

This article was downloaded by:[Bochkarev, N.]  
On: 18 December 2007  
Access Details: [subscription number 788631019]  
Publisher: Taylor & Francis  
Informa Ltd Registered in England and Wales Registered Number: 1072954  
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Astronomical & Astrophysical Transactions

### The Journal of the Eurasian Astronomical Society

Publication details, including instructions for authors and subscription information:  
<http://www.informaworld.com/smpp/title~content=t713453505>

#### The time interferometer: Clean spectra by synthesis of correlation functions

V. V. Vityazev <sup>a</sup>

<sup>a</sup> Astronomy Department, St. Petersburg University, St. Petersburg, Russia

Online Publication Date: 01 November 1996

To cite this Article: Vityazev, V. V. (1996) 'The time interferometer: Clean spectra by synthesis of correlation functions', *Astronomical & Astrophysical Transactions*, 11:2,

111 - 137

To link to this article: DOI: 10.1080/10556799608205460

URL: <http://dx.doi.org/10.1080/10556799608205460>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

# THE TIME INTERFEROMETER: CLEAN SPECTRA BY SYNTHESIS OF CORRELATION FUNCTIONS

V. V. VITYAZEV

*Astronomy Department, St. Petersburg University, 198904, St. Petersburg,  
Petrodvorets, Bibliothechnaya pl. 2, Russia*

*(Received October 25, 1994)*

The paper presents a comparative study of the fundamentals, problems and techniques common to the spectral analysis of time series and interferometry. On the basis of the conceptual identity of the correlogram and visibility data, an attempt is made to adapt the aperture synthesis techniques well known in radio astronomy to the spectral analysis of time series. Two methods of synthesising a correlogram are proposed. The first one reproduces the idea of Ryle's interferometer and can be realized when averaging over statistical ensemble is possible. The second method is based on iterated correlograms and can be applied to a single curve with gaps. It is shown that our method yields, at least by iterations, a correlogram at all the points of the time span without gaps, and consequently clean spectra can be obtained. The method is described in detail and numerical examples are presented to illustrate the application of the algorithm to gapped astrometrical time series and to time series with uneven precision of the measurements.

KEY WORDS Time series, power spectra

## 1 INTRODUCTION

In earlier paper (Vityazev, 1994) a striking similarity of a spectral window and the beam of an interferometer was found and the concept of the Time Interferometer was introduced. In this paper we discuss further the properties of the Time Interferometer as a tool for spectral analysis of gapped time series. A comparative study of time series analysis and interferometry shows many common problems. In particular, a correlogram (time series analysis) and visibility data (interferometry) are essentially one and the same if viewed from the standpoint of the associated mathematics. Thus, the idea of aperture synthesis widely used in radio astronomy can be transferred to the spectral analysis of time series. The basic problem considered in this paper is: a time series, known at some interval of time containing gaps, is assumed to be either a set of realizations or a single curve. The problem then is to obtain the values of the correlation function (the correlogram) at all points of the

initial interval. In Sections 2–4 a review of the fundamentals, concepts, problems and methods common to the time series analysis and interferometry is made. In Section 5 Ryle's aperture synthesis technique is used in the time series analysis under assumption that the time series is given by a set of realizations. The synthesis of a correlogram for a single realization is proposed in Section 6. This procedure is essentially based on the correlation transform successively applied to the initial time series and to the resulting correlograms. As a result, the computation of the final correlogram at all points of the initial time span becomes possible. The efficiency of this technique in the presence of noise is studied in Section 7. The applications of the method to the gapped astrometrical time series and to the time series with non-uniform accuracy of the measurements are shown in Sections 8 and 9.

