Atmos. Chem. Phys., 9, 4197–4206, 2009
www.atmos-chem-phys.net/9/4197/2009/

Atmospheric Chemistry and Physics

Gap filling and noise reduction of unevenly sampled data by means of the Lomb-Scargle periodogram

K. Hocke and N. Kämpfer

Institute of Applied Physics and Oeschger Centre for Climate Change Research, University of Bern, Switzerland

Received: 25 January 2008 – Published in Atmos. Chem. Phys. Discuss.: 4 March 2008
Revised: 25 May 2009 – Accepted: 15 June 2009 – Published: 24 June 2009

Abstract. The Lomb-Scargle periodogram is widely used for the estimation of the power spectral density of unevenly sampled data. A small extension of the algorithm of the Lomb-Scargle periodogram permits the estimation of the phases of the spectral components. The amplitude and phase information is sufficient for the construction of a complex Fourier spectrum. The inverse Fourier transform can be applied to this Fourier spectrum and provides an evenly sampled series (Scargle, 1989). We are testing the proposed reconstruction method by means of artificial time series and real observations of mesospheric ozone, having data gaps and noise. For data gap filling and noise reduction, it is necessary to modify the Fourier spectrum before the inverse Fourier transform is done. The modification can be easily performed by selection of the relevant spectral components which are above a given confidence limit or within a certain frequency range. Examples with time series of lower mesospheric ozone show that the reconstruction method can reproduce steep ozone gradients around sunrise and sunset and superposed planetary wave-like oscillations observed by a ground-based microwave radiometer at Payerne. The importance of gap filling methods for climate change studies is demonstrated by means of long-term series of temperature and water vapor pressure at the Jungfraujoch station where data gaps from another instrument have been inserted before the linear trend is calculated. The results are encouraging but the present reconstruction algorithm is far away from being reliable and robust enough for a serious application.

Correspondence to: K. Hocke
(klemens.hocke@iap.unibe.ch)

1 Introduction

Atmospheric data are often unevenly sampled due to spatial and temporal gaps in networks of ground stations, radiosondes, and satellites. Many measurement techniques such as lidar, solar UV backscatter, occultation depend on weather, solar zenith angle, or constellation geometry and cannot provide continuous data series. Unevenly sampled data have one advantage since aliasing effects at high frequencies are reduced, compared to evenly sampled data (Press et al., 1992).

There are several approaches to come over with undesired data gaps: linear and cubic interpolation and triangulation (appropriate for small gaps), spherical harmonics expansion, Kalman filtering and Bayesian cost functions (using a priori knowledge), and data assimiliation of observations into atmospheric general circulation models. Atmospheric parameters often have periodic oscillations due to forcing processes within the dynamical system of Sun, atmosphere, ocean, and soil. Thus, gap filling by spectral methods can also be considered as an efficient tool.

Adorf (1995) performed a very brief and valuable survey on available interpolation methods for irregularly sampled data series in astronomy. Kondrashov and Ghil (2006) describe the application of singular spectrum analysis for spatio-temporal filling of missing points in geophysical data sets. The leading orthogonal eigenvectors of the lag-covariance matrix of data series are utilized for iterative gap filling. Ghil et al. (2002) give a detailed review on advanced spectral methods for reconstruction of climatic time series (e.g., singular spectrum analysis, maximum entropy method, multitaper method).

The review of Ghil et al. (2002) does not mention the Lomb-Scargle periodogram which provides the least-squares

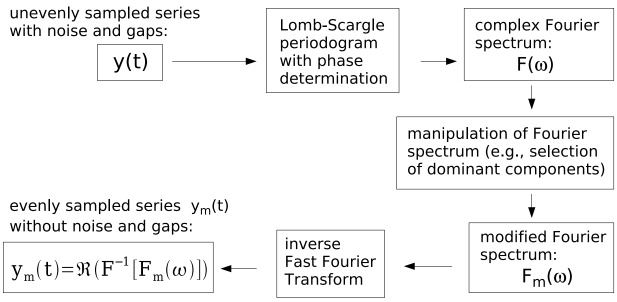

Fig. 1. Flow chart of the reconstruction method: Phases and amplitudes of the spectral components of the series $y(t)$ are calculated with a special version of the Lomb-Scargle periodogram (Hocke, 1998) and a complex Fourier spectrum $F(\omega)$ is constructed from these informations. The Fourier spectrum is modified, e.g., removal of spectral components which are below a certain confidence limit. The real part of the inverse fast Fourier transform of the modified Fourier spectrum $F_m(\omega)$ provides the evenly spaced series $y_m(t)$ without noise and without gaps.

spectrum of unevenly sampled data series (Lomb, 1976; Scargle, 1982; Press et al., 1992). The Lomb-Scargle periodogram reduces to the Fourier transform in case of evenly sampled data. In presence of data gaps, the sine and cosine model functions are orthogonalized by additional phase factors (Lomb, 1976). Scargle (1989) investigated the reconstruction of unevenly sampled time series by application of the Lomb-Scargle periodogram and subsequent, inverse Fourier transform. The potential of the study of Scargle (1989) for gap filling of atmospheric data sets has not been recognized yet, maybe because the article has been published in an astronomical journal. Another reason is that most available computer programs of the Lomb-Scargle periodogram solely derive the spectral power density but not the phase or the complex Fourier spectrum. Hocke (1998) modified the Lomb-Scargle algorithm (Fortran program "period.f") of Press et al. (1992) in such a way that amplitudes and phases are returned as functions of frequency and tested the program by means of artificial time series. The modified algorithm has been applied in various studies, e.g., for estimation of amplitudes and phases of tides, planetary waves, and gravity waves in the mesosphere and lower thermosphere, and for investigation of energy dissipation rates in the auroral zone (Nozawa et al., 2005; Altadill et al., 2003; Ford et al., 2006; Hall et al., 2003).

The present study continues the basic idea of Scargle (1989) and utilizes the Lomb-Scargle periodogram for reconstruction of time series. We explain the derivation of the complex Fourier spectrum from the Lomb-Scargle periodogram (Sect. 2). The complex Fourier spectrum should be reduced to its relevant spectral components before the inverse Fourier transform yields the equally spaced time series without gaps and without noise. The reconstruction method is tested with synthetic data and with real observations of lower mesospheric ozone (Sect. 3). Particularly, the tests with time series of lower mesospheric ozone show that the reconstruction method is not only appropriate for gap filling. The reconstruction method also supports the interpretation of the data and retrieves the regular daily change of lower mesospheric ozone with a high temporal resolution which is required around sunset and sunrise.

