PERGAMON

Computers & Geosciences 25 (1999) 739–750

COMPUTERS & GEOSCIENCES

Spectral analysis with incomplete time series: an example from seismology ☆

Stefan Baisch*, Götz H.R. Bokelmann [1]

Ruhr-Universität Bochum, Institut für Geophysik, Universitätsstr. 150, D-44780, Bochum, Germany

Received 18 November 1998; received in revised form 11 March 1999; accepted 11 March 1999

Abstract

A method for spectral analysis of nonequidistantly spaced time series is presented: the CLEAN algorithm performs an iterative deconvolution of the spectral window in the frequency domain. We demonstrate the capability of the method on synthetic data examples and apply CLEAN to seismological data, in an example where we seek temporal changes in elastic wave velocities. The observed periodic changes of phase differences consist of frequencies, which in principle can be explained by the influence of solid earth tides, but also by other effects with similar periodicities. Only CLEAN enabled us to enlarge the time window over missing data segments until the frequency resolution was accurate enough to rule out solid earth tides as cause for the observed periodic changes. A MATLAB version of the CLEAN algorithm is available from the authors, or from the IAMG server.

Keywords: Missing data; CLEAN algorithm; Fourier transform; Lomb–Scargle normalized periodogram; Temporal variations; Solid earth tides

1. Introduction

A frequent dilemma in spectral analysis is the incompleteness of the data record, either in the form of occasional missing data or as larger gaps. Standard data processing techniques, notably the fast Fourier transformation, require data given on a regular equidistantly spaced grid, thus forcing the analyst to perform an interpolation. Artefacts from such interpolations may be critical or in some cases even dominate the resulting spectra. Standard techniques are even less useful, if the data are per se given on a grid, which is not equidistantly spaced.

This study addresses the computation of spectral information for time series with missing data including the situations of occasional missing data, larger data gaps and nonequidistantly spaced data. This problem has been addressed before in a number of studies, most of them in the astrophysical sciences. Also in geosciences, many problems require a rigorous treatment of the missing data problem. The purpose of this paper is thus to present the missing data problem and illustrate which difficulties can arise by simple interpolation or by an uncritical generalization of the Fourier technique. We illustrate our preferred approach and show an application in seismology,

☆ Code available from http://www.iamg.org/CGEditor/index.htm.

* Corresponding author. Tel.: +49-234-700-7574; fax: +49-234-709-4181.

E-mail address: baisch@geophysik.ruhr-uni-bochum.de (S. Baisch)

[1] Present address: Department of Geophysics, Stanford University, Stanford, CA 94305-2215, USA.

0098-3004/99/$ - see front matter
PII: S0098-3004(99)00026-6

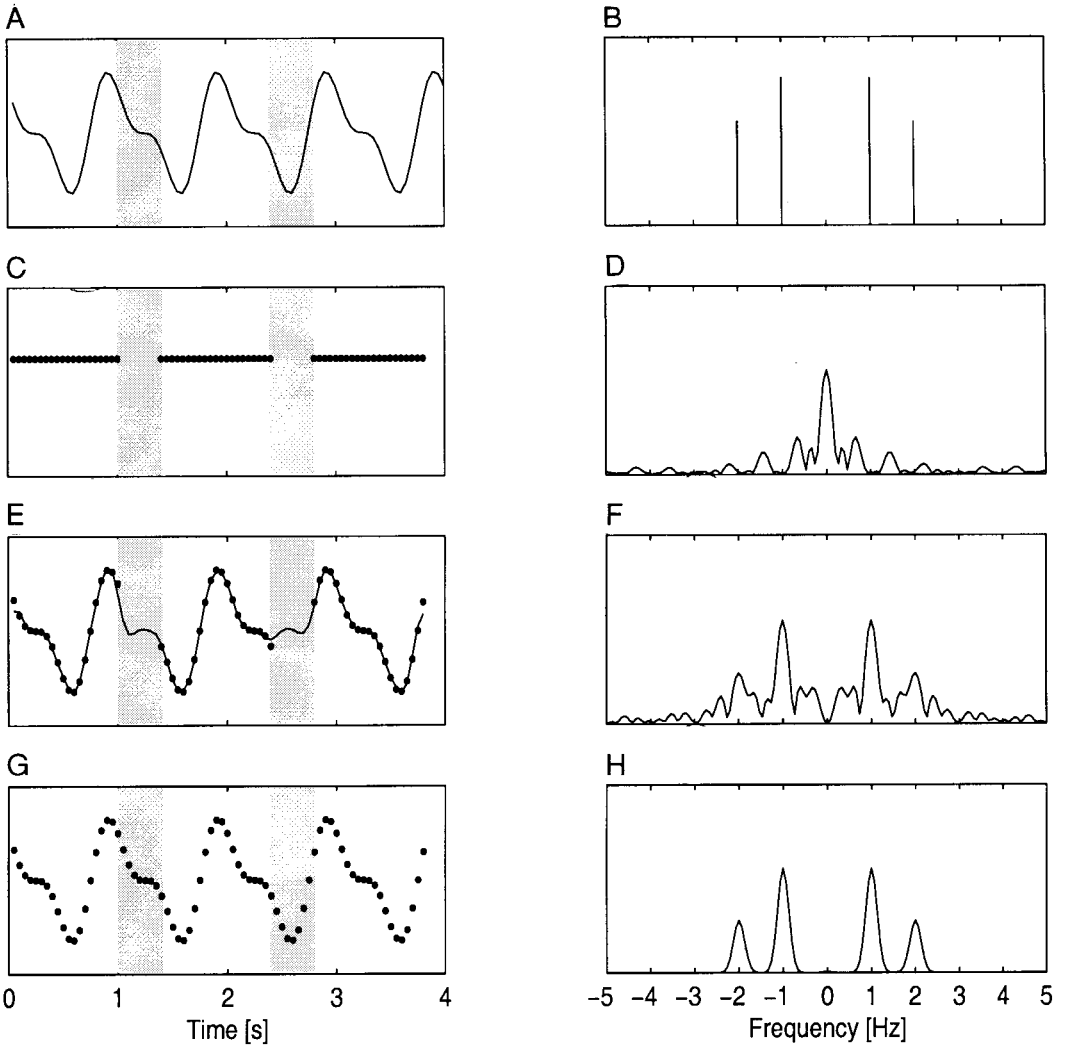

Fig. 1. Illustration of effect of uneven sampling. On left full (continuous) time series (A), sampling function (C) and sampled time series (dots in E), as well as reconstructed data from dirty spectrum (line in E) and from CLEAN spectrum (G) are shown. On right are shown corresponding spectra, of full time series (B), sampling function (D), sampled time series (F) and cleaned spectrum (H). Spectrum of sampled time series (F) corresponds to convolving spectrum of analytical function (B) with spectrum of sampling function (D). Note, that data reconstructed from dirty spectrum closely fit zero values in former gaps (shaded area), whereas reconstruction from CLEAN spectrum fits values of full time series (A) properly.

which could not be solved without a formal treatment of the missing data problem.

