

Calculation of signal spectrum by means of stochastic inversion

T. Nygrén¹ and Th. Ulich²

¹Department of Physical Sciences, University of Oulu, P.O. Box 3000, 90014, University of Oulu, Finland

²Sodankylä Geophysical Observatory, University of Oulu, 99600, Sodankylä, Finland

Received: 21 January 2010 – Revised: 5 June 2010 – Accepted: 28 June 2010 – Published: 6 July 2010

Abstract. The standard method of calculating the spectrum of a digital signal is based on the Fourier transform, which gives the amplitude and phase spectra at a set of equidistant frequencies from signal samples taken at equal intervals. In this paper a different method based on stochastic inversion is introduced. It does not imply a fixed sampling rate, and therefore it is useful in analysing geophysical signals which may be unequally sampled or may have missing data points. This could not be done by means of Fourier transform without preliminary interpolation. Another feature of the inversion method is that it allows unequal frequency steps in the spectrum, although this property is not needed in practice. The method has a close relation to methods based on least-squares fitting of sinusoidal functions to the signal. However, the number of frequency bins is not limited by the number of signal samples. In Fourier transform this can be achieved by means of additional zero-valued samples, but no such extra samples are used in this method. Finally, if the standard deviation of the samples is known, the method is also able to give error limits to the spectrum. This helps in recognising signal peaks in noisy spectra.

Keywords. Ionosphere (Auroral ionosphere) – Radio science (Signal processing)

1 Introduction

The power or amplitude and phase spectrum of a signal is conventionally calculated by means of the Fourier transform. This can be done, if the signal is sampled at equal intervals. If this is not the case, the Fourier transform can only be applied after resampling the signal by means of some interpolation method (Meisel, 1978). If the lengths of sampling intervals

are not too variable, crude estimates of the spectra can be obtained by applying Fourier transform to the sample sequence as if the sampling interval would be constant.

It is a common situation in geophysics, astronomy or biology that sampling is made at unequal intervals. Since interpolation necessarily introduces some error in the signal, methods have been developed which calculate the spectrum directly from the original samples even in such a case.

A usual approach to calculating the signal spectrum from unequally spaced samples is based on least squares fitting of sinusoidal signals to the data points. Barning (1962) searched for the most important frequency by fitting sinusoidals at various frequencies and choosing the frequency giving the best fit. The next step was to subtract the fitted frequency from the signal and to repeat the search using the difference signal. In such a manner a few frequency components from the light curve of a star were obtained. A key idea of this method is to find the most distinctive frequencies, their phases and amplitudes rather than the phases and amplitudes of the signal components at some pre-selected frequency bins.

A more advanced method was presented by Vaníček (1969) and later developed by Vaníček (1971). This was also based on finding frequency components in the order of importance. In this approach the next frequency is found by fitting simultaneously the amplitudes and phases of the previous frequencies and the trial frequency. The method was simplified by Lomb (1976). It was later shown by Scargle (1982) that Lomb's method is equivalent to a periodogram modified in a certain manner. A fast algorithm for calculating the periodogram was developed by Press and Rybicki (1989) and a program code is available in Press et al. (2007).

It has turned out that, with the computing power available in the present time, there is no need to search for the main frequency components. Instead, one can choose a set of frequencies and make a simultaneous fit giving the amplitude and phase of each frequency component simultaneously.



Correspondence to: T. Nygrén
(tuomo.nygren@oulu.fi)

This formalism has been put in matrix form by Craymer (1998).

In this paper an approach to the problem is made starting from stochastic inversion based on Bayesian theory. The work was originally independent of that by Craymer (1998), but it was later noticed that this approach leads to a similar formalism. It is clear that this must happen, because stochastic inversion with Gaussian errors leads to a generalized least squares solution. A new finding is that the number of frequency bins in the spectrum is not limited in the manner generally believed, but the number of frequency bins can even be greater than the number of signal samples. However, this does not improve the spectral resolution, which is determined by the length of the sample sequence. The effect is rather interpolation of the spectrum just as interpolation is achieved in Fourier transform by adding zero-valued samples to the end of the time series. The difference is that the present method does not need any additional zero-valued samples. If error estimates of data values are known, the method also allows the calculation of error estimates of the power spectrum for all frequencies. This can be done also when subsequent data samples are not independent. Sampling a filtered signal may lead to such a sample sequence.

2 Method

Consider a model of a time signal $x(t)$ in terms of sinusoidal oscillations at m angular frequencies $\omega_k, k = 1, 2, \dots, m$ with different amplitudes and phases, i.e.

$$x(t) = \sum_{k=1}^m (a_k \sin \omega_k t + b_k \cos \omega_k t). \tag{1}$$

By sampling the signal at times $t_j, j = 1, 2, \dots, n$, we obtain a sample sequence $\{x(t_1), x(t_2), \dots, x(t_n)\}$, where

$$x(t_j) = \sum_{k=1}^m (a_k \sin \omega_k t_j + b_k \cos \omega_k t_j) + \varepsilon_j. \tag{2}$$

Here sampling is considered as a measurement of the signal value and ε_j is the true measurement error (not an error estimate). Unlike in Fourier transform, only positive frequencies are considered here and, instead of a complex exponential, the phases of the frequency components are expressed in terms of sines and cosines and their relative amplitudes. Since $x(t)$ may contain a DC component, it is reasonable to choose $\omega_1 = 0$. Then Eq. (2) can be written as

$$x(t_j) = b_1 + \sum_{k=2}^m (a_k s_{jk} + b_k c_{jk}) + \varepsilon_j, \tag{3}$$

where

$$s_{jk} = \sin \omega_k t_j \tag{4}$$

and

$$c_{jk} = \cos \omega_k t_j \tag{5}$$

are coefficients which can be calculated when the angular frequencies and the sampling times are known. The amplitudes a_k and b_k are unknowns which are to be determined.