## 2 FUNDAMENTALS OF THE TIME SERIES SPECTRAL ANALYSIS

Suppose that a time series  $X(t)$  is a stationary stochastic process with zero mean defined by a set of realizations,

$$X(t) = [x_p(t)]_{p=1}^N \quad 0 \leq t \leq T < \infty. \quad (2.1)$$

Using  $E$  to denote the mathematical expectation value, one has for the *autocorrelation function* of  $X(t)$ :

$$K(t, t') = E[X(t) X(t')]. \quad (2.2)$$

Due to the assumed properties of  $X(t)$ , the function  $K(t, t')$  depends on the difference  $\tau = t - t'$ ,

$$K(t, t') = k(\tau) = k(-\tau). \quad (2.3)$$

The *power spectrum*  $G(\omega)$  and the auto-correlation function  $k(\tau)$  are related to each other via Fourier transforms:

$$G(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} k(\tau) \exp(-i\omega\tau) d\tau, \quad (2.4)$$

$$k(\tau) = \int_{-\infty}^{+\infty} G(\omega) \exp(i\omega\tau) d\omega. \quad (2.5)$$

To describe gaps in observations we introduce the *time window function*,

$$h(t) = \begin{cases} 1 & \text{if at time } t \text{ the data exists,} \\ 0 & \text{if at time } t \text{ the data is absent.} \end{cases} \quad (2.6)$$

With this notation the observed time series can be represented as

$$Y(t) = h(t)X(t). \quad (2.7)$$

Calculate the *periodogram* of the  $p$ -th realization:

$$d_p(\omega) = \frac{1}{2\pi T} \left| \int_0^T y_p(t) \exp(-i\omega t) dt \right|^2, \quad (2.8)$$

and evaluate the averaged periodogram,

$$D(\omega) = E[d_p(\omega)]. \quad (2.9)$$

Under the conditions stated above, the relationship between the spectra  $D(\omega)$  and  $G(\omega)$  is given by the convolution

$$D(\omega) = \int_{-\infty}^{+\infty} G(\omega') W(\omega - \omega') d\omega', \quad (2.10)$$

where the *spectral window* function  $W(\omega)$  is the periodogram of the time window function

$$W(\omega) = \frac{1}{2\pi T} \left| \int_0^T h(t) \exp(-i\omega t) dt \right|^2. \quad (2.11)$$

Now, we introduce two inverse Fourier transforms

$$k_D(\tau) = \int_{-\infty}^{+\infty} D(\omega) \exp(i\omega\tau) d\omega, \quad (2.12)$$

$$H(\tau) = \int_{-\infty}^{+\infty} W(\omega) \exp(i\omega\tau) d\omega. \quad (2.13)$$

With these notations, one has from Eq. (2.10)

$$k_D(\tau) = H(\tau) k(\tau). \quad (2.14)$$

Figure 1 summarizes the fundamentals of time series analysis considered here. For further details reader is referred to Jenkins and Watts (1968), Deeming (1975a, 1975b), Otnes and Enocson (1978), Marple (1978), Terebizh (1992).

### Comments

1. Equations (2.10) and (2.11) are the basic tools in evaluating the observational spectrum  $D(\omega)$  and in understanding the artifacts introduced by missing data. Suppose the spectrum  $G(\omega)$  has sharp features at  $\omega = \pm\omega_0$ ,

$$G(\omega) = A[\delta(\omega - \omega_0) + \delta(\omega + \omega_0)], \quad (2.15)$$

## TIME SERIES

## Theoretical quantities

Power spectrum  
 $G(\omega)$

Correlation function  
 $k(\tau)$

Frequency  
 $\omega$

Time lag  
 $\tau$

## Estimators

Periodogram  
 $D(\omega)$

Correlogram  
 $k_D(\tau)$

## Connections

Convolution  
 $D(\omega) = W(\omega) \otimes G(\omega)$

Multiplication  
 $k_D(\tau) = C(\tau) k(\tau)$

## Observations

Spectral window  
 $H(\omega)$

No name  
 $C(\tau)$

Figure 1 Fundamentals of spectral analysis of time series.