A number of methods have been proposed for solving the missing data problem. Our study is closely related to that of Roberts et al. (1987) in using the CLEAN method for spectral analysis in their compact notation. In that method, knowledge of the sampling function is used to perform a stepwise deconvolution in the frequency domain. Although the performance of CLEAN was so far tested only for noiseless data, we obtain good results applying it to noisy data, in synthetic as well as in real data.

The first two sections of the paper state the spectral analysis problem of unevenly spaced discrete data, describe the CLEAN method and give the Roberts et al. (1987) algorithm to compute CLEAN numerically. We illustrate the remarkable stability of the technique in a number of examples and finally apply the technique to real data.

The application, which motivated our study, stems from seismology. We investigate temporal changes of elastic propagation velocities beneath the seismological GERESS array in Bavaria, Germany. We use a continuously emitting source (machine noise) to estimate

relative velocity changes over a time span of 17 days. In this application, the missing data problem arises from the fact that seismic-event-contaminated time windows could not be used for the analysis. Thus, only 33% of the data were used for analysis. The velocities (phase differences) clearly show periodic changes with period in the vicinity of 1/day and 2/day. We were particularly interested whether those changes could be caused by solid earth tides and tried to identify the characteristic spectral component of the lunar tide M2 in the frequency domain (at 1.9324/day). Among several approaches, only CLEAN gave sufficient stability to rule out unambiguously solid earth tides as the cause for the observed periodic changes.

2. Problem statement

We begin with the Fourier transform of a function $f(t)$ given for all times t. It is well-known that if $f(t)$ is a square-integrable continuous function, it can be reconstructed from a spectrum $F(\nu)$

$$f(t) = \int_{-\infty}^{\infty} \mathrm{d}\nu F(\nu) \mathrm{e}^{2\pi \mathrm{i}\nu t}.$$

The spectrum itself is given by the Fourier transform

$$F(\nu) = \mathrm{FT}[\, f(t)] = \int_{-\infty}^{\infty} \mathrm{d}t f(t) \mathrm{e}^{-2\pi \mathrm{i}\nu t}.$$

Consider now the discrete case, where data are given only at N times t_r. These times however are entirely arbitrary. The finite number of data points is

$$f_r = f(t_r) \quad r = 1, \ldots, N.$$

Such a discrete sampling may be introduced into the continuous formulation using a window function (or sampling function) $s(t)$, which may be conveniently defined as

$$s(t) = \frac{1}{N} \sum_{r=1}^{N} \delta(t - t_r)$$

with the Dirac delta function $\delta(t)$.

The sampled signal can then be written as $f_s(t) = f(t)\, s(t)$. The Fourier transform of that sampled signal is

$$F_s(\nu) = \mathrm{FT}[\, f_s(t)] = \mathrm{FT}[\, f(t) s(t)]$$

and from the Fourier convolution theorem

$$= F(\nu) \otimes S(\nu) = \int_{-\infty}^{\infty} \mathrm{d}\nu' F(\nu') S(\nu - \nu') \quad . \tag{1}$$

$S(\nu)$ is suitably termed the 'spectral window function'

$$S(\nu) = \int_{-\infty}^{\infty} \mathrm{d}t s(t) \mathrm{e}^{-2\pi \mathrm{i}\nu t} = \frac{1}{N} \sum_{r=1}^{N} \mathrm{e}^{-2\pi \mathrm{i}\nu t_r}.$$

$F_s(\nu)$ is called the 'dirty spectrum'

$$F_s(\nu) = \int_{-\infty}^{\infty} \mathrm{d}t f_s(t) \mathrm{e}^{-2\pi \mathrm{i}\nu t} = \frac{1}{N} \sum_{r=1}^{N} f_r \mathrm{e}^{-2\pi \mathrm{i}\nu t_r} \tag{2}$$

since it contains the spectral information of the signal, but is contaminated by the convolution with the spectral window function $S(\nu)$. This contamination is illustrated in Fig. 1.

There are different categories of sampling functions $s(t)$. For evenly spaced sampling with a time interval Δt

$$s(t) = \sum_{r=-\infty}^{\infty} \delta(t - r\Delta t) \quad r = -\infty, \ldots, \infty$$

the spectral window function becomes

$$S(\nu) = \frac{1}{\Delta t} \sum_{r=-\infty}^{\infty} \delta\left(\nu - \frac{r}{\Delta t}\right).$$

Periodic sampling in the time domain with interval Δt thus causes multiple peaks in the frequency domain with interval $1/\Delta t$. Introducing the Nyquist-frequency $\nu_N = 1/2\Delta t$, the 'dirty spectrum' is

$$F_s(\nu) = \frac{1}{\Delta t}\left[F(\nu) + \sum_{r=1}^{\infty} [F(\nu - 2r\nu_N) + F(\nu + 2r\nu_N)] \right].$$

If the function $F(\nu)$ is zero for $|\nu| \geq \nu_N$, $f_s(t)$ is fully recoverable from $F_s(\nu)$. This is the well-known sampling theorem. If the condition is satisfied, the spectrum up to ν_N is not contaminated. If the sampling function is a box car

$$s(t) = \begin{cases} 1 & \textit{for } t_0 \leq t \leq t_N \\ 0 & \textit{otherwise} \end{cases}.$$

(finite data length $T = t_N - t_0$), the spectral window function is a sinc-function

$$S(\nu) = \frac{\sin(\pi\nu T)}{\pi\nu} \mathrm{e}^{-\pi \mathrm{i}\nu(t_1 + t_N)}$$

which causes smearing in the spectrum (spectral leakage). The width of smearing (frequency resolution) is controlled by the data length T ($\delta\nu \approx 1/T$). Even in these two basic examples of discrete sampling and finite data-length complications occur, which may lead to difficulties in interpreting the raw (dirty) spectra. Thus the technique discussed in this paper is also use-

ful in applications which do not involve data gaps or arbitrary sampling. The focus however is on those situations that tend to conceal further the true spectrum. They are illustrated later using numerical examples.

To understand better the properties of the dirty spectrum, consider Eq. (2). Note that if data points are missing among f_r, the resulting $F_s(\nu)$ is equivalent to the dirty spectrum of a data set, in which all missing data are identically zero. This implicit assumption is not desirable. Indeed, $f_s(t)$ determined by the inverse Fourier transform of the dirty spectrum $F_s(\nu)$ reflects that implicit assumption (see Fig. 1e): it closely fits zero values at these missing times.