Equation (3) means that each sample value is a linear combination of the amplitudes of the sines and cosines of all frequencies. The samples can be arranged as an n -component column vector

$$\mathbf{x} = \begin{pmatrix} x(t_1) \\ x(t_2) \\ \vdots \\ x(t_n) \end{pmatrix}, \tag{6}$$

their errors as an error vector

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \tag{7}$$

and the coefficients b_k and a_k as a single $(2m - 1)$ -component column vector

$$\mathbf{X} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \\ a_2 \\ a_3 \\ \vdots \\ a_m \end{pmatrix}. \tag{8}$$

When the coefficients s_{jk} and c_{jk} are arranged as an $n \times (2m - 1)$ matrix

$$\bar{\mathbf{A}} = \begin{pmatrix} 1 & c_{12} & c_{13} & \dots & c_{1m} & s_{12} & s_{13} & \dots & s_{1m} \\ 1 & c_{22} & c_{23} & \dots & c_{2m} & s_{22} & s_{23} & \dots & s_{2m} \\ 1 & c_{32} & c_{33} & \dots & c_{3m} & s_{32} & s_{33} & \dots & s_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & c_{n2} & c_{n3} & \dots & c_{nm} & s_{n2} & s_{n3} & \dots & s_{nm} \end{pmatrix}, \tag{9}$$

all equations (3) can be written as a single matrix equation

$$\mathbf{x} = \bar{\mathbf{A}} \cdot \mathbf{X} + \boldsymbol{\varepsilon}. \tag{10}$$

Equation (10) poses a direct theory of a linear inversion problem. In stochastic inversion, both the measurements \mathbf{x} , the unknowns \mathbf{X} and the errors $\boldsymbol{\varepsilon}$ are taken as random variables. The method is well known (see e.g. Nygrén et al., 1997) but, for clarity, it is briefly reviewed here.

Assuming that the measurement errors are Gaussian random variables with zero mean and a covariance matrix

$$\bar{\boldsymbol{\Sigma}} = \langle \boldsymbol{\varepsilon} \cdot \boldsymbol{\varepsilon}^T \rangle, \tag{11}$$

the conditional probability density of \mathbf{X} is

$$D(\mathbf{X}|\mathbf{x}) \propto \exp\left[-\frac{1}{2}(\bar{\mathbf{A}} \cdot \mathbf{X} - \mathbf{x})^T \cdot \bar{\boldsymbol{\Sigma}}^{-1} \cdot (\bar{\mathbf{A}} \cdot \mathbf{X} - \mathbf{x})\right], \tag{12}$$

where T indicates a transpose. The maximum point of this Gaussian distribution is obtained by minimising the quadratic form $(\bar{\mathbf{A}} \cdot \mathbf{X} - \mathbf{x})^T \cdot \bar{\boldsymbol{\Sigma}}^{-1} \cdot (\bar{\mathbf{A}} \cdot \mathbf{X} - \mathbf{x})$. The quadratic form gets a minimum at

$$\mathbf{X} = \mathbf{X}_0 = \bar{\mathbf{Q}}^{-1} \cdot (\bar{\mathbf{A}}^T \cdot \bar{\boldsymbol{\Sigma}}^{-1}) \cdot \mathbf{x}, \quad (13)$$

where

$$\bar{\mathbf{Q}} = \bar{\mathbf{A}}^T \cdot \bar{\boldsymbol{\Sigma}}^{-1} \cdot \bar{\mathbf{A}} \quad (14)$$

is the Fisher information matrix. In terms of these notations the conditional probability density in Eq. (12) can be written as

$$D(\mathbf{X}|\mathbf{x}) \propto \exp\left[-\frac{1}{2}(\mathbf{X} - \mathbf{X}_0)^T \cdot \bar{\mathbf{Q}} \cdot (\mathbf{X} - \mathbf{X}_0)\right]. \quad (15)$$

This is a standard Gaussian multivariate distribution with a mean \mathbf{X}_0 and a covariance matrix $\bar{\mathbf{Q}}^{-1}$. Hence the statistically most probable values of the unknowns is given by Eq. (13) and their errors are obtained from the covariance matrix

$$\bar{\boldsymbol{\Sigma}}_{\mathbf{X}} = \bar{\mathbf{Q}}^{-1}. \quad (16)$$

If the data samples are statistically independent and their errors have identical variances σ^2 , Eq. (13) is simplified as

$$\mathbf{X}_0 = (\bar{\mathbf{A}}^T \cdot \bar{\mathbf{A}})^{-1} \cdot \bar{\mathbf{A}}^T \cdot \mathbf{x}. \quad (17)$$

and

$$\bar{\boldsymbol{\Sigma}}_{\mathbf{X}} = \sigma^2 (\bar{\mathbf{A}}^T \cdot \bar{\mathbf{A}})^{-1}. \quad (18)$$

If $\bar{\boldsymbol{\Sigma}}$ is diagonal and the variance of ε_i is σ_i^2 , the quadratic form in the exponential in Eq. (12) is

$$\begin{aligned} & (\bar{\mathbf{A}} \cdot \mathbf{X} - \mathbf{x})^T \cdot \bar{\boldsymbol{\Sigma}}^{-1} \cdot (\bar{\mathbf{A}} \cdot \mathbf{X} - \mathbf{x}) = \\ & \sum_i \frac{1}{\sigma_i^2} \left[\left(\sum_j A_{ij} X_j \right) - x_i \right]^2. \end{aligned} \quad (19)$$

Minimising this gives the conventional least squares method where inverses of variances are used as weights. Thus Eq. (13) is a solution of a generalized least squares method which takes into account not only the variances of the measurement errors but also their covariances. This analysis shows the basic idea behind the least squares method: it maximises the conditional probability of the unknowns if the measurement errors have a Gaussian distribution.