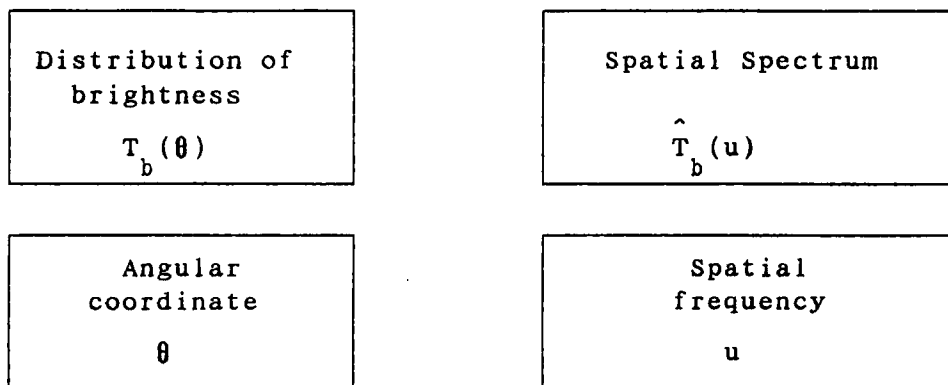
where  $A = \text{const}$  and  $\delta(\omega)$  is the Dirac's delta function. According to Eq. (2.10), we obtain

$$D(\omega) = A[W(\omega - \omega_0) + W(\omega + \omega_0)]. \quad (2.16)$$

Hence, Eq. (2.10) describes the transformation of the initial spectrum  $G(\omega)$  to the

## INTERFEROMETRY

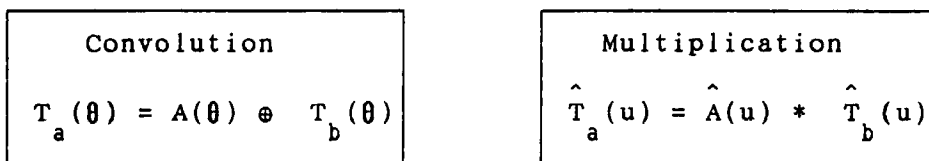
## Theoretical quantities



## Estimators (Images)



## Connections



## Observations

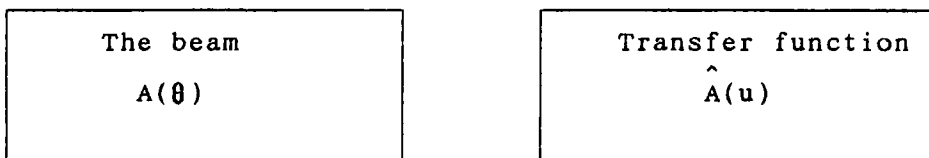


Figure 2 Fundamentals of interferometry.

observed one  $D(\omega)$ . It is important to stress that the transition from  $G(\omega)$  to  $D(\omega)$  is completely determined by the spectral window. If the data are irregularly spaced,

the spectral window usually has a *central peak* at  $\omega = 0$ , *side peaks* centered at some frequencies  $\omega = \bar{\omega}_k$ ,  $k = 1, 2, \dots$ , and *side lobes* of the central peak and of the side peaks. The side lobes are caused by a finite time span of the observations. They exist even for uniformly spaced data. The side peaks appear only if the data are spaced irregularly. According to Eq. (2.12), all these features are present in the observed spectrum. Indeed, the central peak is visible at  $\omega = \omega_0$ , and, showing true features of the initial spectrum, it can be called the true peak. The side peaks are responsible for the spectral features at  $\omega = |\omega \pm \omega_k|$ . The initial spectrum  $G(\omega)$  has no lines at these frequencies. Hence, these features are false and may be regarded as "ghosts".

### 3 FUNDAMENTALS OF INTERFEROMETRY

The *intensity of radiation* from any source on the sky can be described in terms of position  $(\alpha, \delta)$ , wavelength  $(\lambda)$  and time  $(t)$ . Each measurement is an average over a band of wavelengths and over some span of time. Here, the one-dimensional, monochromatic and instantaneous approximation is used to simplify the discussion. The resulting specific intensity  $T_b(\theta)$  that describes the distribution of the source brightness along the arc of a circle ( $\theta$  is the angular coordinate) has a Fourier transform  $\hat{T}_b(u)$ , which is called *the spectrum of spatial frequencies*. Correspondingly, when an interferometer measures *the visibility data*  $\hat{T}_a(u)$ , the image, or the map  $T_a(\theta)$  can be calculated by the Fourier transform of  $\hat{T}_a(u)$ . Two fundamental relations are valid:

