Haar Fluctuations Scaling Analysis Software Without Interpolation

August 13, 2014

Mathematica Function HaarNoInterpolate

1 Basic Summary

This function performs a scaling analysis of fluctuations defined through the Haar Wavelet, that is for a given time interval the difference between the average of the first half and the second half of the interval. This allows us to estimate the mean and the RMS scaling exponents H without using interpolation, which assumes that the curve is continuous and differentiable between points. Such an assumption is problematic since in the small scale limit of a scaling regime, the mean derivatives of order greater than H diverge. Therefore, even linear interpolation may give spurious results as most relevant atmospheric regimes have H smaller than 1 and thus, any linearly interpolated part of the series will locally have H equal to 1. This software instead rejects improper fluctuations on the basis of the ratio $\epsilon = \frac{t_{j+k/2}-t_j}{t_{j+k}-t_j}$ which is the ratio between the j^{th} and the $j+k^{th}$ element in the series; regular sampling implies $\epsilon = \frac{1}{2}$. A parameter ϵ_{min} was chosen so that fluctuations with $\epsilon \in [\epsilon_{min}, 1-\epsilon_{min}]$ would be used, but those with $\epsilon \notin [\epsilon_{min}, 1-\epsilon_{min}]$ would be rejected . It was found that $\epsilon_{min} = \frac{1}{4}$ was a good compromise to increase the number of fluctuations considered without losing too much precision in scale. (See Appendix A)

2 Input

This function requires 5 inputs: $HaarNoInterpolate[Time, Field, Calibration, LowerPointsToDrop, UpperPo$

- $Time$ is the series containing the temporal position of the points in the field under consideration.
- $Field$ is the field under consideration itself, for example a serie of temperature values.
- $Calibration$ is a scalar which multiplies $Time$, allowing to adjust for units. For example, if time is given in kiloyears, $Calibration$ would be set to 1000 to transform $Time$ in years.
- $LowerPointsToDrop$ indicates how many points to drop, starting in order from the highest resolution, before performing the fit for H
- $UpperPointsToDrop$ indicates how many points to drop, starting in order from the lowest resolution, before performing the fit for H

3 Output

The function returns the following output

- The number of fluctuations used in the analysis.
- The total number of points : The number of time resolutions for which fluctuations were calculated.
- The range over which the fit for H was performed.
- H : the mean exponent H
- The RMS exponent $= \frac{\xi(2)}{2}$: the RMS exponent. (If $C_1 = 0$, this is H, otherwise the correction $\frac{K(2)}{2}$ is given.)
- Derivative C_1 : The codimension of the mean calculated from ξ.
- α from psipp(1) : Multifractality index calculated from ξ.
- C_1 graph slope method: The codimension of the mean calculated from the C_1 graph.
- alpha from graph slope : Multifractality index α calculated from the C_1 graph.
- Schmitt α : Multifractality index α calculated using the second derivative of $\xi(q)$ at $q = 0$. (Only valid for $\alpha > 1$)

4 Example 2D

HaarNoInterpolate The analysis was performed on the $EPICA$ temperature time series, which was reconstructed from $\delta^{18}O$ concentrations.
Input

- $Time = Epica[[All, 3]]$ (Series of years BP at which the ice core was sampled)
- $Field = Epica[[All, 5]]$ (Series of the reconstructed temperature)
- $Calibration = 1$ (The time series is already given in years)
- $LowerPointsToDrop = 27$ (Removed the first 27 points before fitting)
- $UpperPointsToDrop = 27$ (Removed the 27 last points before fitting)

Output

```
In[39]:= HaarNoInterpolate[Epica[[All, 3]], Epica[[All, 5]], 1, 27, 27]
          the number of flucutations used in the analysis is = 285345.
          The total number of points is 93
          Fits are over the range   316.228  to  25118.9  in the specified units
         H=   0.483155
         the RMS exponent= ξ(2)/2= 0.508424  the difference with H is K(2)/2 = -0.0252175
         derivative C1=  -0.0318943
         α from psipp(1) =  1.58164
         C1 graph slope method =-0.0318052
         alpha from graph slope = 1.57968
         Schmitt α=  1.65656
```