Unlike the Fourier transform, the inversion solution contains no relation between the number of data samples and the number of frequencies. By choosing the number of frequencies in such a manner that the number of equations in Eq. (10) is equal to, greater than or smaller than the number of unknowns we obtain three different cases. These are best understood by considering the solutions \mathbf{X} of a general linear equation $\mathbf{x} = \bar{\mathbf{A}} \cdot \mathbf{X}$:

1. If the number of equations is equal to the number of unknowns, $\bar{\mathbf{A}}$ is a square matrix (in our case this corresponds to $(2m - 1) = n$ so that $m = (n + 1)/2$ and n must be odd). Then the matrix equation has a unique solution if $\det(\bar{\mathbf{A}}) \neq 0$. If $\det(\bar{\mathbf{A}}) = 0$, two different cases may appear; either \mathbf{x} is or is not in the range of $\bar{\mathbf{A}}$. In the former case the equation can have more than a single solution, since a linear combination of a given solution and any vector in the null space of $\bar{\mathbf{A}}$ is also a solution. Then it is possible to find the shortest vector \mathbf{X} in the solution space. When \mathbf{x} is not in the range of $\bar{\mathbf{A}}$, the equation does not have a solution. Even then it is possible to find the shortest length of the vector \mathbf{X} of all possible least squares solutions (this is the SVD solution).
2. If the number of equations is smaller than the number of unknowns, the problem is underdetermined (in our case $(2m - 1) > n$ so that $m > (n + 1)/2$). Formally, it is then possible to expand $\bar{\mathbf{A}}$ into a square matrix with additional rows of zeros and also to expand \mathbf{x} with additional zeros. This leads to $\det(\bar{\mathbf{A}}) = 0$, and then one can find either the shortest \mathbf{X} in the solution space or the shortest \mathbf{X} of all possible least squares solutions.
3. If the number of equations is greater than the number of unknowns, the problem is overdetermined, unless $\bar{\mathbf{A}}$ is row-degenerate in such a way that the problem reduces to one of the above cases (in our case $(2m - 1) < n$ so that $m < (n + 1)/2$). An overdetermined set of equations has a solution only in the least squares sense.

The solutions in Eq. (13) or Eq. (17) cover all the above cases as long as $\bar{\mathbf{Q}}^{-1}$ exists. Unless the number of data samples is very great, it is practical to choose such a high frequency resolution that the problem becomes underdetermined. It turns out that this choice has an effect similar to expanding the sample sequence by zero-valued samples when Fourier transform is used; it leads to interpolation of the spectrum.

In conclusion, this formalism gives a compact recipe for determining the most likely values of the unknown amplitudes a_k and b_k in Eq. (2). The Fisher information matrix may be close to singular so that it implies a use of regularization. Then the inverse matrix can be obtained by

$$\bar{\mathbf{Q}}^{-1} \approx (\bar{\mathbf{Q}} + \lambda \bar{\mathbf{I}})^{-1}, \quad (20)$$

where λ is a small number and $\bar{\mathbf{I}}$ is a unit matrix. Actually, it would be possible to expand the inversion method to include some a priori information (see Nygrén et al., 1997), which could also help in regularizing unstable solutions. It seems, however, that this is not needed in practice, and therefore it has been omitted from the theory.

When amplitudes a_k and b_k are known, they can be used for calculating the power of the frequency component ω_k , i.e.

$$P_k = a_k^2 + b_k^2. \quad (21)$$

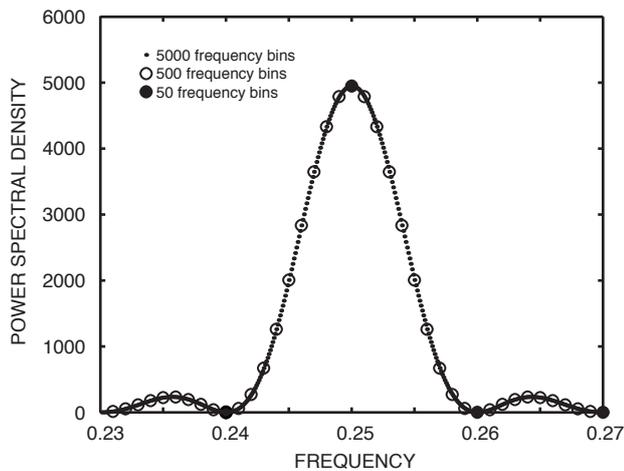


Fig. 1. Power spectrum of a signal at three different resolutions. The frequency of the sinusoidal signal is one fourth of the sampling frequency, and the sample sequence consists of 100 samples. The number of equidistant frequency bins within the frequency interval 0–0.5 is 50, 500 and 5000, respectively.

The amplitude and phase can also be calculated in an obvious way. One should notice that the power spectral density means power per frequency interval, and therefore the power values given by Eq. (21) do not directly give the power spectrum. Consider, for instance, what happens if some frequency grid is first chosen and the inversion is carried out. Then a denser grid is made by adding intermediate frequencies and the inversion is repeated. Now the power (i.e. the amplitudes) at the original frequencies will be smaller than in the first inversion, because a part of the power goes to the new frequencies. This means that, if you simply plot the powers of different frequencies given by Eq. (21) as a function of frequency, the power scale will be different for different frequency grids. However, the proper definition of the power spectrum involves the frequency step in such a way that the energy of the spectrum is equal to the energy of the signal. Here the power of the spectrum is defined as an integral of the power spectral density. Then the power scale of the spectrum will be independent of the frequency step, and we can compare spectra with different frequency grids.