$$T_a(\theta) = A(\theta) \oplus T_b(\theta) \quad (\text{Convolution}), \quad (3.1)$$

$$\hat{T}_a(u) = \hat{A}(u) \cdot \hat{T}_b(u) \quad (\text{Multiplication}), \quad (3.2)$$

where  $A(\theta)$  is *the beam* of an interferometer, and  $\hat{A}(u)$  is *the transfer function*. Equation (3.2) determines an interferometer as *a filter of spatial frequencies*, whereas Eq. (3.1) explains why, due to the convolution of  $T_b(\theta)$  with the beam  $A(\theta)$ , the resulting image is called a dirty map. Figure 2 summarizes the fundamentals of interferometry considered here. For further details the reader is referred to Esepkina *et al.* (1973) and Thompson *et al.* (1986).

### 4 COMPARISON OF THE CONCEPTS, PROBLEMS AND METHODS

From Figures 1 and 2 one can see that in the analysis of time series and in the interferometry the Fourier transforms, the convolution and the correlation are the basic mathematical tools. Moreover, these figures show "who is who" in both sciences. Really, at the first level we introduce rigorous (theoretical) quantities. In the spectral-analysis case they are the power spectrum and the correlation function; in the interferometry their counterparts are the distribution of brightness and the

spatial spectrum. At the second level we have estimators of the strict quantities. In the spectral-analysis, they are the periodogram and the correlogram, whereas in the interferometry these are the map and the visibility data, respectively. Finally, equations which connect the quantities of the two levels are identical (convolution and multiplication) and include the characteristics of observations: the beam and the transfer function and their analogs, i.e. the spectral window and the function  $C(\tau)$ , for which in the spectral analysis no name exists.

In reality, due to finite dimension of mirrors and finite time spans of observations we cannot get the distributions of brightness on the sky and the power spectra of the time series, and what we can do is to find their as good as possible approximations. In optics or in radio astronomy, when the filled apertures are used, the maps are produced directly in the focal plane of a telescope. Analogously, when the time series is given at all points of some interval or at time points regularly spaced within the interval, the evaluation of the periodogram can be made quite easily. When an interferometer is used, the aperture is not solid, and what we can measure is the visibility data, i.e. the estimator of the spectrum of spatial frequencies. The longer the baseline, the less area in the  $(u - v)$ -plane can be filled and the more dirty the resulting map becomes after transformation of the visibility data from the  $(u - v)$ -onto  $(\alpha - \delta)$ -plane (in the one-dimensional case the  $u$ - and  $\theta$ -domains should be considered, but for the sake of convenience we use the standard terminology). To overcome this, various techniques of the *aperture synthesis* are used, and this leads to complete solution of the problem since it allows to fill the  $(u - v)$ -plane completely. If the *aperture synthesis* provides the partial filling of the  $(u - v)$ -plane, the cleaning procedures can be used with the aim to eliminate from the map the artifacts of the "holes" in the  $(u - v)$ -plane. The same problems we meet in the spectral-analysis case, when the time points are distributed irregularly or have long gaps. In this case the correlograms cannot be determined for all values of time lag  $\tau$ , and this would give false features in the resulting periodograms. This is the point where the main idea of the present paper is hidden. It is: whether it is possible to apply the *aperture synthesis* method to spectral analysis of time series?

## 5 SYNTHESIS OF THE CORRELOGRAM

It is known that to fulfill the aperture synthesis one should have an interferometer with changeable baseline. Of all the schemes of the aperture synthesis, the one proposed by Ryle (1960) is the most suitable for us (Figure 3). This tool consists of two antennas  $A$  and  $B$  fixed at the separation  $L$ . The third antenna  $C$  is moving in the interval  $[L/2, L]$ . At each position of the moving antenna one obtains two interferometers  $(AC)$  and  $(CB)$  with the baselines  $l$  and  $L/2 + l$ , and they yield the values of the visibility data  $\hat{T}_a$  at the points  $u = l/\lambda$  and  $u = U/2 + l/2$ , where  $U = L/\lambda$ . Obviously, while antenna  $C$  sweeps all the interval  $[L/2, L]$ , the visibility data  $\hat{T}_a(u)$  become available at all points of the interval  $[0, U]$ .