We can remove this effect by eliminating the spectral window function from the dirty spectrum. A straightforward deconvolution in the frequency domain, however, is not possible due to the (mostly zero) nature of the sampling function (Roberts et al., 1987). This problem can be circumvented by estimating the (complex) amplitude of a cosinusoidal, removing its influence on the dirty spectrum including its sidelobes and iterating over this procedure.

3. The CLEAN algorithm

To illustrate how that can be done without invoking a deconvolution we follow Roberts et al. (1987) considering an example of a single harmonic component

$$f(t) = A\cos(2\pi\hat{\nu}t + \Phi) \quad ,$$

with (real) amplitude A, frequency $\hat{\nu}$ and phase Φ. Transformation into the frequency domain yields the spectrum

$$F(\nu) = a\delta(\nu - \hat{\nu}) + a^{\dagger}\delta(\nu + \hat{\nu})$$

using $\dagger$ for complex conjugation and $a = (A/2)\,e^{i\Phi}$ for (complex) amplitude. Now let $f(t)$ be sampled at N discrete times

$$f_s(t) = f(t)_s(t) = \frac{1}{N}\sum_{r=1}^{N} f(t)\delta(t - t_r) \quad .$$

Using the convolution theorem we find $F_s(\nu) = F(\nu) \otimes S(\nu)$.

For the discrete time series the dirty spectrum Eq. (1) becomes

$$F_s(\nu) = aS(\nu - \hat{\nu}) + a^{\dagger}S(\nu + \hat{\nu}) \quad . \tag{3}$$

If $S(0) = 1$, we have at the peak frequency $\hat{\nu}$

$$F(\hat{\nu}) = a + a^{\dagger}S(2\hat{\nu}) \quad .$$

Writing $F^{\dagger}$ similarly and inserting into $a^{\dagger}$ we see that we can determine the amplitude a of the peak, if we know its frequency $\hat{\nu}$

$$a(\hat{\nu}) = \frac{F_s(\hat{\nu}) - F_s^{\dagger}(\hat{\nu})S(2\hat{\nu})}{1 - \|S(2\hat{\nu})\|^2} \tag{4}$$

The idea of the CLEAN formula is to use Eq. (4) to find the (complex) amplitude of a cosinusoidal and remove its contribution to the spectrum F_s including all sidelobes using Eq. (3).

This is done by choosing the largest peak in the dirty spectrum. This procedure is intuitive, but in practice several difficulties occur. First of all we can not exactly determine $\hat{\nu}$ by simply taking the maximum of the dirty spectrum, since the peak in D is smeared by the window spectrum and the interaction of aliases from the positive and negative frequency range. This problem becomes worse if several signals are present. Therefore it is useful to remove only a fraction of its contribution from the dirty spectrum. If this is done iteratively, small errors in $a(\hat{\nu})$ will be corrected in subsequent iterations.

The iteration scheme given by Roberts et al. (1987) is:

1. Compute the dirty spectrum $F_s(\nu)$.
2. Start the iteration with the initial residual spectrum $R^0 \equiv F_s$.
3. On the i'th iteration find the maximum frequency ν_{peak} in the previous residual spectrum R^{i-1} and calculate its complex amplitude $a(\nu_{\text{peak}})$ using Eq. (4).
4. Use Eq. (3) to calculate the contribution of $a(\nu_{\text{peak}})$ to the dirty spectrum and form the residual spectrum R^i by subtracting a fraction g $(0 < g < 1)$ of the result from R^{i-1}:

$$R^i = R^{i-1} - g[(a^i S(\nu - \nu_{\text{peak}}) + (a^i)^{\dagger}S(\nu + \nu_{\text{peak}}))] \quad .$$

   Store the subtracted fraction ga^i to a clean component array at locations $\nu_{\text{peak}>}$ and $-\nu_{\text{peak}}$.

   Continue the iteration until convergence criteria are reached.
5. After the iteration, convolve the clean component array with a Gaussian function to obtain a reasonable frequency resolution. Finally add the residual spectrum of the last iteration.

The iteration ought to proceed until the signal peaks have been successively removed from the residual spectrum. Roberts et al. (1987) named several criteria to define stopping conditions for the noise-free situation. A straightforward method is the predefinition of a threshold value for R^i, so that the iteration stops, if R^i drops below that value. Any contribution below that level is attributed to noise. In practice the definition of