Finally the limitations of the Lomb-Scargle reconstruction method in presence of non-periodic fluctuations are discussed. A test with long-term series of temperature and water vapor pressure emphasizes the need of data gap filling methods for linear trend estimation of climatic series (Sect. 4). Only a few researchers actively investigate the effect of data gaps on the trend estimation. Funatsu et al. (2008) found that the effect of temporal sampling arising from the fact that lidar measurements are only made in nights without visible cloud cover introduces discrepancies that propagate on the calculation of temperature trends of the middle atmosphere. Such studies and work on reconstruction methods are necessary if we want to achieve higher accuracies in trend estimations.

2 Data analysis

The flow chart of the data analysis for reconstruction of data gaps and noise reduction is illustrated in Fig. 1. In the following, we explain the principle of the Lomb-Scargle periodogram. Special emphasis is put on the construction of a Fourier spectrum which is used for the inverse Fourier transform from the frequency domain back to the time domain. The selection of relevant spectral components has not been considered by Scargle (1989). This idea is related to the selection of principal components or leading orthogonal eigenfunctions of the lag-covariance matrix for filling of data gaps (Kondrashov and Ghil, 2006).

2.1 Lomb-Scargle periodogram

The Lomb-Scargle periodogram is equal to a linear least-squares fit of sine and cosine model functions to the observed time series $y(t_i)$ which shall be centered around zero (Lomb, 1976; Press et al., 1992).

$$y(t_i) = a\cos(\omega t_i - \Theta) + b\sin(\omega t_i - \Theta) + n_i; \quad \text{with} \quad i = 1, 2, 3, \ldots, N \tag{1}$$

where $y(t_i)$ is the observable at time t_i, a and b are constant amplitudes, ω is the angular frequency, n_i is noise at time t_i, and Θ is the additional phase which is required for the orthogonalization of the sine and cosine model functions of Eq. (1) when the data are unevenly spaced.

According to the addition theorem of sine and cosine functions, Eq. (1) can also be written as

$$y(t_i) = A\cos(\omega t_i - \Theta - \phi) + n_i; \tag{2}$$

with $\phi = \arctan(b, a)$ and $A = \sqrt{a^2 + b^2}$.

This writing style shows that a cosine function with amplitude A and phase $\varphi=\Theta+\phi$ is fitted to the observed time series. For the construction of the Fourier spectrum we need both, amplitude and phase.

The power spectral density $P(\omega)$ of the Lomb-Scargle periodogram is given by

$$P(\omega)=\frac{1}{2\sigma^2}\left(\frac{R(\omega)^2}{C(\omega)}+\frac{I(\omega)^2}{S(\omega)}\right), \tag{3}$$

$$R(\omega)\equiv\sum_{i=1}^{N}y(t_i)\cos(\omega t_i-\Theta), \tag{4}$$

$$I(\omega)\equiv\sum_{i=1}^{N}y(t_i)\sin(\omega t_i-\Theta), \tag{5}$$

$$C(\omega)\equiv\sum_{i=1}^{N}\cos^2(\omega t_i-\Theta), \tag{6}$$

$$S(\omega)\equiv\sum_{i=1}^{N}\sin^2(\omega t_i-\Theta). \tag{7}$$

σ^2 is the variance of the y series. The phase Θ is calculated with the four quadrant inverse tangent

$$\Theta=\frac{1}{2}\arctan\left(\sum_{i=1}^{N}\sin(2\omega t_i),\sum_{i=1}^{N}\cos(2\omega t_i)\right). \tag{8}$$

Scargle (1982) presented this formula of Θ as the exact solution for orthogonalization of the sine and cosine model functions of Eq. (1) in case of unevenly sampled data ($\sum\cos(\omega t_i-\Theta)\sin(\omega t_i-\Theta)=0$).

Equations (1–8) describe the Lomb-Scargle periodogram (Lomb, 1976; Scargle, 1982; Press et al., 1992; Bretthorst, 2001a). Scargle (1982) or Hocke (1998) explained that the variables R and I are closely related with the amplitudes of Eq. (1) ($a=\sqrt{2/N}R/\sqrt{C}$ and $b=\sqrt{2/N}I/\sqrt{S}$).

2.2 Construction of the complex Fourier spectrum

For calculation of the complex Fourier spectrum (as required for the FFT algorithm of the program language Matlab), the normalization of the spectral power density $P(\omega)$ of Eq. (3) is removed by multiplication of P with σ^2 (variance of the y series). The amplitude spectrum A_{FT} is

$$A_{FT}(\omega)=\sqrt{\frac{N}{2}\sigma^2P(\omega)}, \tag{9}$$

where N is the dimension of the complex Fourier spectrum F.

It is better to start the time vector with $t_1=0$ (before calculation of the Lomb-Scargle periodogram). Within the program period.f of Press et al. (1992), the phase is calculated with respect to the average time $t_{ave}=(t_1+t_N)/2$. The change of the phase due to the time shifts can be easily corrected for each spectral component (Scargle, 1989; Hocke, 1998). The phase φ_{FT} of the complex Fourier spectrum (inclusive phase correction ωt_{ave}) is given by

$$\varphi_{FT}=\arctan(I,R)+\omega t_{ave}+\Theta, \tag{10}$$

following the phase expression in Eq. (2).

The real part of the Fourier spectrum is

$$R_{FT}=A_{FT}\cos\varphi_{FT}. \tag{11}$$

The imaginary part of the Fourier spectrum is

$$I_{FT}=A_{FT}\sin\varphi_{FT}. \tag{12}$$

The FFT-algorithm of the program language Matlab requires a complex vector F of such a format:

$$F=[\text{complex}(0,0),\text{complex}(R_{FT},I_{FT}),\ \ldots\ \text{reverse}[\text{complex}(R_{FT},-I_{FT})]]. \tag{13}$$

The first number is the mean of the time series (zero mean), then comes the complex spectrum, and finally the reverse, complex conjugated spectrum is attached.