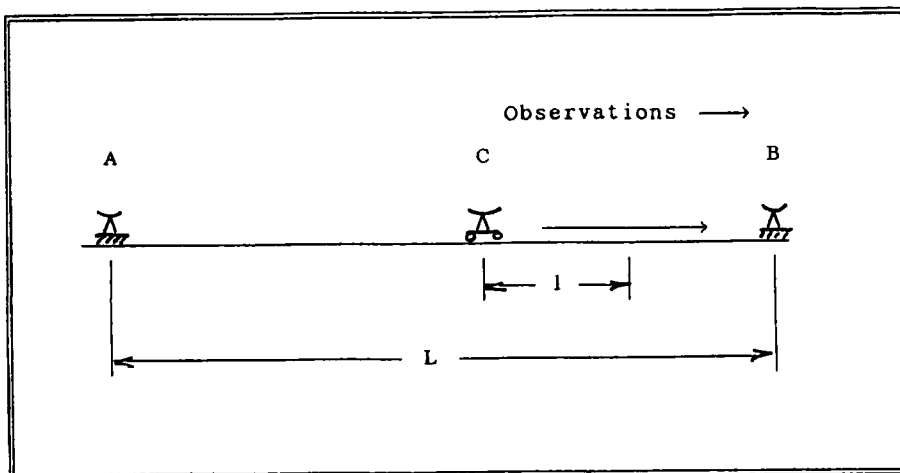


Figure 3 Ryle's scheme of aperture synthesis.

Now we apply this idea to the time series analysis. It is not difficult to show that

$$D(\omega) = \frac{1}{\pi} \int_0^T C(\tau) k(\tau) \cos(\omega\tau) d\tau, \quad (5.1)$$

where (5.2)

$$C(\tau) = \frac{1}{T} \int_0^{T-\tau} h(t) h(t+\tau) dt. \quad (5.3)$$

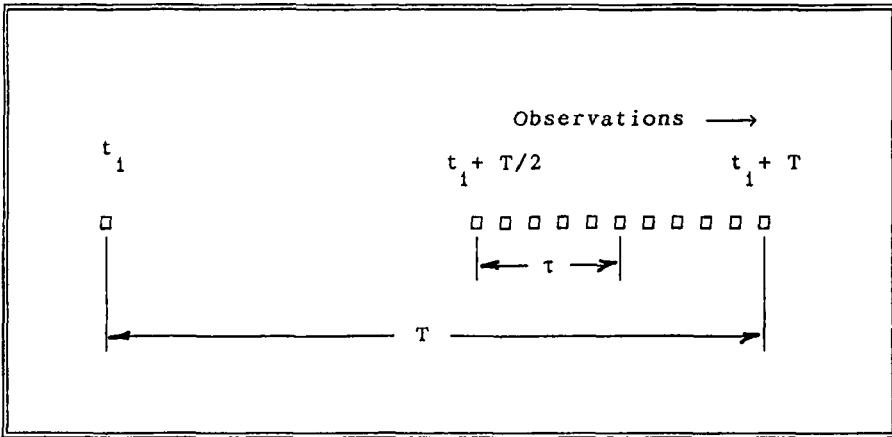
When the time series  $X(t)$  is known at each point within the interval  $[0, T]$ , the function  $h(t)$  equals unity elsewhere, and from Eqs. (5.1) and (5.2) we get

$$D(\omega) = \frac{1}{\pi} \int_0^T (1 - \tau/T) k(\tau) \cos(\omega\tau) d\tau. \quad (5.4)$$

Comparison of Eqs. (5.3) and (2.10) tells us that in the absence of gaps the spectral window  $W(\omega)$  can be written in the form

$$W(\omega) = \left[ \frac{\sin(\omega T/2)}{\omega T/2} \right]^2 \quad (5.5)$$

This spectral window has no side peaks, and consequently the power spectrum corresponding to (5.4) is clean. Now we see that the principal problem in evaluating the periodogram (5.1) is to evaluate the correlation function  $k(\tau)$  at all points of



