specific comments below. Revisions will be made where they are appropriate. That said, we are a bit skeptical to the use of the phrase “firm conclusions.” Many of the results we have obtained are of the negative type, i.e., we show that one cannot reject a null hypothesis, and hence renders the alternative hypothesis (e.g., the hypothesis of two scaling regimes) unsupported by the data. This means that we point out that previously published conclusions are subject to Type-I statistical errors (false positives = detecting an effect that is not present). The scientific literature is littered with Type-I errors, presumably because detecting new effects are perceived as much more exciting than proving that exciting findings may have less exciting explanations.

Reviewer: I suggest that the manuscript be subjected to minor revisions before it can be published.

Response: The paper will be subject to a major revision. The revision will be outlined at the end of this response.

2 Specific comments

Reviewer: Page 1202, line 15: Is this an important model for temperature prediction on centennial timescales?

Response: A recent paper proposes that the macroweather-climate regime model (SLIM) can be used as a basis for temperature prediction up to a few decades, (*Lovejoy et al.*, 2015). According to the multiple scale regimes hypothesis, prediction on centennial time scales the alternative SLIM-model should be based on a model with $\beta \sim 1.5$ or larger. The properties of such a prediction model would be considerably different

from one with e.g., $\beta \sim 0.7$. This issue will be discussed in greater depth in the revision.

Reviewer: Page 1205, line 10: It is not obvious why a Lomb-Scargle periodogram should deal better with irregularly sampled data than conventional techniques or other techniques such as spectral estimation via "correlation slotting" . There is evidence that it does not (Rehfeld Kurths, 2011). Uneven sampling causes bias when overcome by standard interpolation (leading to a red bias), but the LS periodogram also shows considerable high-frequency biases (blue bias) which may be worse. The argument that "the data is not standard, therefore the methods have to be special" is not necessarily true. A much more critical point, with a potentiallly larger impact on the scaling question, is the question of signal-to-noise ratios, and/ or the global representativeness of the used proxy records (see e.g. Laepple and Huybers, 2014).

Response: We agree with the reviewer that it is not obvious a priori that the Lomb-Scargle periodogram (LSP) performs better than other spectral techniques, so this will be discussed more in the revised version of our paper. We have looked into the papers by *Rehfeld et al.* (2011) and *Broersen et al.* (2000), and find that irregularly sampled data causes various problems for all spectral techniques. Slotting can be problematic because the covariance estimators may not be positive semi-definite, and could hence give negative values in the spectrum. The skill of the LS periodogram is as demonstrated by *Rehfeld et al.* (2011) dependent on the skewness of the distribution of sampling intervals. The bias will therefore differ from dataset to dataset, so we cannot determine just by reading these other papers which method is the most suitable for our purpose. In the originally submitted manuscript, the skill of the LSP for our data is illustrated in figure 1 of the main article. Surrogate data was generated with sampling times equal to the low resolution GRIP Holocene/last glacial period. The figure does indeed show a slight underestimation for all values of β, due to the blue bias.

We have recently become aware that one important aspect of the ice core data has been overlooked when we designed the surrogate data: namely the gradual, natural smoothing of the ice core proxy data as one goes back in time. Our surrogate data are sampled at irregular time steps, but as point measurements. In reality, the given values of stable isotopes from ice cores are generally averages over the time period covered by the sliced equal-length samples. The real proxy values will therefore exhibit less variability than if they were instantaneous measurements of the true temperature. This smoothing effect will increase as one goes back in time due to compression of the core. The smoothing effect constitutes a red bias that works in the opposite direction of the blue bias introduced by the LSP. Taking this into account we have created new surrogate data that resembles proxy data to a larger degree than before. The new surrogates are created by first generating regularly sampled time series, which are then averaged over time intervals equal to the GRIP time intervals to obtain irregularly sampled surrogate data. The results are shown in Figure 1 of this response, and demonstrates that the blue bias from the LSP is approximately balanced by the red bias from the smoothing.

Despite our findings of the generally good performance of the LSP for our purpose, a second look at the interpolation + standard periodogram approach shows that it is actually adequate for most of the data studied in our paper. Hence, we have decided to use it in the revised version of the paper, and put results from LSP, including figure 1, in the revised supplementary material. Hopefully this will also meet the complaints of both reviewers that the paper is “too technical."

Reviewer: Page 1206, line 22: This length limit assumes independence of the data points - which is not justified for proxy records.

Response: This “rule of the thumb” is not based on an assumption of independence of the data points, but on experience with application of various estimators on ensembles

of surrogate fGns with relevant scaling exponents ($0 < \beta < 1$). For scales longer than $N/4$ the statistical spread of the fluctuation function or the power spectral density gets so large that it contributes to unacceptable uncertainty in the estimate of the scaling exponent. The rule is illustrated by the increasing uncertainty spread towards lower frequencies in our plots of power spectra. Here it is also clear that the spread increases for lower β (weaker dependence between points).