3 Test results

In order to demonstrate that the number of frequency bins can be much larger than the number of samples, the inversion method was tested using a single sinusoidal signal with unit amplitude. The sampling interval was taken as time unit and the frequency of the signal was one fourth of the sampling frequency. A short sequence of only 100 samples was used in calculating the spectrum. Three different equidistant sets of frequencies were chosen, consisting of 50, 500 and 5000 points within the interval 0–0.5. The resulting spectra

are shown in Fig. 1 at a narrow frequency interval around the nominal frequency of 0.25. It is seen that the frequency resolution given by 50 frequency bins give only a single point at the spectrum peak, whereas the other two spectra produce the line shape determined by the length of the sample sequence. The interpolating capability of the method is well demonstrated. The figure also shows that the power scale of the spectrum is independent of the frequency resolution when scaling is made in the way discussed above. This result also shows that, although it would be possible to use non-equidistant frequency bins, there is no real need to that, because the frequency resolution can be increased to fulfill all needs.

For further testing of the method, artificial signals composed of a set of sinusoidal signals at five frequencies were generated. Figure 2 shows results of spectral analysis of signals at frequencies 0.015, 0.11, 0.235, 0.25 and 0.49 with random phases and respective amplitudes 0.5, 0.7, 0.5, 1.0 and 0.8. One hundred samples are taken at equal intervals and the sampling interval is used as a time unit. The top panel shows the signal as a continuous line and the samples are indicated by the markers. The middle panel shows the power spectral densities calculated from all samples by means of Fourier transform. For the Fourier transform, the sample sequence is padded by 990 zero-valued samples to interpolate the spectra to 501 frequencies ranging from 0 to 0.5. The bottom panel shows the power spectral density calculated by the inversion method. However, only the samples shown by the filled dots in the top panel are used by the inversion method. Thus the samples shown by the open dots, which have been randomly chosen, are not used. Neither is any padding by zero values made. The inversion spectrum is calculated at the same 501 frequencies used in the Fourier analysis so that the number of frequencies is more than five times the number of samples.

In Fig. 2, the number of samples used in inversion analysis is 90 so that 10% of samples have been dropped out. These have been chosen randomly. The comparison of inversion and Fourier spectra shows that the five spectral peaks are clearly found by inversion. The effect of dropping random samples out is to create noise peaks at random frequencies. The fact that the peaks in the inversion spectrum are smaller than those in the Fourier spectra must be due to the scaling of the spectrum in the way discussed above. When the total power is kept constant, a part of the power goes to noise peaks and a smaller amount is left for the main peaks. If all samples are used in calculating the inversion spectrum, the result is very similar to the Fourier spectrum.

Figure 3 shows two more examples of inversion spectra from signals with the same frequencies and amplitudes. The phases are randomly chosen, and therefore the signals are not identical but this does not affect the Fourier spectrum. Thus only the inversion spectra are shown. In the first example (two top panels), 25 randomly chosen samples out of the original 100 samples are dropped out and in the second

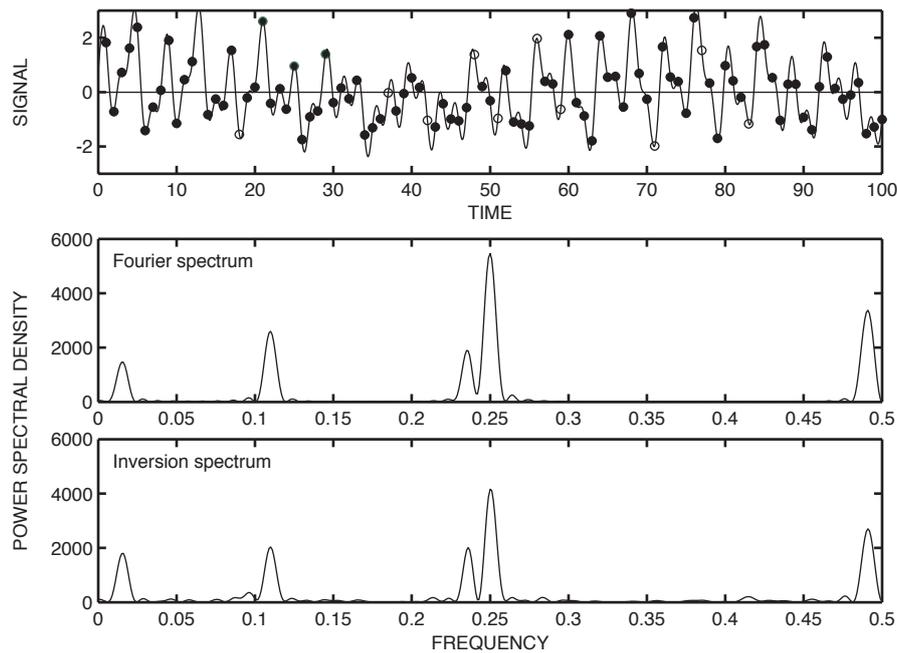


Fig. 2. Top panel: Test signal composed of sinusoidal signals at frequencies 0.015, 0.11, 0.235, 0.25 and 0.49 and respective amplitudes 0.5, 0.7, 0.5, 1.0 and 0.8. The phases are chosen randomly. The number of samples is 100 and they are indicated by markers. Middle panel: Power spectral density calculated from all samples in the top panel by means of Fourier transform. Bottom panel: Power spectral density calculated from the 90 samples indicated by filled dots in the top panel.

example (two bottom panels) the number of dropped samples is 41. The noise obviously increases with increasing number of dropped samples. Even in these cases the five spectral peaks are reproduced, although a large fraction of the original samples are not used in calculating the spectra. In the topmost example, however, the spectral peak at the lowest frequency (0.015) is only slightly higher than the noise peaks so that it could be taken as noise. In the bottom example the same peak is clearly higher than the noise peaks, regardless of the fact that the number of missing data points is higher. This reveals that the properties of the signal (random phases) and the true sampling times (randomly chosen set of signal samples) have random effects on the inversion spectrum.

The effect of random noise in the signal is portrayed in Fig. 4. The top panel shows the signal and sample values obtained by adding Gaussian random noise with standard deviation of 0.3 to the true signal values. The Fourier spectrum in the middle panel contains the true spectral peaks and weak noise at random frequencies. The inversion spectrum is calculated by dropping out 26 random samples from the original sample sequence. The true spectral peaks are clearly visible also in this case. The noise level is increased due to the gaps in the data.