For the frequency grid, we used an oversampling factor ofac=4 in order to have a smaller spacing of the frequency grid points allowing an improved determination of the frequencies of the domimant spectral components. The frequencies are at

$$\omega_j=\frac{2\pi}{(t_N-t_1)\text{ofac}}j,\quad\text{with } j=1,2,3,..,\frac{N\text{ofac}}{2}. \tag{14}$$

There are other small details, and the reader is referred to our Matlab program lspr.m which is an extension of the program period.f (Press et al., 1992) and which is provided as auxiliary material of the present study.

2.3 Reconstruction

Once the complex Fourier spectrum is in the right format, the data analysis is quite simple. The amplitude spectrum can be easily analysed and modified. For example, all spectral components with amplitudes below a certain confidence limit can be set to zero, and/or spectral components within certain frequency bands can be easily removed. Thus, there are many possibilities for modifications of the time series in the frequency domain, and several examples are shown in the next section. The inverse Fourier transform of the modified Fourier spectrum F_m quickly provides the result in the time domain (Fig. 1). The real part of $F^{-1}(F_m)$ is an evenly spaced time series with reduced noise and composed by the selected spectral components. We have not investigated yet, but it is likely, that the phase information could be utilized for modification of the Fourier spectrum. For example, the phase spectrum may contain valuable information on phase coupling of atmospheric oscillations.

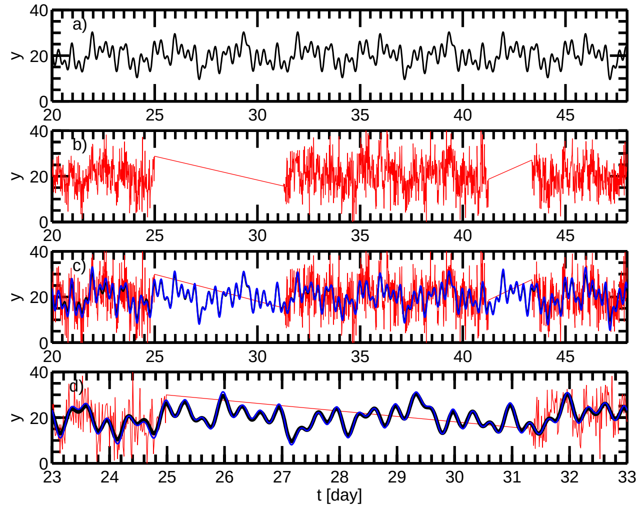

Fig. 2. Test of the reconstruction method with artificial data: **(a)** Artificial series consisting of 5 superposed sine waves (black); **(b)** Artificial series of a) plus noise and two data gaps (red); **(c)** Blue line is reconstructed from the informations of the red line. The reconstructed series is in agreement with the the true signal curve (black line, almost hidden by the blue line); **(d)** Zoom of c) for a better comparison of the true signal (black) and the reconstructed signal (blue).

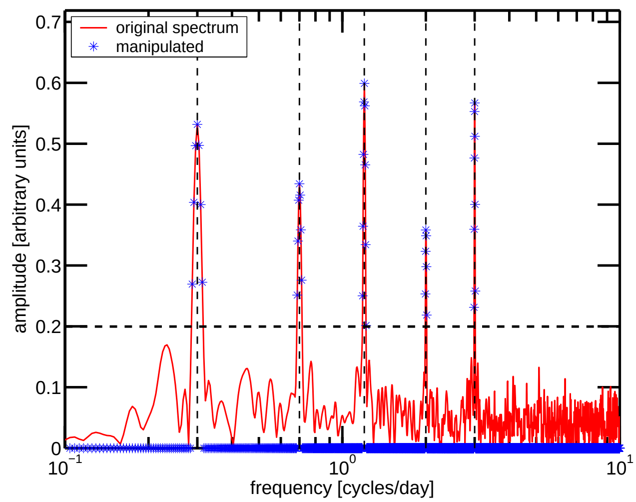

Fig. 3. The original amplitude spectrum of $F(\omega)$ (red) is derived from the Lomb-Scargle periodogram of the noisy series $y(t)$ in Fig. 2b. The modified Fourier spectrum $F_m(\omega)$ is shown by blue symbols (amplitudes below the horizontal, dashed black line are set to zero). Vertical dashed lines indicate the true frequencies of the 5 sine waves of the artificial series $y(t)$ of Fig. 2a. Inverse fast Fourier transform of $F_m(\omega)$ provides the smooth and gap-free series which are shown by blue lines in Fig. 2c and 2d.

3 Results

3.1 Test with synthetic data

The periodic signal within the synthetic data series consist of 5 superposed sine waves with frequencies from 0.3 to 3 cycles/day. The time interval is 60 days long, the sampling time is 20 min, and the number of data points is 3718. A segment of the signal curve is shown in Fig. 2a. Random noise is added to the signal curve. The random noise is by a factor of 2.5 larger than the signals. The combined series of periodic signals, random noise, and data gaps are depicted in Fig. 2b. The series will be used for a test of data gap filling by the Lomb-Scargle periodogram.

For preparation, we subtract the arithmetic mean from the series of Fig. 2b and multiply the series with a Hamming window. Then we apply the data analysis as described by Fig. 1 and obtain the blue curve in Fig. 2c which is almost identical with the true signal curve (black line in Fig. 2c). For the sake of a better comparison of the blue and the black line, Fig. 2d shows a zoom of Fig. 2c. Our data analysis successfully reduced the noise of the red line in Fig. 2b which was the input series and correctly filled the large data gaps. At the edges of the data segment, the reconstructed values are usually of minor quality because of the Hamming window. A workaround is to enhance the window size and to use only the middle part of the reconstructed data segment.

The modification of the Fourier spectrum (Fig. 1) is not mentioned by Scargle (1989) and is described in more detail by the amplitude spectra of Fig. 3. The red line is the amplitude spectrum as obtained from the red time series of Fig. 2b by means of the Lomb-Scargle periodogram. The blue symbols denote the modified spectrum which is used for the inverse fast Fourier transform providing the blue line of Fig. 2c. All spectral components with amplitudes smaller than the dashed horizontal line of Fig. 3 are removed by setting their amplitudes to zero.