Out[39]= {

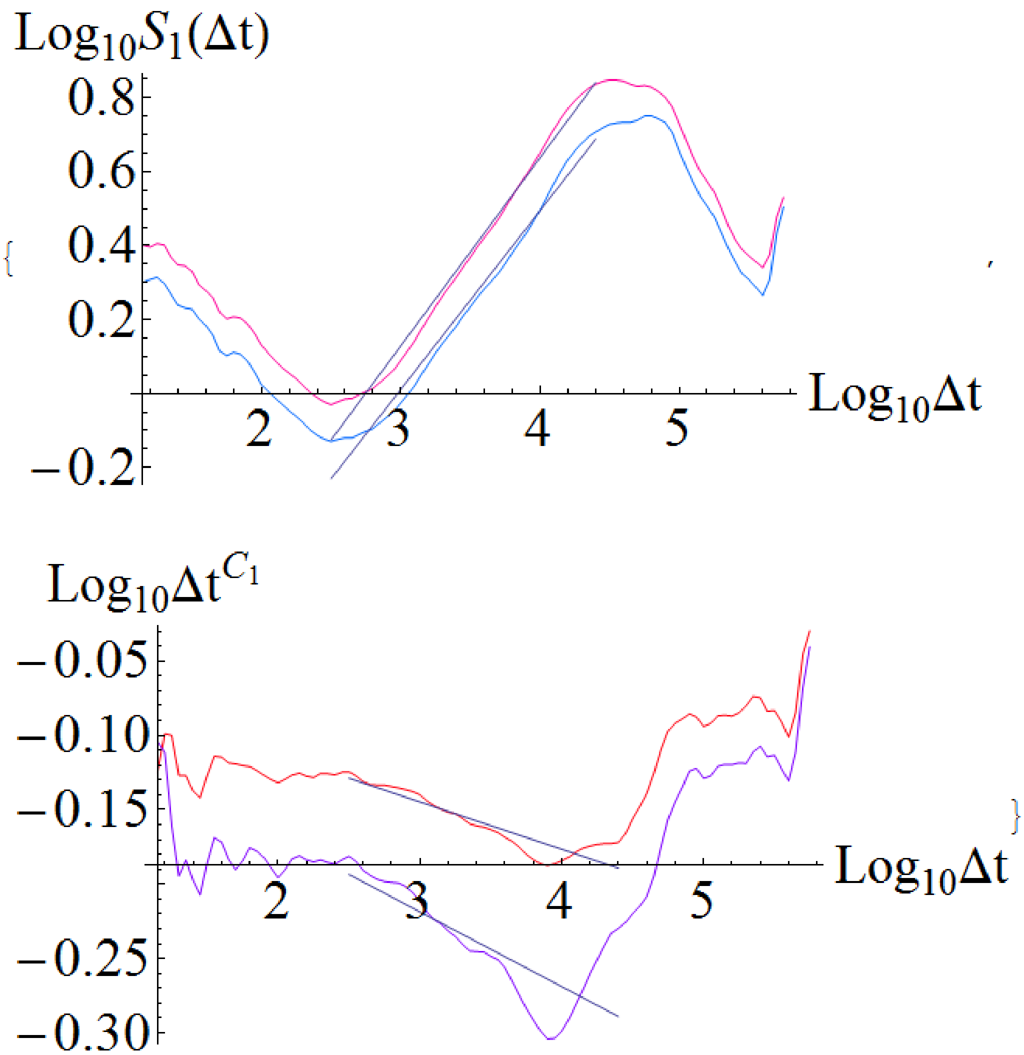

Figure 1: Input and related output of the function HaarNoInterpolate applied to the temperature series reconstructed from the *EPICA* core. Notice that the fitted lines on the graphs only appear over range on which the fit was performed for H, which is from 316 *years* to 25119 *years*. We can distinguish three scaling regimes: *Macroweather*, *Climate* and *Macroclimate*. H was found to be 0.48 for the *Climate* regime which was investigated here, that $H > 0$ means fluctuations increase with scale. In the *Macroweather* and *Macroclimate* regimes, $H < 0$, meaning that fluctuations tend to cancel each other out at those time scales, i.e. less than 316 *years* or more than 25119 *years*.