Reviewer: Page 1210, line 12: Which figures in the supplement are referred to?

Response: It is referred to section S1 in the supplement, and fig S1 and S2. We will make this clear in the revised version.

Reviewer: Why is the case of $\beta = 1$ so awfully difficult? This is a significant drawback of the methodology!

Response: In figure 1 of our main article, a negative bias is seen for synthetic surrogate data with β slightly above unity. This bias is explained in more detail in section S1 (p1 l 5-10) but we realize that we do not explicit refer to this explanation in section 3.2 of the main article. It is not completely clear to us if the reviewer finds the explanation in the supplementary material unsatisfactoy, or if the main error is that we do not explicitly refer to the supplement for details about the bias. In the revised paper we will refer to the right section in the supplementary material for a detailed explanation of this bias. To make sure the bias for synthetic data with β=1 is properly explained, we comment on this below.

Synthetically generated discrete fGns and fBms exhibit scaling only asymptotically, i.e., for time scales much larger than the sampling time. At the shortest time scales the variability is higher than what we expect from a continuous process. This effect is larger for the case β just above 1 than for the other values of β in our study. It leads to a flattening of the spectrum as the frequency approaches the Nyquist frequency and hence reduces the estimated β if this frequency range is used in the fitting of a straight line to the spectrum in a log-og plot. For very long time series it is no problem to leave out this range in the fitting, and then there will be no bias. For short time series, including those with approximately 2000 data points used here, it is not possible to completely avoid this effect without increasing the uncertainty of the estimate (a smaller fitting range enhances uncertainty).

Reviewer: And: Is the LS-Periodogram actually doing a better job here than interpolation+(multitaper) spectral estimation?

Response: The main motivation for using the LSP instead of interpolation and a standard spectral estimator has been to avoid the red bias on high frequencies associated with interpolation. The LSP may be affected by the blue bias on the same frequencies. A red or blue bias in the power spectrum may incorrectly be interpreted as a scale break, and it may influence the frequency range we choose for estimation of a spectral exponent. The periodogram is our preferred estimator for the power spectral density, and we have therefore repeated our scaling analysis using interpolation plus periodogram for the high-resolution GRIP data set and the EPICA data set with lower temporal resolution. The other ice core data sets analysed in the supplementary material have similar temporal resolution as EPICA. The interpolation + periodogram approach works very well for both data sets in the frequency range up to the frequency corresponding to inverse of the mean sampling interval, i.e., the interpolation has very little effect for this frequency range. The spectra obtained are very similar to the LSP's, and the estimates of the scaling exponent β are similar. Since the two methods both perform well for our purpose, we choose to employ the more standard technique in the main article.

Reviewer: P1211, Line 15: Why are the wavelet analyses necessary? What are they supposed to contribute to the main message of the manuscript?

Response: This is discussed at length in our response to reviewer Lovejoy.

Reviewer: P1211, line 24: Are these reconstructions actually independent, or do they use the same background proxy series for reconstruction?

Response: First, let us acknowledge a small error that was discovered after the mansucript was published. On page 1212 line 10 it is referred to the Mann et al. (2009) reconstruction, but this is not correct. The reconstruction that was used is *Mann et al.* (2008), the error will be corrected both in the main text and in the bibliography.

Regarding the independence of the proxy records, all papers except one provide the detailed information necessary to track the origin of the records. *Briffa et al.* (2001) use tree-ring records but does not state explicitly the source. For the remaining 6 proxy/multiproxy reconstructions, there are data sets that are used in more than one reconstruction, including tree-ring records and low-resolution proxy records. Independence is therefore not guaranteed, but the reconstruction methods are different. The records are therefore processed in different ways, which may influence the scaling of the resulting reconstructions.

Reviewer: Page1212, line 23: If these are temperature reconstructions, what were the calibration models? Linear regression based on O18? Or nonlinear models? Were they adjusted for Holocene and Glacial differences? Same for l. 4 on p1213 (deuterium)

Response: This is an error from our side. The records that are used are not converted to temperature, they are δ^{18}O records (GRIP) and the EPICA record is δD. What was meant in the present text was that the δ^{18}O/ δD proxies are related to temperature. This will be corrected in the revised paper.

Reviewer: Page 1216, line 10: Why should DO-events flatten the spectrum on centennial and shorter timescales? This is not obvious (to me).

Response: This was an inaccurate statement from our side, and we will remove this claim from the revised paper. The DO events affect the entire spectrum spectrum. Preliminary spectral analysis of the stadial and interstadial stages separately suggest scaling with $\beta \sim 1$ in both stages, so the discontinuities associated with the transitions between the two stages and their distribution in time, is what alters the spectrum an gives rise to what appears as a scale break. It should be mentioned that we have checked that this is not an artifact due to some blue bias of the LSP, since we observe the same flattening for interpolation + periodogram. Further detailed analysis of the scaling structure of the glacial state (which is very complex and very interesting) is beyond the scope of the present paper.