In case of geophysical data, confidence limits can be utilized for separation between relevant spectral components and noise. The relevant spectral components have to exceed the confidence limit which is selected with respect to the mean power of the spectrum. As already mentioned, more sophisticated methods of spectrum modification are also thinkable. In the following we give an example with real data, showing the consequence if the spectrum is not modified.

3.2 Test with observational data of lower mesospheric ozone

3.2.1 SOMORA Ozone Microwave Radiometer

The stratospheric ozone monitoring radiometer (SOMORA) monitors the radiation of the thermal emission of ozone at 142.175 GHz. SOMORA has been developed at the Institute of Applied Physics, University Bern. The instrument was

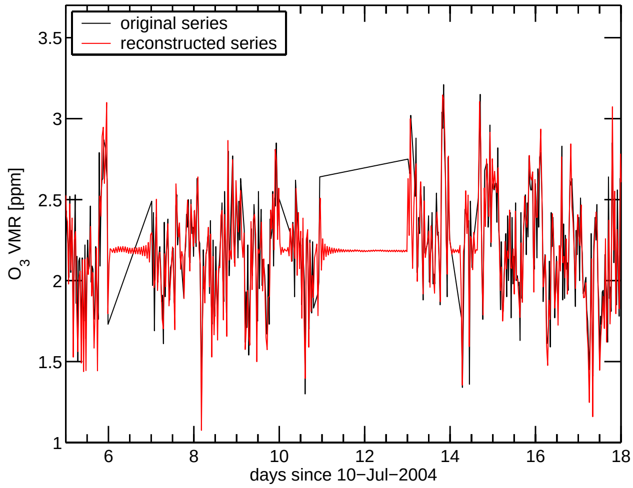

Fig. 4. Test of the reconstruction method with data of the ground-based microwave radiometer SOMORA in Payerne (Switzerland): Time series of O_3 volume mixing ratio at h=52 km which are shown by the black line. The red line is the result if the inverse Fourier transform is applied to the complete spectrum $F(\omega)$ without modification. The spectrum is shown in Fig. 5.

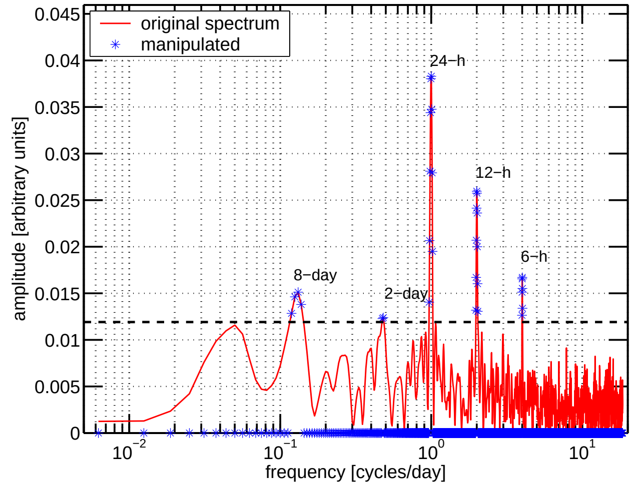

Fig. 5. Complete amplitude spectrum of $F(\omega)$ (red) and modified spectrum of $F_m(\omega)$ (blue) only considering values above the 98% confidence limit which is shown by the horizontal dashed line. The spectra are derived from the O_3 series of Fig. 4 by means of the Lomb-Scargle periodogram. The result of inverse Fourier transform of the complete spectrum $F(\omega)$ is shown in Fig. 4 by the red line. The result of inverse Fourier transform of the modified spectrum $F_m(\omega)$ is shown in Fig. 6 by the red line. It is better to take the modified spectrum $F_m(\omega)$ for reconstruction of the ozone series.

first put into operation on 1 January 2000 and was operated in Bern (46.95 N, 7.44 E) until May 2002. In June 2002, the instrument was moved to Payerne (46.82 N, 6.95 E) where its operation has been taken over by MeteoSwiss. SOMORA contributes primary data to the Network for Detection of Atmospheric Composition Change (NDACC).

The vertical distribution of ozone is retrieved from the recorded pressure-broadened ozone emission spectra by means of the optimal estimation method (Rodgers, 1976). A radiative transfer model is used to compute the expected ozone emission spectrum at the ground. Beyond 45 km altitude, the model atmosphere is based on monthly averages of temperature and pressure of the CIRA-86 climatology, adjusted to actual ground values using the hydrostatic equilibrium equation. Either a winter or summer CIRA-86 ozone profile is taken as a priori. The retrieved profiles minimize a cost function that includes terms in both the measurement and state space. The altitude h is fixed in both the forward and inverse models, and altitude-dependent information is extracted from the measured spectra by reference to the used pressure profile. SOMORA retrieves the volume mixing ratio of ozone with less than 20% a priori contribution in the 25 to 65 km altitude range, with a vertical resolution of 8–10 km, and a time resolution (spectra integration time) of ∼30 min. More details concerning the instrument design, data retrieval, intercomparisons, and applications can be found in Calisesi (2000), Calisesi (2003), Calisesi et al. (2005), Hocke et al. (2006), and Hocke et al. (2007). Here, we use SOMORA measurements of ozone volume mixing ratio in the lower mesosphere at h=52 km having a time resolution of about 30 min.

3.2.2 Lower mesospheric ozone in summer and winter

The data segment has a length of 40 days, the number of data points is 1479, and and the start of the series is on 10 July 2004. The lower mesospheric ozone series of the SOMORA radiometer are depicted by the black line in Fig. 4. Two major gaps appear around day 6.5 and day 12. If we apply the data analysis (Fig. 1) without the modification of the spectrum, the red line is retrieved which well captures the original series. However the noise is not reduced and the gaps are just filled by the arithmetic time-mean. This is certainly not desired but the example gives us some insight into the Lomb-Scargle periodogram method. The variance of the evenly sampled series (red line) minimizes at the locations of the data gaps.

Reduction of the spectrum to its dominant spectral components seems to be necessary for an improved filling of the data gaps. Thus all spectral components with amplitudes smaller than the 98%-confidence limit (dashed horizontal line) are set to zero in Fig. 5 where the full spectrum is shown by the red line, and the modified spectrum is given by the blue symbols. The confidence level of 98% corresponds to the mean power of the background spectrum multiplied by the factor 2.3 (given by the inverse of the normal cumulative distribution function). Thus the possibility that the selected spectral components belong to the background spectrum (noise) is less than 2%.