Appendix A of [*Lovejoy*, 2014]

An interpolation-free algorithm for estimating Haar fluctuations:

Paleotemperatures are typically nonuniformly sampled in time. Sometimes – such as in the case of the Epica series used in fig. 1b – the problem is due to the compression of the ice with depth and the problem can be somewhat alleviated by sampling the deeper reaches of the core at higher rates (e.g. the high resolution section of the GRIP core shown in fig. 2b). However, the usual remedy is to interpolate the series and then to resample it at a uniform temporal interval/resolution. While for many purposes this may be adequate, for either spectral or fluctuation analyses it may lead to biases and spurious results. The reason is that interpolation assumes that the curve is not only continuous between points, but also that the series $T(t)$ is differentiable (in the common case of cubic spline interpolation, up to third order!). However in the small scale limit, in a scaling regime, the mean derivatives of order $>H$ diverge. Since we have found empirically that all the relevant atmospheric regimes have $H<1$, even linear interpolation may give spurious results. Indeed, any linearly interpolated part of the $T(t)$ series will at least locally have $H=1$ since over such segments, $\Delta T(\Delta t) \approx \Delta t$. Therefore if these regions are too numerous, including the fluctuation statistics over linear segments will introduce biases.

One of the many advantages of Haar fluctuations is that they are quite easy to estimate without any interpolation while accurately taking into account the resolution of the data. We now describe the simple algorithm used in figs. 4b-e, 5 (note that several of these series were already analysed but using interpolation). Assume that there are N measurements of temperature $T(t_i)$ at time t_i where i is an index 1 through N. Define the running sum S_i:

$$S_i = \sum_{j \le i} T\left(t_j\right) \qquad \text{(A1)}$$

Consider an index j and an even number k. The j, k fluctuation $\Delta T_{j,k}$ over the interval $[t_j, t_{j+k}]$ can be estimated as follows. First determine the sums of the $T(t_i)$ over the first and second halves the interval:

$$\Delta S^{(1)} = S_{j+k/2} - S_j; \quad \Delta S^{(2)} = S_{j+k} - S_{j+k/2} \qquad \text{(A2)}$$

in the case of regular sampling, the ratio:

$$\varepsilon = \frac{t_{j+k/2} - t_j}{t_{j+k} - t_j} \qquad \text{(A3)}$$

has the value $\varepsilon = 1/2$.

The Haar fluctuation is simply the average of the first half minus the average of the second half of the interval and can thus be estimated as:

$$\Delta T_{j,k} = \frac{2}{t_{j+k} - t_j}\left(S^{(1)} - S^{(2)}\right) \quad \text{(A4)}$$

However, if ε is too far from 1/2, this estimate may be poor. Therefore, in the calculation of the statistical moments we should only keep the corresponding fluctuations on condition that $\varepsilon_{min}<\varepsilon<(1-\varepsilon_{min})$ where $0<\varepsilon_{min}<1/2$ is a parameter that can be adjusted so as to make the condition as restrictive as we like: exactly uniform sampling corresponds to the limit ε_{min} ->1/2. Decreasing ε_{min} has the effect of losing precision in the scale Δt, hence it smooths the $S(\Delta t)$ curve. However, taking ε_{min} too close to ½ will result in the rejection of too many fluctuations with the consequence that the statistics will be poor. In the present case, it was found that generally ε_{min} = 1/4 was a reasonable compromise (see fig. 1). One can check the accuracy by seeing how much the statistics change when ε_{min} is varied (if they don't vary much then the choice of ε_{min} is acceptable). Note also that as usual, the fluctuations are multiplied by an extra "calibration" constant (taken throughout this paper = 2). This ensures that they are quite close to differences in regions where H>0 and close to tendencies (averages with the means removed) in regions where H<0. Once the fluctuations are estimated, $S_q(\Delta t)$ can be estimated by "binning" the fluctuations into "bins" with Δt regular spaced logarithmically. For each bin, the various powers of ΔT are averaged, in our implementation of the algorithm we used 20 bins per order of magnitude in Δt (the software available from http://www.physics.mcgill.ca/~gang/software/index.html).