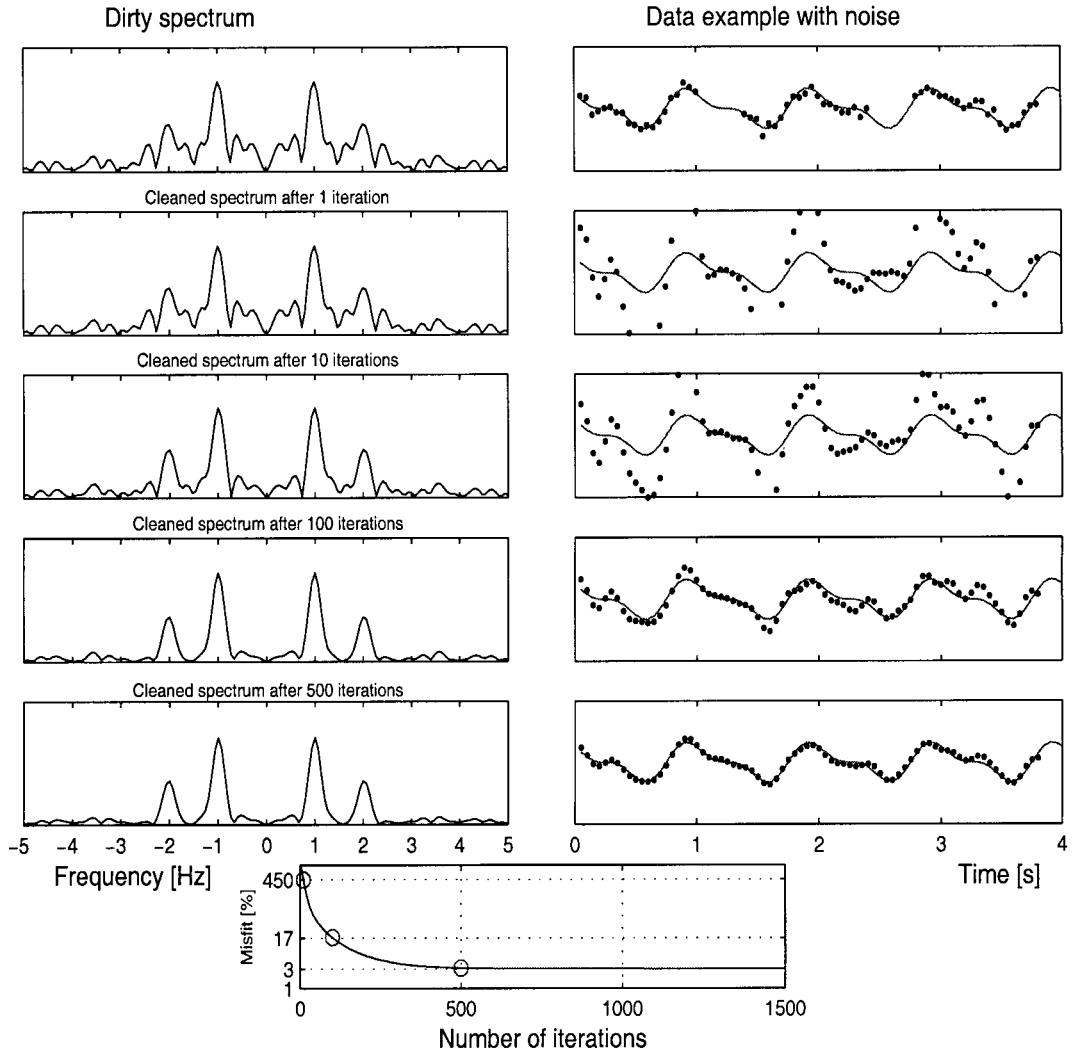

Fig. 2. Data example of Fig. 1 with normally distributed noise added. Standard deviation of noise is 31% that of noiseless data. On left, dirty spectrum and cleaned spectra after 1, 10, 100 and 500 iterations are shown ($g = 0.1$). Right shows data reconstructed from corresponding spectra (dots). Line marks noiseless continuous data. Bottom subplot shows misfit to data depending on number of iteration steps. Circles indicate misfit after 10, 100 and 500 iterations.

the noise level by visual inspection is critical and strongly dependent on the spectral window function.

An automatic but expensive procedure is to base the cutoff criterion on the misfit to the data, which can be computed only in the time domain. However, we feel that it is most convenient to iterate until a predefined maximum number of iteration steps is reached. That number should be chosen large enough to ensure that the iteration does not stop before the influence of all signal peaks has been removed. On the other hand, it appears that the result is insensitive to over-iteration (see the following section).

A limitation of the CLEAN algorithm is, that according to Eq. (4) the spectrum of the sampling function must be known up to twice the frequency of the dirty spectrum. For this reason, cleaning of the dirty spectrum is possible only up to half of the maximum frequency $\nu_{\max}$.

4. Synthetic examples

CLEAN has been tested for the noise-free situation (Roberts et al., 1987) with remarkable success (see also the results in Fig. 1). Fig. 2 now demonstrates the capability of CLEAN for signals buried in noise. The

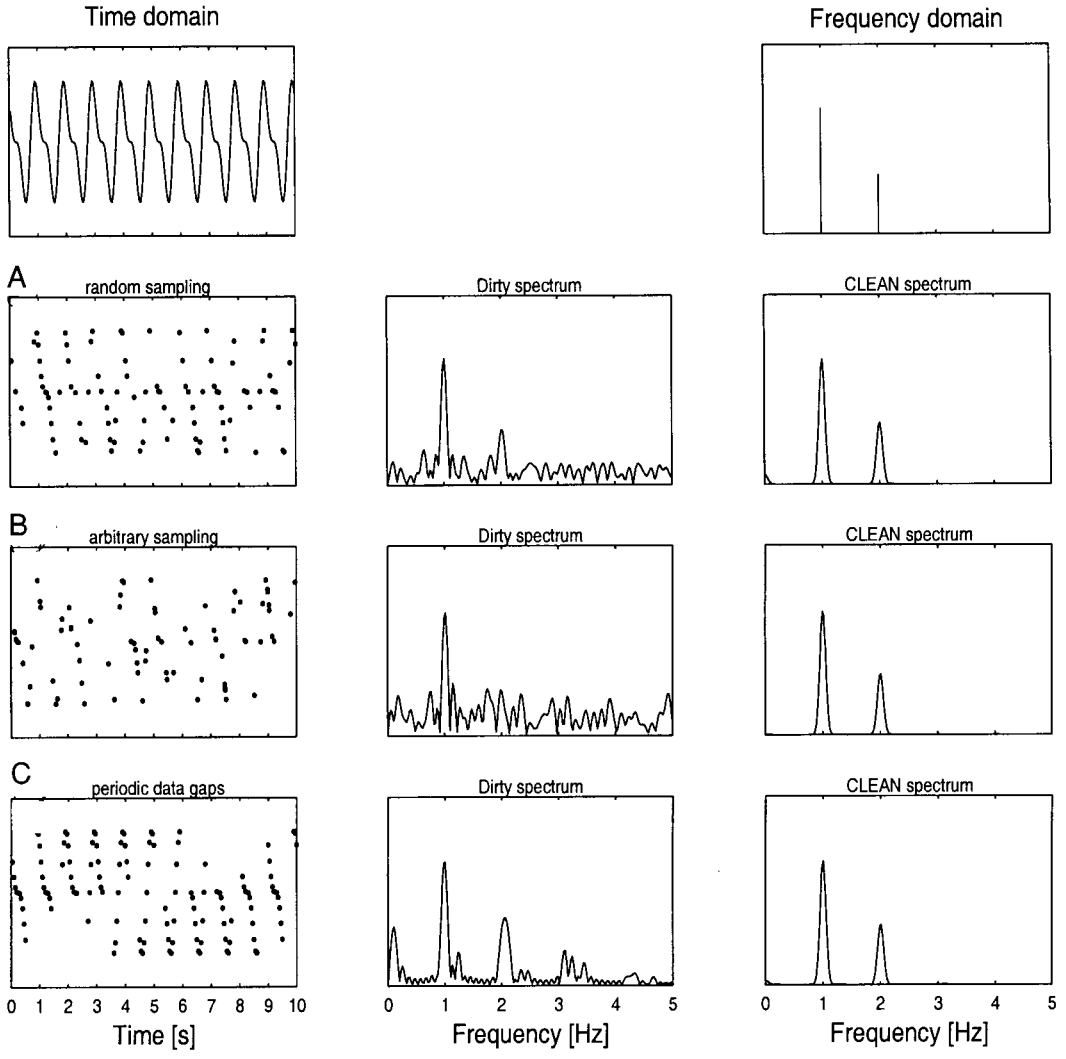

Fig. 3. Application of CLEAN to different categories of missing data: (A) random sampling from regular grid, (B) arbitrary sampling and (C) sampling with periodic data gaps. On left, sampled data in time domain are shown (dots). Corresponding dirty spectra are shown in center subplots and cleaned spectra on right. In all cases, CLEAN successfully recovers spectral information. On top, full time series (left) and its spectrum (right) are shown.

data example is the same as in Fig. 1 with 31% normally distributed noise added. On the left-hand side of Fig. 2 spectra are shown after different numbers of CLEAN iterations. Note that the misfit to the data calculated in the time domain approaches a constant value of 3% which is reached after approximately 400 iteration steps. Therefore it is uncritical if the iteration continues after all signal peaks have been removed from the spectrum.