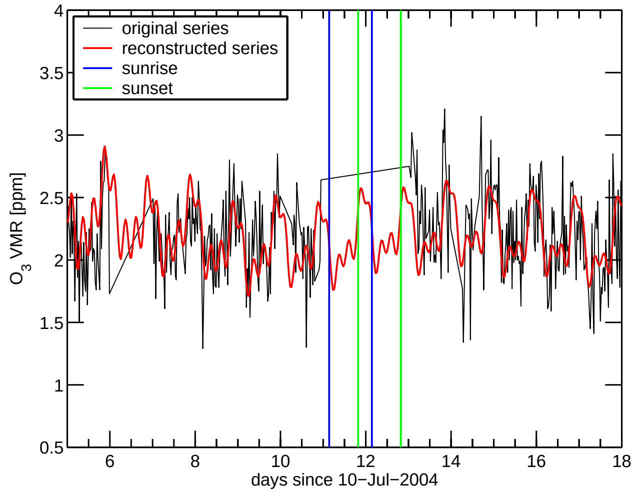

Fig. 6. Comparison of the O_3 series (black) with the reconstructed series (red) using the modified spectrum of Fig. 5. The reconstruction method provides good predictions or estimates of O_3 VMR within the gaps of the original series, e.g., large gap around day 12. The daily variation of lower mesospheric ozone (h=52 km)]is obviously well described by the variation of the reconstruced series (red): Rapid increase of ozone after sunset (time of sunset is indicated by the green vertical line) and rapid decrease of ozone after sunrise (blue vertical line).

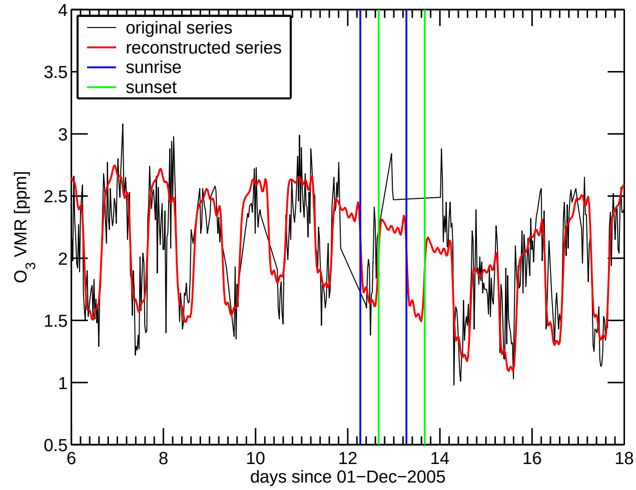

Fig. 7. Same as Fig. 6 but for measurements in winter.

Application of the fast Fourier transform to the modified spectrum yields the series of Fig. 6. The red line is the reconstructed series which has been retrieved from the black, original series. The data gap around day 12 is well filled. The time points of sunrise and sunset are independently determined and are marked by blue and green vertical lines respectively. Due to fast recombination of atomic and molecular oxygen after sunset, the ozone volume mixing ratio is larger during nighttime. The reconstructed series well captures the regular daily variation. Additionally a planetary wave-like oscillation with a period of about 8 days is also present in the reconstructed series of Fig. 6. If we go back to the spectrum in Fig. 5, we see spectral components due to planetary wave-like oscillations (e.g., 8-day, 2-day) as well as the diurnal, semidiurnal, and 6-h components. Because of these periodic signals, the reconstruction by spectral methods yields good results in Fig. 6.

The reconstruction method is also applied to lower mesospheric ozone observed at h=52 km in winter (December 2005) by the SOMORA radiometer. The daily variation of the red curve in Fig. 7 is different to Fig. 6, since the nights are longer in winter. Additionally, planetary wave activity is usually stronger in winter. However the red reconstructed series of Fig. 7 well represent the daily variations and the planetary wave-like oscillations in winter, too. The major gap around day 13 is well filled by the Lomb-Scargle reconstruction method. We like to emphasize that the reduction of noise and the filling of data gaps are valuable for a better recognition of basic processes such as the impact of sunrise and sunset on lower mesospheric ozone. A good estimate of the regular daily variation of lower mesospheric ozone with a high temporal resolution is needed for analysis of gravity wave-induced short-term variance of ozone (Hocke et al., 2006). Digital filtering of the ozone series in the time domain cannot correctly distinguish between sudden changes of the ozone distribution around the solar terminator and gravity wave-induced short-term fluctuations of ozone. A reconstruction by means of spectral methods can solve this problem (Fig. 7).

4 Need of gap filling methods for climate change studies

4.1 Limitations of the Lomb-Scargle reconstruction method

As shown in the previous section the Lomb-Scargle reconstruction method gives convincing results if the data series mainly consist of periodic oscillations. In case of ozone series of the lower stratosphere, the contributions of nonperiodic fluctuations are much stronger, and it is not clear if the application of the Lomb-Scargle reconstruction method is useful. The literature on the Lomb-Scargle periodogram mainly investigates the problem of assessment of the “false alarm probability” (Baluev, 2008; Horne and Baliunas, 1986). Gaps and nonperiodic fluctuations of the data series can induce false spectral components in the Lomb-Scargle periodogram which are often misinterpreted as geophysical signals. The Lomb-Scargle periodogram often overestimates the high-frequency components since these spectral components minimize the variance within the gap intervals (Fig. 4).

Hernandez (1999) found that the conventional Lomb-Scargle significance test gives lower confidence levels than the Schuster-Walker and Fisher tests. Scargle (1989) assumed that there may be kinds of data which cause difficulties and advised the reader to use caution in application of the Lomb-Scargle periodogram, especially in cases of unusual sampling. A simple recipe which is valid for all data series appears impossible. The solution for gap filling of a long-term series is probably to switch between various methods (e.g., linear interpolation, filtering, spectral methods) and to use variable data window lengths depending on the gap size distribution and the oscillation periods. Iterative procedures might be also useful for optimal filling of data gaps within a long-term series, and non-fixed basis alternatives are considerable (Kondrashov and Ghil, 2006). At this point, the question arises if such efforts are necessary and if gap filling is really required for climate change studies? This question will be partly answered in the following.

