

Interactive comment on “Gap filling and noise reduction of unevenly sampled data by means of the Lomb-Scargle periodogram” by K. Hocke and N. Kämpfer

Anonymous Referee #1

Received and published: 27 March 2008

General Comments:

The paper reports results of applying a modified Lomb-Scargle periodogram for gap-filling of unevenly sampled datasets. Such datasets are common in geosciences, and so this manuscript should be of interest to a wide audience. I find very interesting, though not necessarily clear, in how this spectral estimation method has been modified for the purpose of reconstruction, and especially its impressive results on synthetic and real data. After relatively minor revisions are made as outlined below it should be published.

I would like to commend the authors for spirit of "reproducible research" and supplying

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



the Matlab script with the underlying algorithm. It allows to verify synthetic examples in the paper, as well as to apply the method to time series with different spectral properties than the ones considered here. In doing so one can test method capabilities and limitations, and consider further improvements. I encourage authors to supply as auxiliary material the data for Figs 4-7 as well.

I find that the algorithm performs very well (at least with the synthetic data) when 1) the "signal" consists of pure harmonic signals, and is not quasi-periodic (modulated in amplitude for example), 2) the time series is relatively long and high resolution, so there is a good number of repetitions of the signal 3) the fraction of the missing data is relatively small, such as for Fig.2 example in the paper (~10 days from 60 days). 4) The noise is mostly "white".

Some of these restrictions are probably inherently due to the fixed cosine/sine basis and periodogram estimation, which may be not optimal in case of short and very noisy time series. It would be good for a wider audience if authors could comment on some of these points, as well as on the multivariate extension of the presented methodology, especially since non-fixed basis alternatives, i.e. as in Kondrashov and Ghil (2006), are less restrictive.

Specific Comments:

Section 2.2.

I find the description of this key innovation to the method very short and somewhat confusing, especially since a key reference Hocke (1998) can not be located in online journal access! Some of the code-level implementation details should be included in the Matlab script. Instead, authors should better explain connection of Eq. (11-13) with Eqs (2-3). Here are a few other points which are not clear:

- is "n" in Eq. (10) equal to N?
- p. 4608, l. 18, Why is it important to start with $t_1=0$?

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

- p. 4608, l. 20 Why t_{ave} is needed? It is not convenient to look for a fortran code as a reference.

- Why oversampling factor equal to 4 is needed in Eq. (15) ?

Section 3.

Suggest to mention number of data points for all the time series.

discussion of Figs.3 and 5: authors should explain how they determine confidence levels in the text.

p. 4610, l. 21: When running Matlab demo, there are unwanted effects in the reconstruction when using Hamming window; the variance is increasing towards the both ends of the time series.

p. 4611 l. 8; Should comment on when noise in data is colored (red).

Technical corrections:

p. 4612, l. 17: should be time-mean.

Interactive comment on Atmos. Chem. Phys. Discuss., 8, 4603, 2008.

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)