The next examples demonstrate the performance of CLEAN to different categories of sampling. Harmonic content of the data and parameters for CLEAN (500 iterations with gain factor 0.1) are the same for all examples.

We start with randomly sampled data: in Fig. 3A 94 randomly distributed points were eliminated from an equidistantly spaced grid of 200 points. This leaves the minimum time spacing fixed by the sampling interval. The resulting dirty spectrum shows both spectral components, with the weaker component at 2 Hz nearly covered by the 'noise' of the spectral window. After applying CLEAN, the harmonic content of the data had been correctly unmasked.

In the second example (Fig. 3B) the data had been sampled at 80 entirely arbitrary points. In contrast to the first example the original data grid was not regular. Therefore the maximum frequency, which carries independent information of the time series, is considerably

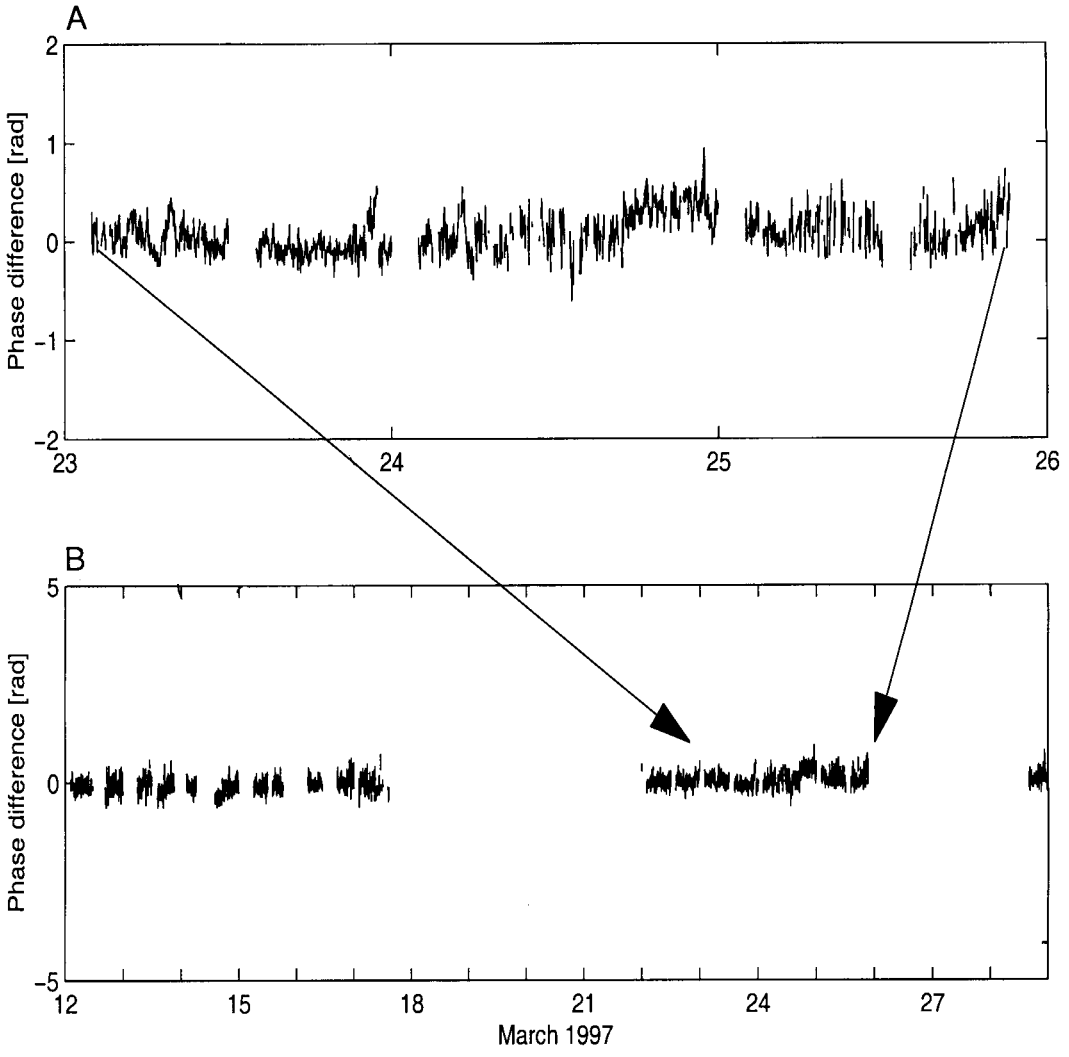

Fig. 4. Estimated phase differences between GERESS stations D3 and D9. (A) shows subwindow with only small data gaps, which is taken from 17 days time series in (B).

larger than in the first example (smaller minimum time spacing) and the computation time increases. The dirty spectrum is noisier than in the first example and the harmonic component at 2 Hz is entirely covered by the 'noise' of the spectral window. Again CLEAN succeeds in recovering the true harmonic component of the data.

The last example deals with the problem of regularly sampled data with periodically distributed data misses. In practice, reasons for periodic data misses may arise from (a) the instrumentation itself, as trigger signals, battery or hard disc changes etc. or (b) from periodically occurring contaminations of the data. The synthetic data in Fig. 3C extract data groups of 7 samples each from the regularly sampled time series. The spacing between the gaps is 11 samples. The dirty spectrum shows the signal peaks at 1 and 2 Hz and artefacts of the spectral window function. An interpretation of the dirty spectrum is difficult, since the spectral window yields distinct peaks, one of them with a similar magnitude than the weaker signal peak. As in the former two examples, CLEAN successfully recovers the spectral information.