4.2 Trend analysis of climatic series with and without gap filling

The precise estimation of the mean states of atmospheric, oceanic and other parameters is the precondition for the correct determination of temporal trends which is required in any kind of climate change study (Schneider, 2001). Data gaps occur in many climatic series. The distribution of data gaps is usually not stochastic but depends for example on weather conditions and human-induced factors such as vacation time of the operator of the measurement instrument. Thus data gaps can introduce systematic biases into the data series. In addition, quality control checks of the data retrieval can enhance the amount of systematic data gaps. Presently, most trend studies are based on multiple linear regression analyses which are fitting a linear trend and various sine waves to unevenly spaced data series. The hope is that the data gaps play no role or that the gaps might be automatically corrected by the multiple regression analysis but this is certainly not the case. Iteration of the multiple linear regression would be a workaround where the data gaps are filled by means of the fitted regression curve. The repeated application of multiple linear regression to the gap-free series will give an improved estimate of the linear trend.

For investigation of the influence of data gaps on the linear trend, the long-term recordings of daily mean temperature and water vapor pressure at Jungfraujoch station (z=3580 m) in Switzerland are analyzed. Both data series of MeteoSwiss are complete, and the “true values” of their linear trends can be determined, unbiased by data gaps. Then the real data gaps of another measurement instrument are introduced into the complete Jungfraujoch series. The gaps are taken from the ground-based ozone microwave radiometer GROMOS at Bern (Dumitru et al., 2006). The completeness of the GROMOS data ranges between 65 and 85% depending on season. The annual distribution of the data completeness is shown in Fig. 8 averaged over the time interval from January 1997 to January 2007. Compared to infrared and optical remote

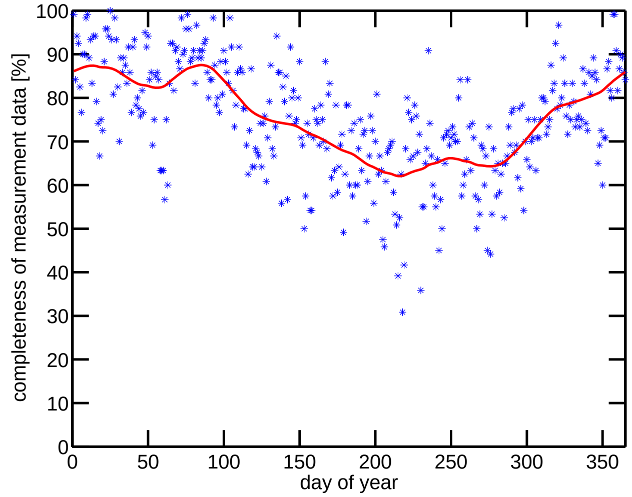

Fig. 8. Completeness of the ozone data of the microwave radiometer GROMOS at Bern as function of day of year and averaged for the time interval from 1997 to 2007. The GROMOS radiometer is almost continuously operated, and the time resolution of the retrieved ozone profiles is 2 h. In case of high water vapor loading of the troposphere, the retrieval rejects the ozone spectrum, and a data gap occurs. The blue symbols indicate the daily values of completeness while the red line shows the monthly values of completeness.

sensing techniques, ozone microwave radiometers have “all weather capability”. However the attenuation of the stratospheric ozone emission by absorption through tropospheric water vapor leads to a reduced data quality of the ozone line spectra during times of high water vapor loading of the atmosphere. The retrieval algorithm of GROMOS rejects more spectra during humid summers than during dry winters.

After inserting the data gaps into the temperature time series, gap filling methods can be tested with the advantage that the true value of the linear trend is known. This true value can be recalculated if the data gaps have been filled in a right manner. The result is shown in Fig. 9. The red line indicates the true linear trend of the original complete temperature series which is around 1 Kelvin per decade at Jungfraujoch station from January 1997 to January 2007 (influences from interannual and decadal oscillations of the temperature are not considered here). The black line indicates the linear trend which has been calculated from the incomplete temperature series (with data gaps from GROMOS). An underestimation of about 30% is present. The green line shows the trend value which is calculated when the gaps have been filled by linear interpolation. The underestimation of the trend is now about 10%. Gap filling with linear interpolation always gave a better result than “no gap filling” though the latter is often applied in climate trend studies.

Application of the Lomb-Scargle reconstruction method is a bit complicated and requires a longer description. The length of the moving window is selected as 0.75 year for the T series and 0.5 year for the e series (a smaller window

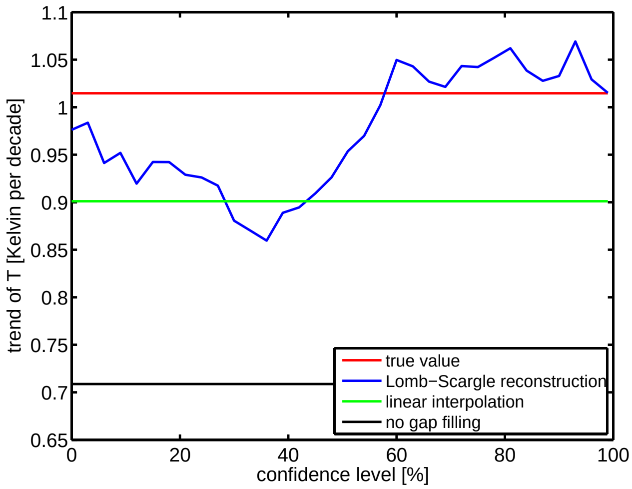

Fig. 9. Linear trend of the temperature at Jungfraujoch station (MeteoSwiss) for the time from January 1997 to January 2007. The red line is the true value obtained for the complete series. The black line denotes the linear trend which is calculated after the data gaps of the GROMOS microwave radiometer (Fig. 8) were inserted into the temperature series. The green line shows the linear trend which is obtained when the data gaps have been filled by linear interpolation. The blue curve shows the linear trend as function of the confidence level when the data gaps are filled by means of the Lomb-Scargle reconstruction method.