In previous cases the missing samples are at random places so that no long data gap is present in the sequence. Figure 5 shows an example containing a single gap of 25 samples. The Fourier spectrum from the complete sample sequence

would be similar to that shown in Fig. 2. However, if only the incomplete data were available, Fourier transform could only be applied separately to the two pieces of the sample sequence and, consequently, the spectral resolution would be essentially worse. The inversion spectrum in Fig. 5 shows all five spectral peaks with no loss of spectral resolution. Still, the presence of the gap produces some noise, and the peak at the lowest frequency is close to the noise level.

It is also interesting to investigate a case with no fixed sampling time. Then a question arises what would be the highest frequency to be used in the inversion. A possible choice would be to take the shortest sampling interval as a time unit and to extend the spectrum up to one half of its inverse. This may lead to mirroring of spectral peaks much in the same way as negative frequencies appear in Fourier transform. Figure 6 portrays an example where sampling times are randomised to make long sampling intervals at some places and very short sampling intervals in other places. In this case Fourier transform could only be used after interpolation or by neglecting the uneven sampling interval and calculating the transform as if the samples were taken at equal spacing. The top panel of Fig. 6 shows the signal and samples, some of which lie so close to each other that the respective dots overlap. The middle panel shows the Fourier spectrum obtained by neglecting the uneven sampling interval and the bottom panel the inversion spectrum. The result indicates that the Fourier spectrum is unable to reproduce the small peak at a frequency of 0.235. Also, a considerable amount

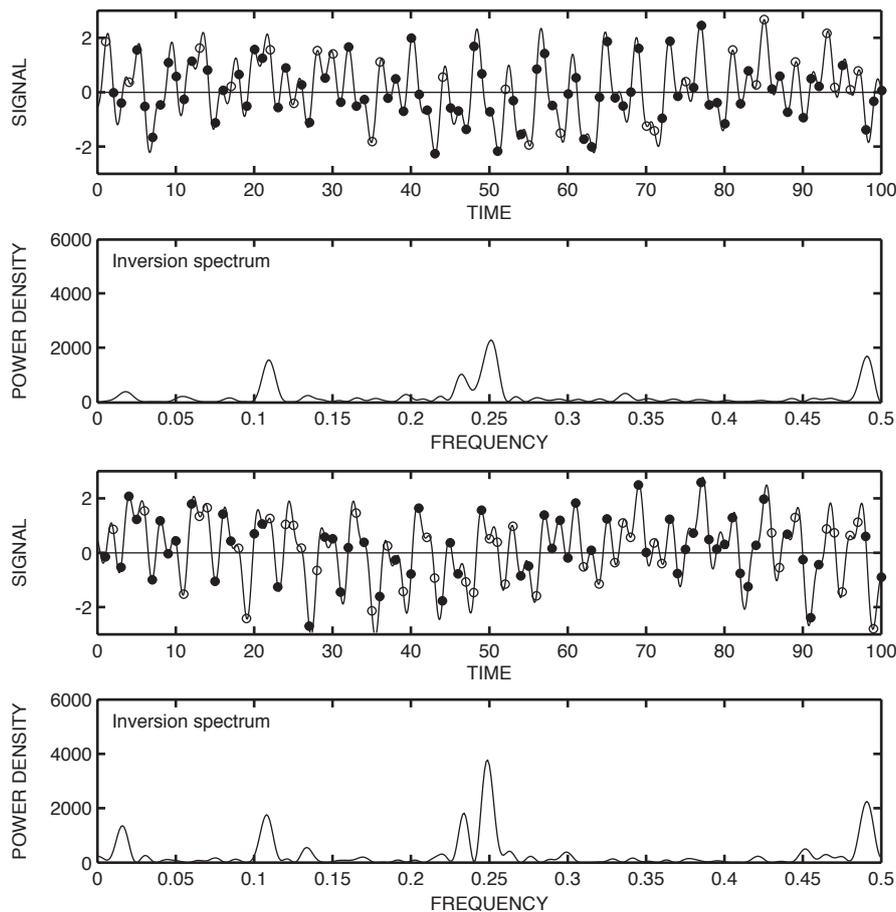


Fig. 3. Two examples similar to that in Fig. 2. The frequencies and amplitudes of the signals are the same as in Fig. 2, but the phases are different. Thus the Fourier spectra are identical and are not shown. In the first example (two top panels) the number of dropped samples is 25 and in the second example (two bottom panels) 41.

of noise appears in the Fourier spectrum. The noise level in the inversion spectrum is essentially lower and the peak at a frequency of 0.235 is better visible.

4 Spectra of ionosonde data

The F- and E-layer critical frequencies f_oF2 and f_oE from the Sodankylä ionosonde data were taken for testing the inversion method with real observations. Sodankylä is an auroral site located at $67^{\circ}22' N$, $26^{\circ}38' E$. In addition, the F-layer peak altitude $hmF2$ was calculated from the values of f_oF2 , f_oE and $M(3000)F2$ (the maximum usable frequency for a 3000-km radio path, not shown), using the method presented by Bradley and Dudeney (1973). The data come from the period 1957–2006, which covers about 4.5 solar cycles. Hourly monthly medians averaged over 10:00–14:00 LT were used in the analysis. Since months have different lengths, this leads to variable sampling intervals. The number of samples is different for different parameters, but it is smaller than 590 in each case. The shortest sampling interval was used for cal-

culating the highest frequency. Since the minimum sampling interval is a bit shorter than $1/12$ th of a year, the maximum frequency will be slightly higher than 6/year. The spectra were calculated at 1200 frequencies distributed evenly from zero to this maximum frequency. At an accuracy of four digits, this leads to a frequency step of 5.004×10^{-3} /year.