5. Applying CLEAN to seismological data

The investigation of temporal changes of the Earth's stress field and the related quantitative understanding of tectonic processes remains one of the most interest-

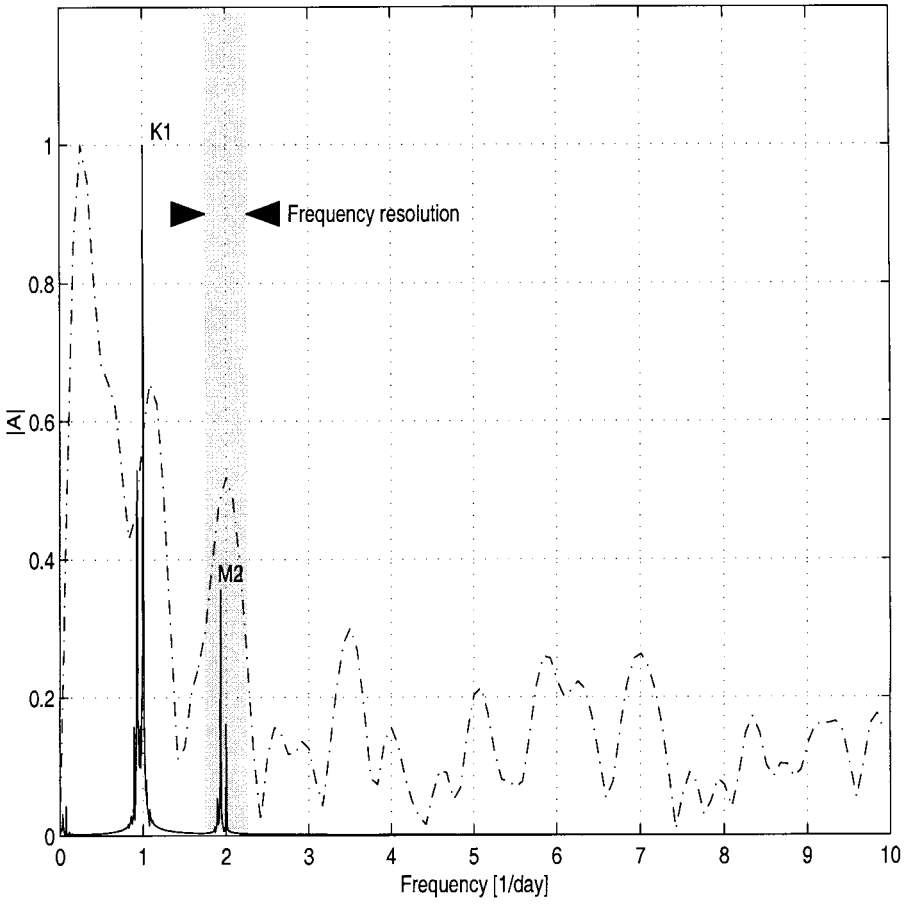

Fig. 5. Spectrum of phase differences calculated for 4-day time window (dash–dotted line). Solid line shows spectrum of predicted solid earth tides with K1 and M2. For comparison, both spectra are normalized to maximum. Shading indicates frequency resolution for phase differences.

ing topics in geophysics. A main drawback of such investigations are the difficulties that occur with the measurement of stress. In situ measurements require boreholes and are limited to the uppermost kilometers of the Earth's crust, whereas an indirect measurement often involves parameters that are only weakly related to stress. Frequently used parameters for stress measurements are seismic velocities, although the functionality of a stress-velocity relation cannot be easily formulated for a complex, heterogeneous medium (e.g. Eisler, 1967; 1969; Aki et al., 1970; Reasenberg and Aki, 1974).

In this study we search for temporal changes of seismic velocities in elastic-wave observations from extremely narrow-band machine sources (see Bokelmann and Baisch, 1999). It has been shown that these waves may propagate observably over hundreds of kilometers in Central Europe. Our main interest is to find out whether our monitoring technique of velocity changes is sensitive enough to render stress induced effects. Therefore we compare observed velocity changes with theoretical changes of the Earth's stress field due to solid earth tides (cf. DeFazio et al., 1973; Bungum, 1977). Our technique comprises monitoring of the relative phase of the seismic signal, recorded by two or more instruments at the Earth's surface, which can be easily related to relative changes in the signal's velocity.

Least-squares estimation of phases (see Appendix A) is easily influenced by the presence of seismic events, which forces us to cut out event-contaminated time windows. Threshold-testing results in a 17-day time series with about 67% of the data missing (Fig. 4B). The parameters for the threshold testing are chosen conservatively for the sake of demonstrating the capability of the CLEAN technique with a challenging data set.

We have chosen a station pair of the GERESS array

Table 1
Strongest solid earth tides for Germany (latitude 48.3°) after Wenzel (1995)

Name	Frequency (1/day)	Amplitude (prediction for rigid earth) (m^2/s^2)
O1	0.9295	0.983
P1	0.9973	0.457
K1	1.0027	1.382
M2	1.9324	1.061
S2	2.0	0.494

with a relative distance of about 3 km (for details about GERESS see Harjes, 1990). Although the phase differences are noisy, one may see indications of periodic changes at low frequencies. However, reliable information about the periodicities requires inspection of the frequency domain. As a start, we picked out the longest, only weakly contaminated data subset of 4 days (Fig. 4A), linearly interpolated the missing data and transformed into frequency domain. Beside a major peak at 0.25/day, the spectrum (Fig. 5) shows two broad peaks in the vicinity of 1/day and 2/day respectively, for which the dominant spectral com-