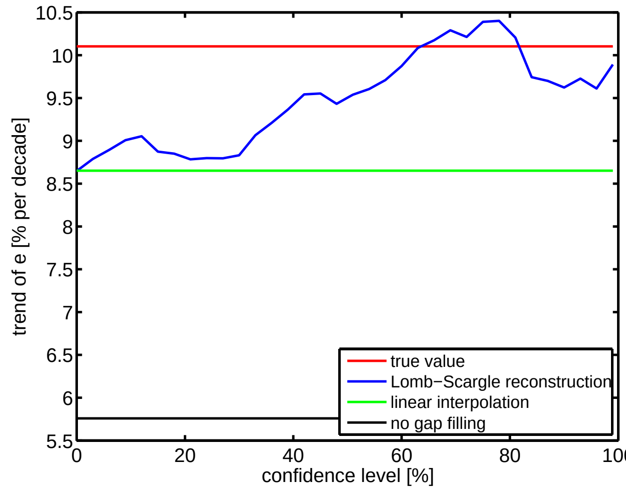

Fig. 10. Same as Fig. 9 but for the linear trend of the water vapor pressure e at Jungfraujoch. The estimated trends agree better with the true value (red line) when the data gaps have been filled before the trend is calculated.

size is also possible). Before calculation of the periodogram, the linear trend and the mean of the data segment are subtracted, and the data segment is multiplied with a Hamming window. Then the confidence level is selected for the modification of the spectrum, and the inverse Fourier transform is solely applied to the spectral components which exceed the confidence level. Linear trend and the mean are added, and the middle part of the data segment is stored in the series of the reconstructed values. Then the data window is moved a step forward, and the same procedure is repeated. Finally a reconstructed series has been generated for the interval from 1997 to 2007. Now the data gaps of the temperature series can be filled by using the reconstructed series (only at the places of the data gaps).

We performed this analysis for various confidence levels and obtained the blue line in Fig. 9 which is better than ”simple linear interpolation”. The Lomb-Scargle reconstruction method offers more possibilities, e.g., variation of the data window size and the confidence level. The problem is that the optimal choice of the reconstruction parameters (window size, confidence level, ...) depends on the individual time series and even on the data segment within the selected time series. So a more adaptive algorithm is desirable in order to ensure a robust reconstruction for all possible cases. However the advantage of the Lomb-Scargle reconstruction method compared to “linear interpolation” is small. In addition linear interpolation of values from a low-pass filtered series as done by Hocke and Kämpfer (2008) may compensate the small difference. Nevertheless the application of the Lomb-Scargle reconstruction method to a long-term series with non-periodic and periodic fluctuations is feasible as the examples have shown.

The time series of water vapor pressure e at Jungfraujoch station is the second test, and the results are shown in Fig. 10. From 1997 to 2007, the water vapor pressure increased by about 10%. Assuming conservation of relative humidity and using the Clausius-Clapeyron relationship, an increase of water vapour pressure of about 6–7% would be expected for a temperature change of 1 Kelvin as it is observed at Jungfraujoch station from 1997 to 2007 (Bengtsson et al., 2004). Again, the black line (ignoration of data gaps) yields the worst result: underestimation of the trend by about 40%. Gap filling by linear interpolation gives a small underestimation of the trend by 15%, and the Lomb-Scargle reconstruction gives even better agreements with the true value (with exception of the 100% confidence level).

5 Conclusions

The reconstruction of unevenly sampled data series with data gaps has been performed by means of the Lomb-Scargle periodogram, with subsequent modification of the Fourier spectrum, and inverse fast Fourier transform. This reconstruction method is reasonable for data series with various periodic signals. Compared to reconstruction methods in the time domain, the Lomb-Scargle reconstruction provides the amplitude and phase spectrum which are utilized for the modification of the time series in the spectral domain. Thus the data user has analytical support by the spectrum and a high degree of freedom for modifications, adjustable to specific data sets and retrieval purposes.

The reconstruction method was tested with synthetic data series and ground-based measurements of lower mesospheric ozone having a strong diurnal variation. Further the reconstruction method was tested by means of long-term series of temperature and water vapor pressure at Jungfraujoch station where data gaps from another instrument were inserted. These series contain non-periodic and periodic fluctuations. The test showed that the error of the linear trend estimation is reduced when the data gaps are filled by linear interpolation or by the Lomb-Scargle reconstruction method before the trend is calculated. The choice of "no gap filling" was the worst choice for determination of the linear trends since climate observations have usually periodic variations and the data gap distribution is usually non-stochastic. "No gap filling" was only a good choice in case of artificial time series with random fluctuations.

The Lomb-Scargle periodogram is the basic element of the reconstruction method. Recently the Lomb-Scargle periodogram has been derived from the Bayesian probability theory, and the method of the Lomb-Scargle periodogram has been expanded for the nonsinusoidal case (Bretthorst, 2001a,b). We expect that the importance of data reconstruction methods for climate change research will increase since observational networks, historical measurement series, and present measurement techniques have non-stochastic data gaps which can introduce considerable errors in the trend estimation.

However the algorithm of the Lomb-Scargle reconstruction method as decribed here is not robust and reliable enough for a serious application. We encountered several problems and were not able to reject all doubts of the reviewers (please see reviewer comments in ACPD). Particularly the selection of reconstruction parameters such as data window size and confidence level is too arbitrary yet. The selection should be more adaptive so that characteristics of individual measurement series can be optimally considered by the algorithm. Combination of the Lomb-Scargle reconstruction with other gap filling methods is also thinkable. Further ideas and work are needed before a solid algorithm (or software toolbox) can be developed. This algorithm would be the basis for a detailed statistical validation of the reconstruction method.

Acknowledgements. We are grateful to the Swiss Global Atmosphere Watch program (GAW-CH) for support of the study within project SHOMING. The ozone radiometer SOMORA is operated by MeteoSwiss in Payerne. The temperature and water vapor pressure series at Jungfraujoch are provided by CLIMAP-net of MeteoSwiss. We provide preliminary Matlab programs for demonstration and discussion of the Lomb-Scargle reconstruction http://www.atmos-chem-phys.net/9/4197/2009/acp-9-4197-2009-supplement.zip. We thank the reviewers for their interest, constructive comments, and improvements.

Edited by: W. Ward

References

Adorf, H.-M.: Interpolation of irregularly sampled data series – a survey, in: Astronomical data analysis software and systems IV, edited by: Shaw, R. E., Payne, H. E., and Hayes, J. J. E., ASP conference series, 77, 1–4, 1995.