While the above procedure essentially solves the problem of "holes" in the series, it does not remove possible biases that arise from systematic sampling nonuniformities such as those arising from cores with high temporal sampling rates near the surface and systematically lower rates at depth. When applied to such series, the small Δt part of the S(Δt) function will be sampled from the top part of the core where all the high resolution data lie. Therefore the high frequencies will be biased towards the near surface statistics. However, if the statistics are fairly homogeneous in time - as they typically are (see fig. 5)– then this is unimportant (see however [*Lovejoy and Schertzer*, 2013] for evidence of exceptional Holocene statistics in Greenland).

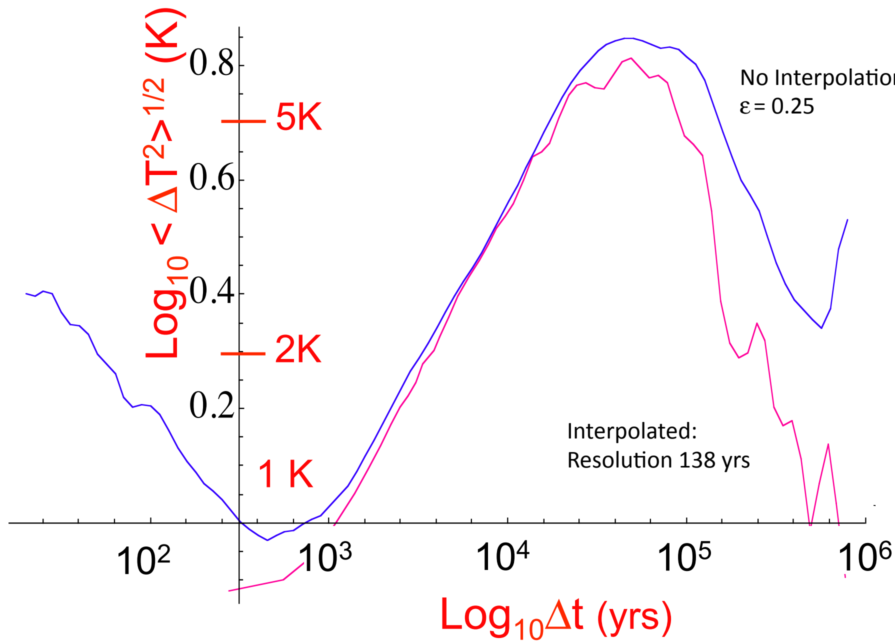

Figure 1: A comparison of the Epica analysis using a uniform sampling on the linearly interpolated data using the same number of data points as in the original series (5788 points, interpolated resolution 138 yrs), magenta, and the result of the interpolation free algorithm described here using ε_{min} = 0.25 (blue). The main differences are at the small and large Δt's. The magenta interpolated curve is reproduced from [*Lovejoy*, 2013].

References

Lovejoy, S. (2013), What is climate?, *EOS*, *94, (1), 1 January*, p1-2.

Lovejoy, S. (2014), A voyage through scales, a missing quadrillion and why the climate is not what ou expect, *Climate Dyn.*, *(submitted, 2/14)*.

Lovejoy, S., and D. Schertzer (2013), *The Weather and Climate: Emergent Laws and Multifractal Cascades*, 496 pp., Cambridge University Press, Cambridge.