Figure 7 shows the F-region critical frequency in the top and the calculated inversion spectrum in the bottom panel. Only a single gap, indicated by the vertical grey stripe, is present in the sample sequence. The top panel shows clearly the solar cycle and annual variations. The spectrum is presented in logarithmic scale and it indicates a strong peak at 0.09508/year (as taken from the calculated spectrum), corresponding to a period of 10.52 years. This is due to the solar cycle. The annual variation and its harmonics are indicated by the peaks at 1/year, 2/year, 3/year and 4/year (as taken from the calculated spectrum, the annual peak lies at 0.9958/year). Harmonics necessarily appear, because annual variation is not sinusoidal. An interesting feature is the presence of two peaks around the annual variation. They lie

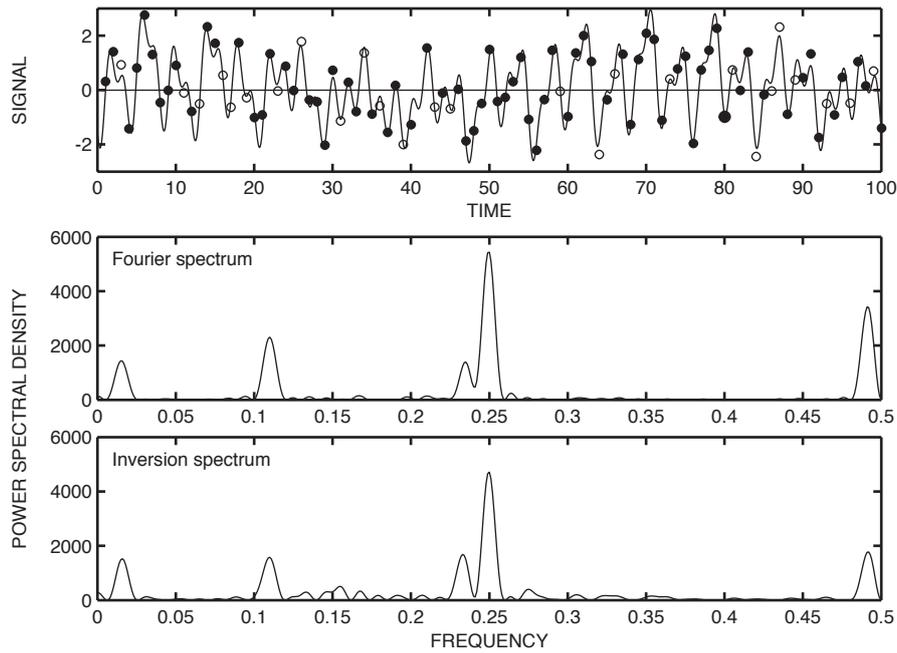


Fig. 4. Same as Fig. 2 but random Gaussian noise with a standard deviation of 0.3 has been added to the samples. The number of dropped samples is 26.

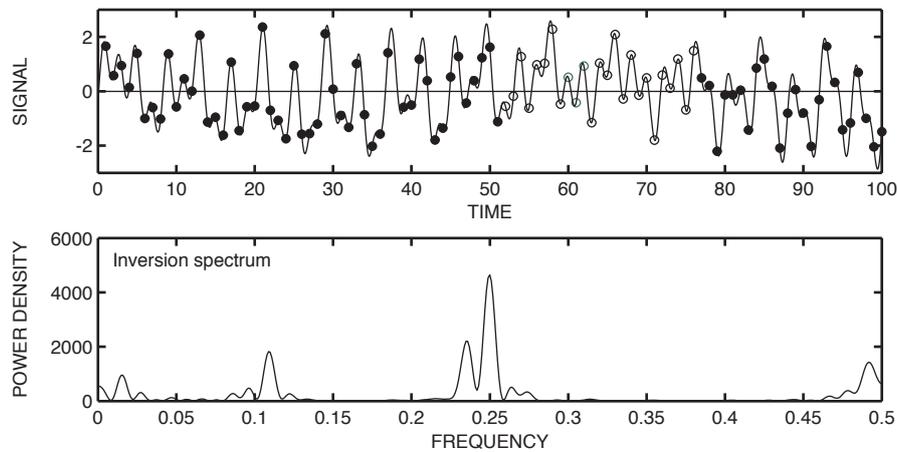


Fig. 5. Same as Fig. 2, but the inversion spectrum is calculated from a sample sequence with a gap of 25 samples in the sample sequence.

at frequencies 0.9058/year and 1.0909/year. The frequency of the first peak departs from the frequency of the annual variation by 0.09008/year and that of the second one by 0.09508/year. Hence the frequency differences are close to the frequency of the solar cycle, and an obvious explanation of the phenomenon is non-linear mixing between the solar cycle and the annual variation.

Similar results for the E-layer critical frequency are seen in Fig. 8. In this case there are more gaps in the data, but the number of frequencies in the spectrum as well as the frequency resolution are the same. Both the solar cycle and the annual variation together with its harmonics are visible

here, too. A difference is that the annual variation is now more pronounced than the solar cycle variation, a fact which is also seen in the top panel. Unlike in the case of *foF2*, there is no indication of mixing with the solar cycle.