Reviewer: Page 217, lines 24/25: Basically, what you write here is that there is no power in this test of beta differences. So why use it?

Response: This simple test on synthetic LRM processes demonstrates that it is relatively easy to detect a scale-break in a time series. On the other hand, without testing the statistical significance of the break, the data analysis is not enough to claim that the scale-break is a real phenomenon. The fGn is defined from only one scaling

exponent, but spurious breaks may occur at random without being significant.

Reviewer: Page 1218. These paragraphs are nice to read.

Response: Thank you! They will be expanded.

Reviewer: Page 1219, line 28: This sentence, and the following one, is hard to digest, in particular to the nice and easy to read discussion before.

Response: In the revision, we shall expand the discussion section and in particular the part dealing with intermittency and multifractality.

3 Outline of revision

In the abstract we add the modified sentences:

- A model for internal variability with only one regime is simpler, and allows more certain predictions on time scales of centuries when combined with existing knowledge of radiative forcing.
- Spectra from a number of late Holocene multiproxy temperature reconstructions, and one from the entire Holocene, have also been analysed, without identifying a significant scale break.

In section 1 (Introduction) we make the definitions of "scaling regime" and "scale break" more precise, and we give a very brief introduction to the concept of fluctuation analysis and the Haar fluctuation function.

In section 2 (The concept of multiple scaling regimes) we only make a few minor updates.

In section 3 (Methods) we move the description of the Lomb-Scargle periodogram to the Supplementary Material, and we add one subsection on structure functions and the scaling function and one on the Haar fluctuation function. Both are accompanied with a figure where the methods are applied to the instrumental global temperature record and used to contrast the differences between modelling the record as a process consisting of two scaling regimes or modelling it as a fractional noise plus a trend.

In Section 4 (Data) we add a subsection on the Marcott multiproxy series.

In section 5 (Results) we expand the discussion of the Moberg multiproxy record and add two figures. One employs structure function and scaling function analysis to show the non-intermittent nature of this record, and also to demonstrate the effect of introducing a model that involves an oscillatory trend. In another figure, the contrast between such a model and a two-scaling regime model is shown by using the Haar fluctuation analysis on ensembles of realisations of these two models.

A new subsection is added where the Marcott reconstruction is analysed, correcting for the gradual smoothing (reduction of the high-frequency variability) of the proxy signal as one goes backward in time towards the early Holocene (see the discussion in our response to Lovejoy).

In section 6 (Discussion and conclusion) we will include of some of the material in the response to Lovejoy. In particular we will discuss in more depth the ambiguities arising when performing scaling analysis on data from evolving or otherwise non-stationary systems. Here we will add a wavelet scalogram of GRIP data illustrating the difference between Holocene and glacial climate, and also the change in variability between stadials and interstadials during the ice age. This scalogram also show that the longest interstadials exhibits a variability very similar to the Holocene.

In the Supplementary Material some of the material and figures shown in our response to Lovejoy will be included.

This work was funded by project no. 229754 under the the Norwegian Research Council KLIMAFORSK programme.

References

K. R. Briffa and T. J. Osborn and F. H. Schweingruber and I. C. Harris and P. D. Jones and S. G. Shiyatov and E. A. Vaganov: Low-frequency temperature variations from a northern tree ring density network, J. Geophys. Res.-Atmos, 106, D3, doi:10.1029/2000JD900617, 2001.

Broersen, P.M.T. and S. de Waele and R. Bos: The accuracy of time series analysis for laser-doppler velocimetry, 10th International Symposium on Application of Laser Techniques to Fluid Mechanics, Lisbon, July 10-13, 2000

Lovejoy, S. and del Rio Amador, L. and Hébert, R.: The Scaling LInear Macroweather model (SLIM): using scaling to forecast global scale macroweather from months to decades, Earth System Dynamics Discussions, 6, 1, doi: 10.5194/esdd-6-489-2015, 2015

Mann, Michael E. and Z. Zhang and M. K. Hughes and R. S. Bradley and S. K. Miller and S. Rutherford and F. Ni: Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, PNAS, 106, 36, doi: 10.1073/pnas.0805721105, 2008.

Marcott, S. A., Shakun, J. D., Clark, P. U., and Mix, A. C.: A reconstruction of the regional and global temperature for the past 11.300 years, Science, 339, 1198, doi: 10.1126/science.1228026 (2013).

Rehfeld, K. and N. Marwan and J. Heitzig and J. Kurths: Comparison of correlation analysis techniques for irregularly sampled time series, Nonlinear Processes in Geophysics, 18, 389-404, doi:10.5194/npg-18-389-2011, 2011

Fig01.pdf