Atmos. Chem. Phys. Discuss., 11, C5948–C5951, 2011
www.atmos-chem-phys-discuss.net/11/C5948/2011/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Comparing the effectiveness of recent algorithms to fill and smooth incomplete and noisy time series" *by* J. P. Musial et al.

J. Musial

jan.musial@giub.unibe.ch

Received and published: 4 July 2011

We are very grateful to anonymous referee #2 for the useful comments that allowed us to improve our study. Primarily, we have added new quality fit criteria namely: chi-squred test, mean bias error (or just mean error) and autocorrelation of residuals. The new analysis performed did not change the main findings about the performance of the selected gap-filling and smoothing methods, but it significantly supports our conclusions. It will be enclosed in the final version of the article. As far as minor comments of referee #2 are concerned these are our replies are as follows:

1.) About your third paragraph, I think that your use of "broad, slow variations" vs

"random" , "signal" and "noise" is rather ambiguous since, at least in climatology, the "signal" could be strongly aperiodic.

In this paragraph we state that "In many (but not all) practical cases, the broad, slow variations that offer some degree of predictability are of greater interest than the fast changes, which often appear as random, unpredictable events of lesser consequence, such as uncertainties in the measurements. By analogy with such fields as acoustics and radar, the interesting variations in the time series are called the signal' and all other variations are referred to as 'noise'.". Here we do not assume a periodicity of a signal and we just want to point out that in most cases (but not all) slow, broad variations associated with lower frequencies are of greater interest than rapid changes at higher frequencies which are often (but not always) caused by random, unpredictable events.

2.) Personally, I don't know the IMSL routine CSSMOTH : is a GCV method ? Did you use this method for each time series analyzed ? Does it increase the computational cost of this method ? Please clarify the end of this paragraph.

We have added some comments about this routine at the end of paragraph. This routine, as well as others mentioned, were used across all experiments. On the basis of the performance test (added in the new version of article) it could seen that smoothing spline is extremely robust and the GCV procedure does not have significant influence on the CPU time . More details about CSSMOTH can be found on: http://idlastro.gsfc.nasa.gov/idl_html_help/IMSL_CSSMOOTH.html

3.) A flow-chart will be useful to present each step used in SSA.

For more details about the SSA method (including some graphical representation) please refer to the book by Golyandina, N., Nekrutkin V., and Zhigljavsky A.: Analysis of Time Series Structure: SSA and Related Techniques, Chapman and Hall, Washington D.C., 2001.

4.) I think, as said above, that your hypothesis that high frequency variations are

random is too restrictive. I add that "high-frequency" and "low-frequency" is not precise enough (section 3.3).

Terms high-frequency and low-frequency are indeed ambiguous as they depend on a particular time series or a particular type of analysis. Therefore, it is impossible to derive a precise definition or critical value of a change between high-and low-frequencies. We do not impose that always high frequency variations are random but we state that “typically” this is the case.

5.) please refer to your equations in the first paragraph of section 3.3

Suitable references were added.

6.) what is the increment of % noise and % of missing entries, 1% ?, 5% ? 10% ?

The increment of percentage of noise and gaps on both axes is 10%. Suitable comment was added to the text.

7.) adapt the y-axis of fig. 1.

Y axes on figures 1-4 were adapted to show small significant differences between models rather then to present the extremely high standard deviations associated with erroneous reconstructions of a original dataset.

8.) "is more depend" on the last line sounds odd. Please check.

This spelling mistake was corrected.

9.) about the winter-gap scenario, I think that your comparison between Lomb- Scargle (LS) and Kondrashov and Ghil (KS) methods on one hand and spline on the other hand is unfair since what happens for a summer-gap scenario ? I think that spline doesn’t produce spurious peak as soon as you have enough data to reconstruct the autumn decrease and the spring increase, but this methods won’t be superior to other ones if you don’t get these tendencies. I understand that spurious peak appears because there is an intermittent "bi-annual" cycle which is simulated for each year by LS and

SSA methods. But I think that the superiority of spline is mostly a matter of chance in your example. Your use of term "spurious" is perhaps ambiguous in that context and "intermittent" could be a better choice. This comment applies also to what is written in the second paragraph of section 5.3.

One example of the summer-gap scenario was added to the text. It could be seen that in this case the problem of spurious, intermittent peaks remains, but this time second minimum is introduced, as opposed to winter gap scenario, where a second maximum occurs. We do agree that a spline function reconstruction of a time series with the winter gap scenario depends only on few points just before and after a gap (this was emphasized in the description of the continuous gap scenario). We use the term ”spurious” in a sense of: “Lacking authenticity or validity in essence or origin; not genuine; false” (definition from: http://www.thefreedictionary.com/spurious). However, we have added term “intermittent” for an easy of better understanding.

10.) Can you give us an objective measure of the "considerable computational cost" for example as CPU for each method and how it increases with length of observed or simulated time series, and possibly the amount of missing entries to be filled. This would help to scale the advantage or disadvantage of each method in an objective way. The same comment applies to what is said at the end of the 2nd paragraph p.7 ("it will be seen that these methods ... ") and also page 11 ("on the other hand, the high computational requirements of the SSA gap-filling algorithm").

A graph presenting the CPU time of selected gap-filling and smoothing techniques as a function of time series length was added.

Yours sincerely,

Jan Musial

Interactive comment on Atmos. Chem. Phys. Discuss., 11, 14259, 2011.