Figure 9 portrays the F layer peak altitude *hmF2* and the corresponding spectrum. Now the data have even more gaps than in Fig. 8. This is due to the method applied in calculating *hmF2*; it puts some restrictions on the values of *foE* and *foF2* which can be used by the method. Even in this case the spectrum clearly shows the solar cycle and the annual variation with its two harmonics. There is also non-linear mixing with the solar cycle variation but, surprisingly, the side peaks

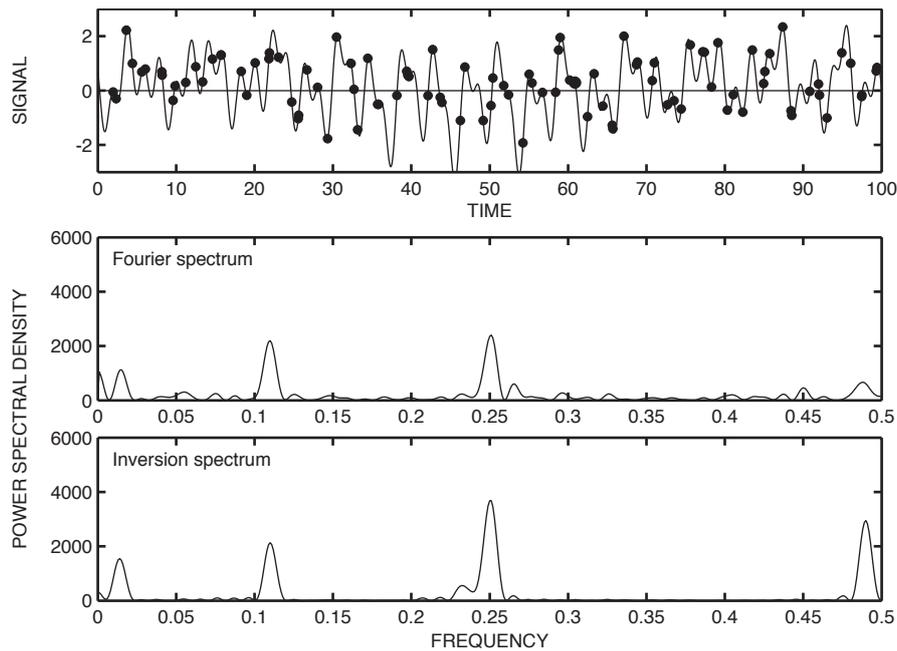


Fig. 6. Same as Fig. 2, but sampling times are randomised. The Fourier spectrum is calculated from all samples as if the sampling interval would be equal, but the true sampling times are used in calculating the inversion spectrum.

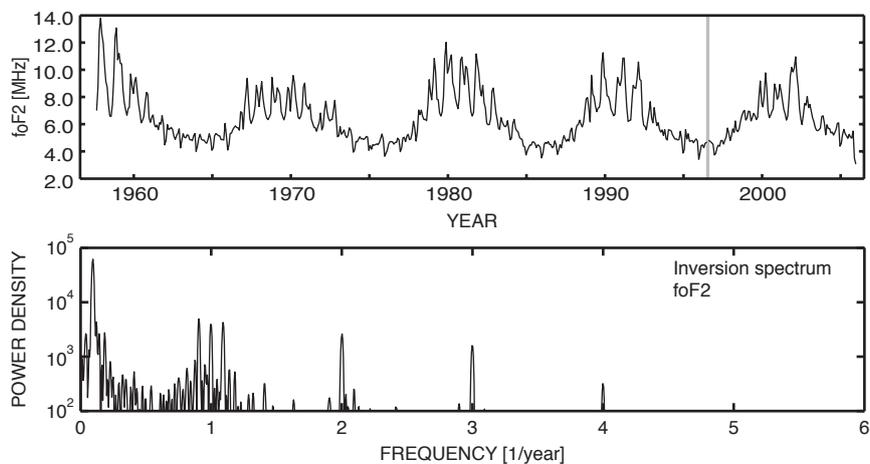


Fig. 7. Top panel: The F layer critical frequency f_oF2 from 1957–2006, as observed by the ionosonde at Sodankylä, Finland. The data points are mean values at 10:00–14:00 LT from the middle day of each month. Bottom panel: The inversion spectrum calculated from the data in the top panel. The whole spectrum consists of 1200 frequencies and the frequency resolution is 5.004×10^{-3} /year.

appear on both sides of the first harmonic of the annual variation rather than the annual variation itself. The reason for this is not understood.

Finally, Fig. 10 shows an example of a spectrum of real data sequence with very long gaps. The data consists of the values of $hmF2$ obtained from Sodankylä ionograms using real height analysis. The data covers the years 1958–1994, but real heights have not been determined from times of the gaps in the figure. The sampling step is variable but close to 30 days. Again, the results show the solar cycle variation.

Although the noise level is quite high, the annual and semi-annual variations are still visible. The high noise level is also seen in the top panel.

5 Discussion

The examples shown in this paper demonstrate the power of the inversion method in calculating spectra from data sequences with unequal sampling intervals, randomly missing data points or even long gaps. Comparisons with Fourier

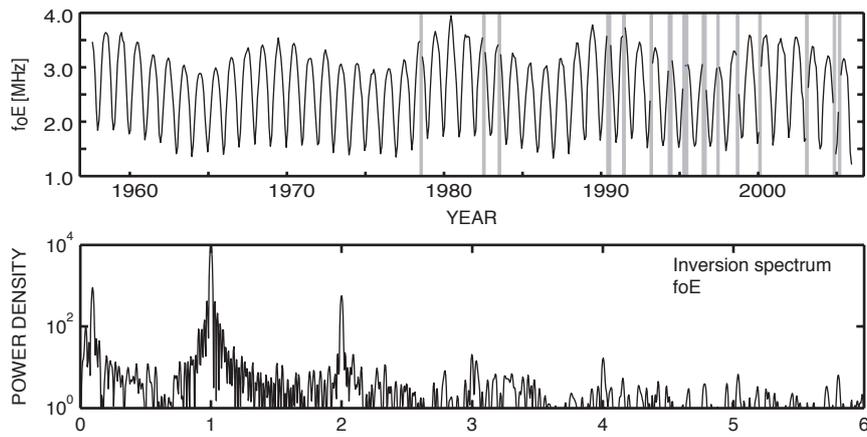


Fig. 8. Same as Fig. 7 for the E layer critical frequency f_oE . There are more gaps in the data than in Fig. 7.