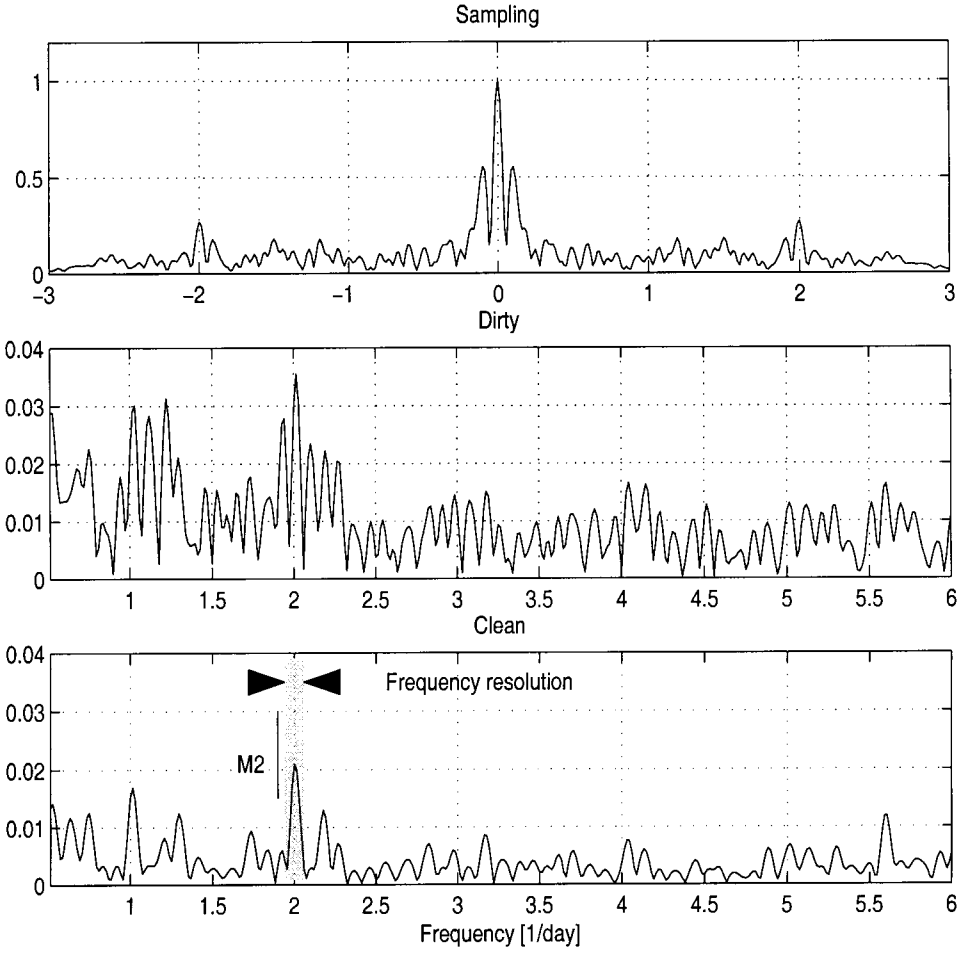

Fig. 6. Spectrum of sampling function for 17-day time window (upper subplot), dirty spectrum of phase differences (middle) and cleaned spectrum of phase differences (lower subplot) after 300 CLEANS with gain factor of 0.1. Note that M2 (short line) does not match 2/day peak within frequency resolution (shaded area).

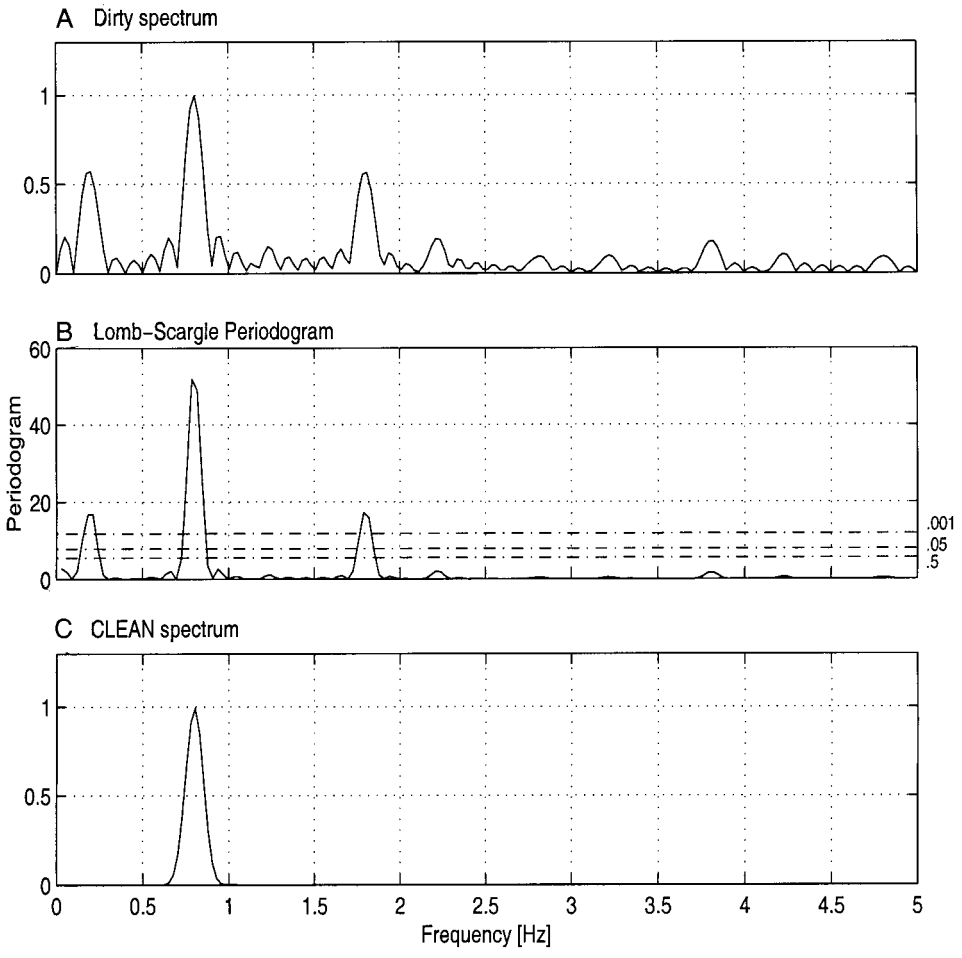

Fig. 7. Comparison of Lomb–Scargle normalized periodogram and CLEAN. (A) shows dirty spectrum of synthetic time series, $\psi = \cos(2\pi 0.8t)$ sampled at 110 points with periodic data gaps. Lomb–Scargle normalized periodogram (B) closely fits dirty spectrum including two sidelobes at 0.2 and 1.8 Hz. In contrast, CLEAN correctly removes sidelobes and fits only true signal at 0.8 Hz (C). Broken lines labeled at right-hand side of (B) denote false alarm probability, e.g. 0.001 stands for false alarm probability of 0.1%.

ponents of the earth tides K1 and M2 lie well within the frequency resolution of $\delta\nu = 0.25$/day (see Table 1).

Unfortunately there are several other periodic effects, which might influence our data in a similar way. Meteorological effects, such as humidity and temperature changes, have spectral energy at exactly 1/day (diurnal) and 2/day (semidiurnal). Also the source function of the seismic signal itself shows periodic changes in these bands, due to the load of the electric power network. In principle, the latter effects should not enter phase differences gathered at a fixed frequency. However, remnant frequency variations might produce some leftover effects of the source phase. To understand the cause of the observed periodic changes, we need to extend the time window for a higher frequency resolution until the characteristic frequency of M2 (1.9324/day) can be separated from that of other effects with 2/day periodicities.

Extending the time window to 17 days would allow that frequency resolution, but it comes along with a considerable increase of missing data (Fig. 4B) and thus a complicated spectrum of the sampling function (Fig. 6A).

Computation of the dirty spectrum yields clusters of peaks around 1/day and 2/day with similar magnitude (Fig. 6B). Using CLEAN the peaks at 1/day and 2/day become much more pronounced and remain as the strongest peaks in the spectrum. With that superior frequency resolution the observed spectrum clearly does not match M2. Instead both peaks show diurnal and semidiurnal frequencies. The observed periodic effects are thus apparently not due to the solid earth

tides; instead one of the other factors mentioned above is the cause.

6. Discussion

For the given data, the decision whether the periodic changes of phase differences are due to elastic velocity changes caused by tidal stresses or not required the application of the CLEAN technique. However, there exist other methods to evaluate the spectral content of nonequidistantly sampled time series. One of them is the Lomb–Scargle normalized periodogram (e.g.Press and Rybicki, 1989; Schulz and Stattegger, 1997), which acts on a per-point instead of a per-time interval basis. In addition, the normalization of the Lomb–Scargle periodogram enables a simple calculation of the significance level of any peak. Especially for arbitrary sampled time series which contain a single harmonic component, the Lomb–Scargle normalized periodogram is a useful algorithm.