**Figure 4** The synthesis of a correlogram.

the interval  $[0, T]$ . When we have no gaps this can be done easily with the help of Eqs. (2.2) and (2.3). Really, every two points spaced at the distance  $\tau$  may be called the Time Interferometer with variable baseline  $\tau$ , and each such pair yields the estimation of the correlation function (the correlogram) by averaging the products  $X(t)X(t + \tau)$  over the set of realizations. This follows from our assumption that  $X(t)$  is a stationary stochastic process, and it is very important for our study. To proceed further, assume that we have a gap in observations. Let the length of the gap be  $l$  and the longest distance between the borders of the gap and the boundary points of the interval  $[0, T]$  be  $a$ . If  $l \leq a$  then the correlation function  $k(\tau)$  can be evaluated at all points  $\tau \in [0, T]$ , otherwise only on the subintervals  $[0, a]$  and  $[l, T]$ . Figure 4 illustrates the former situation. Here we assume that initially the time series is given at two points  $t_1$  and  $t_1 + T/2$ . This pair of points (the Time Interferometer) allows us to get the correlation function only at the point  $\tau = T/2$ . Let us make new observations at the points  $t_1 + T/2 < t < t_1 + T$ . It is clear that each new point  $t = t_1 + T/2 + \tau$  yields two additional values of the correlogram, namely, at the points  $\tau$  and  $T/2 + \tau$ . Obviously, when the observations will cover the interval  $[t_1 + T/2, t_1 + T]$ , we shall obtain the values of the correlogram at all points of the interval  $[0, T]$ , and the periodogram  $D(\omega)$  calculated from Eq. (5.3) will be clean.

It is instructive to compare Figures 3 and 4. The fixed antennas  $A$  and  $B$  are the counterparts of the boundary points  $t_1$  and  $t_1 + T$ , while each new position of the moving antenna  $C$  is nothing else but the new point of observations. Since Ryle's interferometer makes the synthesis of the visibility data, it is a good reason to call the estimation of the function  $k(\tau)$ , shown in Figure 4, the synthesis of the correlogram.

Return now to the situation when  $l > a$ . In this case the correlogram turns out to be evaluated at all points except the new gap of the length  $l - a$ . Now, for the

periodogram  $D(\omega)$  one has

$$D(\omega) = \frac{1}{\pi} \int_0^T (1 - \tau/T) \bar{h}(\tau) \cos(\omega\tau) d\tau, \quad (5.6)$$

where the correlation window  $\bar{h}(\tau)$  is defined as

$$\bar{h}(\tau) = \begin{cases} 1, & \text{if } 0 \leq \tau < a \text{ or } 1 < \tau \leq T, \\ 0, & \text{if } a < \tau < 1. \end{cases} \quad (5.7)$$

It is clear that the spectral window corresponding to the correlation window (5.6) has side peaks and Eq. (5.5) gives us a dirty spectrum in which the artifacts of the new gap can be seen. Nevertheless, one can expect that this new spectrum will be less contaminated than the periodogram calculated directly from Eqs. (2.8) and (2.9). Thus we see that to clean the spectrum completely one needs to perform more observations until the condition  $l \leq a$  becomes true.

## 6 SYNTHESIS OF THE CORRELOGRAM FOR A SINGLE REALIZATION

The method considered in the previous section is based on the possibility to evaluate the function  $k(\tau)$  by averaging the product  $X(t)X(t+\tau)$  over a set of realizations. In astronomy, we usually have time series that are defined by a single curve. In this section we consider the kind of the synthesis which can be done in such cases.