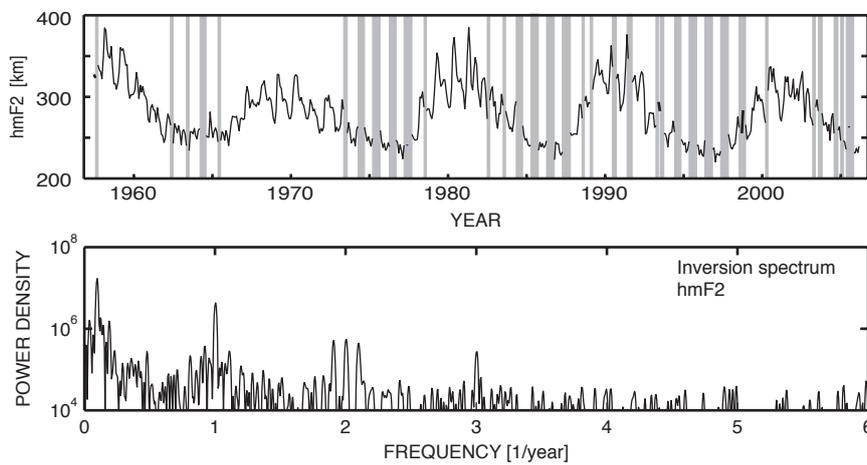


Fig. 9. Same as Fig. 7 for the F-layer maximum altitude $hmF2$. There are even more gaps in the data than in Fig. 8.

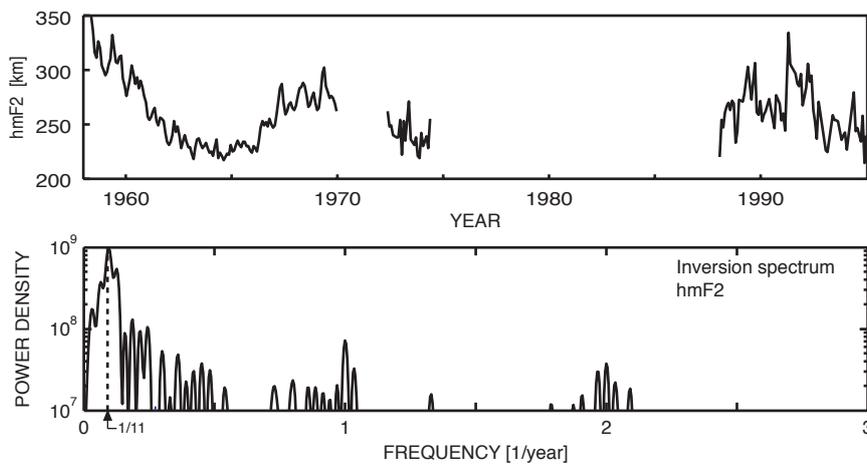


Fig. 10. Same as Fig. 7 for the F-layer maximum altitude $hmF2$ obtained by means of real height analysis of individual ionograms. The data contains large gaps.

spectra calculated from complete time series show that the spectral resolution (i.e. the widths of individual spectral peaks) is not affected by missing data points. This is only determined by the time span of the sample sequence; even long gaps in the data have no effect on the widths of the spectral peaks. Gaps in data or missing data points only create random noise peaks. If a gap is too long, weak spectral peaks may be embedded in noise.

The method described in this paper was actually presented earlier by Craymer (1998). However, a new aspect in the present work is that the number of frequency bins in the spectrum is not limited by the number of samples as believed before. This is convincingly demonstrated by a number of examples. The use of stochastic inversion as a starting point also gives a deeper insight in the method. Furthermore, if error estimates of data samples are available, the method also gives a possibility to calculate an error limit for each spectral point. This, however, is not investigated here.

Acknowledgements. This paper was presented at the 14th International EISCAT workshop in Tromsø, 2009. The data were collected and analysed by the staff of Sodankylä Geophysical Observatory during a period of several decades. Support from Space Institute at the University of Oulu is gratefully acknowledged.

Topical Editor K. Kauristie thanks two anonymous referees for their help in evaluating this paper.

References

- Barning, F. J. M.: Numerical analysis of the light-curve of 12 Lacertae, *Bull. Astron. Inst. Neth.*, 17(1), 22–28, 1963.
- Bradley, P. A. and Dudeney, J. R.: A simple model of the vertical distribution of electron concentration in the ionosphere, *J. Atmos. Terr. Phys.*, 35, 2131–2146, 1973.
- Craymer, M. R.: The least squares Spectrum, its inverse transform and autocorrelation function: Theory and some applications in geodesy (<ftp://geod.nrcan.gc.ca/pub/GSD/craymer/pubs/thesis1998.pdf>), PhD Dissertation, University of Toronto, Canada (1998).
- Lomb, N. R.: Least-squares frequency analysis of unequally spaced data, *Astrophys. Space Sci.*, 39, 447–462, 1976.
- Meisel, D. D.: Fourier transforms of data sampled at unequal observational intervals, *Astronom. J.*, 83(5), 538–545, 1978.
- Nygrén, T., Markkanen, M., Lehtinen, M., Tereshchenko, E. D., and Khudukon, B. Z.: Stochastic inversion in ionospheric radiotomography, *Radio Sci.*, 32, 2359–2372, 1997.
- Press, W. H. and Rybicki, G. B.: Fast algorithm for spectral analysis of unevenly sampled data, *Astrophys. J.*, 338, 277–280, 1989.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P.: Numerical recipes, 685–692, Cambridge University Press, 2007.
- Scargle, J. D.: Studies in astronomical time series analysis II. Statistical aspects of spectral analysis of unevenly spaced data, *Astrophys. J.*, 263, 835–853, 1982.
- Vaniček, P.: Approximate spectral analysis by least-squares fit, *Astrophys. Space Sci.*, 4, 387–391, 1969.
- Vaniček, P.: Further development and properties of the spectral analysis by least-squares, *Astrophys. Space Sci.*, 12, 10–33, 1971.