Altadill, D., Apostolev, E. M., Jacobi, C., and Mitchell, N. J.: Six-day westward propagating wave in the maximum electron density of the ionosphere, Ann. Geophys., 21, 1577–1588, 2003, http://www.ann-geophys.net/21/1577/2003/.

Baluev, R. V.: Assessing the statistical significance of periodogram peaks, Mon. Not. R. Astron. Soc., 385, 1279–1285, 2008.

Bengtsson, L., Hagemann, S., and Hodges, K. I.: Can climate trends be calculated from reanalysis data?, J. Geophys. Res., 109, D11111, doi:10.1029/2004JD004536, 2004.

Bretthorst, G.: Generalizing the Lomb-Scargle periodogram, in: Maximum Entropy and Bayesian Methods in Science and Engineering, edited by: Mohammad-Djafari, A., AIP conference proceedings, USA, 2001a.

Bretthorst, G.: Generalizing the Lomb-Scargle periodogram – The nonsinusoidal case, in: Maximum Entropy and Bayesian Methods in Science and Engineering, edited by: Mohammad-Djafari, A., AIP conference proceedings, USA, 2001b.

Calisesi, Y.: Monitoring of stratospheric and mesospheric ozone with a ground-based microwave radiometer: data retrieval, analysis, and applications, Ph.D. thesis, http://www.iap.unibe.ch/publications/, Philosophisch-Naturwissenschaftliche Fakultät, Universität Bern, Bern, Switzerland, 2000.

Calisesi, Y.: The Stratospheric Ozone Monitoring Radiometer SOMORA: NDSC Application Document, IAP Research Report 2003-11, http://www.iap.unibe.ch/publications/, Institut für angewandte Physik, Universität Bern, Bern, Switzerland, 2003.

Calisesi, Y., Soebijanta, V. T., and van Oss, R. O.: Regridding of remote soundings: Formulation and application to ozone profile comparison, J. Geophys. Res., 110, D23306, doi:10.1029/2005JD006122, 2005.

Dumitru, M. C., Hocke, K., Kämpfer, N., and Calisesi, Y.: Comparison and validation studies related to ground-based microwave observations of ozone in the stratosphere and mesosphere, J. Atmos. Solar Terr. Phys., 68, 745–756, 2006.

Ford, E. A. K., Aruliah, A. L., Griffin, E. M., and McWhirter, I.: Thermospheric gravity waves in Fabry-Perot interferometer measurements of the 630.0 nm OI line, Ann. Geophys., 24, 555–566, 2006, http://www.ann-geophys.net/24/555/2006/.

Funatsu, B. M., Claud, C., Keckhut, P., and Hauchecorne, A.: Cross-validation of Advanced Microwave Sounding Unit and lidar for long-term upper-stratospheric temperature monitoring, J. Geophys. Res., 113, D23108, doi:10.1029/2008JD010743, 2008.

Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., Robertson, A. W., Saunders, A., Tian, Y., Varadi, F., and Yiou, P.: Advanced spectral methods for climatic time series, Rev. Geophys., 40, 1003, doi:10.1029/2000RG000092, 2002.

Hall, C. M., Nozawa, S., Meek, C. E., Manson, A. H., and Luo, Y.: Periodicities in energy dissipation rates in the auroral mesosphere/lower thermosphere, Ann. Geophys., 21, 787–796, 2003, http://www.ann-geophys.net/21/787/2003/.

Hernandez, G.: Time series, periodograms, and significance, J. Geophys. Res., 104, 10355–10368, 1999.

Hocke, K.: Phase estimation with the Lomb-Scargle periodogram method, Ann. Geophys., 16, 356–358, 1998, http://www.ann-geophys.net/16/356/1998/.

Hocke, K. and Kämpfer, N.: Bispectral Analysis of the Long-Term Recording of Surface Pressure at Jakarta, J. Geophys. Res., 113, doi:10.1029/2007JD009356, 2008.

Hocke, K., Kämpfer, N., Feist, D. G., Calisesi, Y., Jiang, J. H., and Chabrillat, S.: Temporal variance of lower mesospheric ozone over Switzerland during winter 2000/2001, Geophys. Res. Lett., 33, L09801, doi:10.1029/2005GL025496, 2006.

Hocke, K., Kämpfer, N., Ruffieux, D., Froidevaux, L., Parrish, A., Boyd, I., von Clarmann, T., Steck, T., Timofeyev, Y., Polyakov, A., and Kyrölä, E.: Comparison and synergy of stratospheric ozone measurements by satellite limb sounders and the ground-based microwave radiometer SOMORA, Atmos. Chem. Phys., 7, 4117–4131, 2007, http://www.atmos-chem-phys.net/7/4117/2007/.

Horne, J. H. and Baliunas, S. L.: A prescription for period analysis of unevenly sampled time series, Astrophys. J., 302, 757–763, 1986.

Kondrashov, D. and Ghil, M.: Spatio-temporal filling of missing points in geophysical data sets, Nonlin. Processes Geophys., 13, 151–159, 2006, http://www.nonlin-processes-geophys.net/13/151/2006/.

Lomb, N. R.: Least-squares frequency analysis of unequally spaced data, Astrophys. Space Sci., 39, 447–462, 1976.

Nozawa, S., Brekke, A., Maeda, S., Aso, T., Hall, C. M., Ogawa, Y., Buchert, S. C., Röttger, J., Richmond, A. D., Roble, R., and Fuji, R.: Mean winds, tides, and quasi-2 day wave in the polar lower thermospshere observed in European Incoherent Scatter (EISCAT) 8 day run data in November 2003, J. Geophys. Res., 110, A12309, doi:10.1029/2005JA011128, 2005.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: Numerical recipes in Fortran, Cambridge University Press, Cambridge, USA, 2 edn., 1992.

Rodgers, C. D.: Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation, Rev. Geophys. Space Phys., 14, 609–624, 1976.

Scargle, J. D.: Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data, Astrophys. J., 263, 1982.

Scargle, J. D.: Studies in astronomical time series analysis. III. Fourier transforms, autocorrelation and cross-correlation functions of unevenly spaced data, Astrophys. J., 343, 874–887, 1989.

Schneider, T.: Analysis of incomplete climate data: Estimation of mean values and covariance matrices and interpretation of missing values, J. Climate, 14, 853–671, 2001.