With this aim, we assume that the time series is given as a product of two functions

$$y(t) = h(t)f(t), \quad 0 \leq t \leq T, \quad (6.1)$$

where  $T$  is the length of the realization. Now we introduce the periodogram  $D(\omega)$  and the correlogram  $\chi(\tau)$  of the time series (6.1) by the following equations:

$$D(\omega) = \frac{1}{2\pi T} \left| \int_0^T y(t) e^{-i\omega t} dt \right|^2, \quad (6.2)$$

$$\chi(\tau) = \frac{1}{T} \int_0^{T-\tau} y(t) y(t+\tau) dt. \quad (6.3)$$

It can be easily shown that both functions are related by

$$D(\omega) = \frac{1}{\pi} \int_0^T \chi(\tau) \cos(\omega\tau) d\tau. \quad (6.4)$$

As before, the spectral window  $h(t)$  is defined by Eq. (2.6). If the time series  $f(t)$  is a continuous function then the spectral window  $h(t)$  can be considered as having values 0 or 1 at some subintervals of the time span  $[0, T]$  in such a way that

$$T_h = \int_0^T h(t) dt \leq T, \quad (6.5)$$

where  $T_h$  is the total length of the observation period and  $T - T_h$  is the length of gaps.

Assume that  $f(t)$  is a polyharmonic process

$$f(t) = \sum_{k=1}^n A_k \cos(\omega_k t + \varphi_k), \quad (6.6)$$

where  $A_k$ ,  $\omega_k$  and  $\varphi_k$  are the amplitude, the frequency and the phase of the  $k$ -th harmonic. It is not difficult to show that for sufficiently large  $T$  the correlogram (6.1) of the function (6.6) is given by

$$\chi(\tau) = H(\tau) \sum_{k=1}^n (A_k^2/2) \cos(\omega_k \tau), \quad (6.7)$$

where

$$H(\tau) = \frac{1}{T} \int_0^{T-\tau} h(t) h(t + \tau) dt. \quad (6.8)$$

Substitution of (6.8) into (6.4) yields

$$D(\omega) = \frac{1}{4} \sum_{k=1}^n A_k^2 [W(\omega - \omega_k) + W(\omega + \omega_k)], \quad (6.9)$$

where the spectral window  $W(\omega)$  is

$$W(\omega) = \frac{1}{\pi} \int_0^T H(\tau) \cos(\omega \tau) d\tau. \quad (6.10)$$

Again, the spectral window  $W(\omega)$  depends on the time window  $h(t)$ . If there are no gaps in observations, we have

$$H(\tau) = 1 - \frac{\tau}{T} \quad (6.11)$$

and the spectral window, defined by Eq. (5.4), yields a clean power spectrum. Otherwise, when some observations are missing, the spectral windows having side-peaks at frequencies  $\bar{\omega}_p$ ,  $p = 1, 2, \dots$ , will produce dirty spectra, i.e. the spectra contaminated by false features at the frequencies  $|\omega_k \pm \bar{\omega}_p|$ .

Comparison between Eqs. (6.1) and (6.7) shows that the correlogram of  $f(t)$  multiplied by the time window  $h(t)$  has its own system of gaps, defined by the function

$$h^{(1)}(t) = \begin{cases} 1, & \text{if } H(\tau) > 0, \\ 0, & \text{if } H(\tau) = 0. \end{cases} \quad (6.12)$$

It is easy to understand that the new time window  $h^{(1)}(t)$  defines the new gap, the length of which is less than that of the initial time window, i.e.

$$T_h^{(1)} < T_h. \quad (6.13)$$

For this reason it is always possible to fill the gaps completely at least by iterations. Thus we see that the correlation transform defined by Eq. (6.3) reconstructs the polyharmonic function (6.6) inside the interval  $[0, T]$  at the subinterval which is longer than the set on which the function was available initially. Of course, this reconstruction is not complete: the information on phases is lost and due to the amplitude squaring the weak harmonics may be lost as well. Nevertheless, the cicatrizing of the gaps leads us to the clean spectra.

For this purpose we introduce the following iterative procedure:

$$\chi^{(\nu)}(\tau) = \frac{1}{T} \int_0^{T-\tau} \chi^{(\nu-1)}(\tau') \chi^{(\nu-1)}(\tau' + \tau) d\tau', \quad (6.14)$$

$$\nu = 1, 2, \dots,$$

where

$$\chi^{(0)}(\tau) = \frac{1}{T} \int_0^{T-\tau} h(t) h(t + \tau) f(t) f(t + \tau) dt. \quad (6.15)$$