In other situations, if the spectral window has distinct sidelobes, the periodogram may lead to misinterpretation, since it fits the dirty spectrum to a certain extent. This can be seen most clearly in examples of regular sampled data with periodic data gaps. Fig. 7A shows the dirty spectrum of a single harmonic sampled with the same function as in Fig. 4C. Any peak beside that at 0.8 Hz is due to artefacts from the spectral window. The Lomb–Scargle normalized periodogram (Fig. 7B) evaluates two of those artefacts as highly significant signal peaks. In contrast, CLEAN properly removes the sidelobes (Fig. 7C) and shows only the signal peak at 0.8 Hz.

For the given data example with it's complex spectral window function (Fig. 6A) the Lomb–Scargle normalized periodogram is not applicable; only CLEAN can be expected to treat properly the missing data. Thus, we emphasize that CLEAN is a powerful tool for the spectral estimation of any finite regularly sampled time series with missing data.

7. Conclusions

We introduced the CLEAN algorithm for spectral analysis of nonevenly spaced time series to the geophysical context, and demonstrated the capability of the algorithm with synthetic examples. In principle, CLEAN can be applied to data of all different kinds of sampling, evenly and unevenly spaced. For evenly spaced data, CLEAN simply removes the artefacts that stem from the finiteness of the time window. But the focus of this paper lies on nonequidistantly sampled data. We distinguish two different forms of sampling, regular sampling with missing data and entirely arbitrary sampling. CLEAN can be applied to both situations, but the latter case may require a considerable increase of computational time. In the example of regularly sampled data with data misses, CLEAN shows a remarkably stable performance. In all examples we tested, CLEAN successfully recovered the spectral information from the dirty spectra. For the situation of a well-known spectral window CLEAN proved to be superior to the Lomb–Scargle normalized periodogram, which closely fitted the dirty spectrum.

The data example from seismology required the treatment with CLEAN to exclude the solid earth tides as a cause for the observed periodic changes in elastic wave velocities.

Acknowledgements

We gratefully acknowledge the Institut für Geophysik and the Bundesanstalt für Geowissenschaften und Rohstoffe for help in assembling the data set for this study. We are particularly grateful to D.H. Roberts for discussion and software. Comments on the manuscript made by K. Stattegger and an anonymous reviewer are greatly appreciated.

Appendix A. Least-squares estimation of amplitudes and phase differences

A monochromatic signal with frequency ν received at coordinates $\vec{x}$ can be described by a seismogram $\psi(t)$ of the form

$$\psi(t) = A\cos(2\pi\nu t - \vec{k}\vec{x} + \varphi) + \eta(t) \equiv A\cos(2\pi\nu t + \Phi) + \eta(t), \quad (A.1)$$

where $\vec{k}$ denotes the wave vector and φ the phase of the signal. $\eta(t)$ is a measure of the 'noise' which includes any process different from the signal. In the following we will take $\Phi = -\vec{k}\vec{x} + \varphi$ as the signal phase.

To solve Eq. (A.1) for phase Φ and amplitude A we rewrite Eq. (A.1) as

$$\psi(t) = A\cos(\Phi)\cos(2\pi\nu t) - A\sin(\Phi)\sin(2\pi\nu t) + \eta(t) \quad (A.2)$$

and obtain a linear equation system

$$\begin{bmatrix} \psi(t) \\ \psi(t+\Delta t) \\ \psi(t+2\Delta t) \\ \vdots \end{bmatrix} = G\vec{m} + \vec{\eta}, \tag{A.3}$$

where

$$G = \begin{bmatrix} \cos(2\pi\nu(t+\Delta t)) & \sin(2\pi\nu(t+\Delta t)) \\ \cos(2\pi\nu(t+2\Delta t)) & \sin(2\pi\nu(t+2\Delta t)) \\ \cos(2\pi\nu(t+3\Delta t)) & \sin(2\pi\nu(t+3\Delta t)) \\ \vdots & \vdots \end{bmatrix}$$

and

$$\vec{m} = \begin{bmatrix} A\cos(\Phi) \\ -A\sin(\Phi) \end{bmatrix}$$

from which we obtain A and Φ.

Note, that Eq. (A.3) assumes the exact knowledge of the signal frequency. In our context, the signal frequency changes slightly with time leading to a phase drift. Therefore we use phase differences between individual receiver pairs, which are independent of the source phase φ.

References

Aki, K., DeFazio, T., Reasenberg, P., Nur, A., 1970. An active experiment with earthquake fault for an estimation of the in situ stress. Bulletin of the Seismological Society of America 60 (4), 1315–1336.

Bokelmann, G.H., Baisch, S., 1999. Nature of narrow-band signals at 2.083 Hz. Bulletin of the Seismological Society of America 89 (1), 156–164.

Bungum, H., 1977. Precise continuous monitoring of seismic velocity variations and their possible connection to solid earth tides. Journal of Geophysical Research 82 (33), 5365–5373.

DeFazio, T.L., Aki, K., Alba, J., 1973. Solid earth tide and observed change in the in situ seismic velocity. Journal of Geophysical Research 78 (8), 1319–1322.

Eisler, J.D., 1967. Investigation of a method for determining stress accumulation at depth. Bulletin of the Seismological Society of America 57 (5), 891–911.

Eisler, J.D., 1969. Investigation of a method for determining stress accumulation at depth-II. Bulletin of the Seismological Society of America 59 (5), 43–58.

Harjes, H.P., 1990. Design and siting of a new regional array in central Europe. Bulletin of the Seismological Society of America 80 (6), 1801–1817.

Press, W.H., Rybicki, G.B., 1989. Fast algorithm for spectral analysis of unevenly sampled data. Astrophysical Journal 338, 277–280.

Reasenberg, P., Aki, K., 1974. A precise, continuous measurement of seismic velocity for monitoring in situ stress. Journal of Geophysical Research 79 (2), 399–406.

Roberts, D.H., Lehár, J., Dreher, J.W., 1987. Time series analysis with CLEAN. I. Derivation of a spectrum. Astronomical Journal 93 (4), 968–989.

Schulz, M., Stattegger, K., 1997. Spectrum: spectral analysis of unevenly spaced paleoclimatic time series. Computers & Geosciences 23 (9), 929–945.

Wenzel, H.G., 1995. Gezeitenpotential. In: DGG-Seminar Gezeiten, Deutsche Geophysikalische Gesellschaft Sonderband II/1995, pp. 1–18.