For the polyharmonic function (6.6), provided that  $T$  is sufficiently large, the iterated correlogram has the following structure:

$$\chi^{(\nu)}(\tau) = H^{(\nu)}(\tau) \sum_{k=1}^n a_{k\nu} \cos(\omega_k \tau), \quad (6.16)$$

where

$$a_{k\nu} = 2 \left[ \frac{A_k}{2} \right]^{2\nu+1} \quad (6.17)$$

and  $H^{(\nu)}(\tau)$  can be computed by the following procedure

$$H^{(\nu)}(\tau) = \frac{1}{T} \int_0^{T-\tau} H^{(\nu-1)}(\tau') H^{(\nu-1)}(\tau' + \tau) d\tau'. \quad (6.18)$$

$$\nu = 1, 2, \dots,$$

where

$$H^{(0)}(\tau) = \frac{1}{T} \int_0^{T-\tau} h(t) h(t + \tau) dt. \quad (6.19)$$

Now, each step of the procedure yields a periodogram

$$D^{(\nu)}(\omega) = \frac{1}{\pi} \int_0^T \chi^{(\nu)}(\tau) \cos(\omega\tau) d\tau, \quad (6.20)$$

which, due to Eq. (6.9), can be written in the form

$$D^{(\nu)}(\omega) = \frac{1}{2} \sum_{k=1}^n a_{k\nu} [W^{(\nu)}(\omega - \omega_k) + W^{(\nu)}(\omega + \omega_k)], \quad (6.21)$$

where

$$W^{(\nu)}(\omega) = \frac{1}{\pi} \int_0^T H^{(\nu)}(\tau) \cos(\omega\tau) d\tau, \quad (6.22)$$

$$\nu = 0, 1, 2, \dots$$

Equation (6.22) tells us that in the process of iterations the profiles of the spectral lines change their form. To study these changes we assume that the correlation window  $h(t)$  can be described by the polyharmonic function

$$h(t) = \alpha_0 + \sum_{j=1}^{\infty} \gamma_j \cos(\bar{\omega}_j t + \varphi_j), \quad (6.23)$$

where  $\alpha_0$  and  $\gamma_j$  characterize the presence of the gap (when the interval  $[0, T]$  is filled completely  $\alpha_0 = 1$ ,  $\gamma_j = 0$ ,  $j = 1, 2, \dots$ ). Now, for  $T \gg 1$  straightforward calculations yield

$$H^{(0)}(\tau) = \frac{T-\tau}{T} \left[ \alpha_0^2 + \sum_{j=1}^{\infty} \frac{\gamma_j^2}{2} \cos(\bar{\omega}_j \tau) \right], \quad (6.24)$$

and at the  $\nu$ -th step we get

$$H^{(\nu)}(\tau) = R^{(\nu)}(\tau) \alpha_0^{2^{\nu+1}} \left[ 1 + \sum_{j=1}^{\infty} \beta_{j\nu} \cos(\bar{\omega}_j \tau) \right], \quad (6.25)$$

where

$$\beta_{j\nu} = 2 \left[ \frac{\gamma_j}{2\alpha_0} \right]^{2^{\nu+1}}, \quad (6.26)$$

$$R^{(\nu)}(\tau) = \frac{1}{T} \int_0^{T-\tau} R^{\nu-1}(\tau') R^{(\nu-1)}(\tau' + \tau) d\tau', \quad (6.27)$$

$$\nu = 1, 2, \dots,$$

$$R^{(0)}(\tau) = 1 - \frac{\tau}{T}. \quad (6.28)$$

We see that the influence of the gap on  $H^{(\nu)}(\tau)$  diminishes (when  $\gamma_j < 2\alpha_0$  as  $\nu$  increases. For this reason, in the process of iterations, the side peaks in the spectral windows  $W^{(\nu)}(\tau)$  become weaker, and subsequently the false peaks in the periodograms  $D^{(\nu)}(\omega)$  disappear. Thus one can say that the amplitude squaring suppresses the effects of gaps in the resulting spectra. The iterations must be halted when the side peaks in the spectral windows are negligible compared to the intensity of the central peak. The formalization of this will be given in the next section.

## 7 SYNTHESIS OF THE CORRELOGRAM IN THE PRESENCE OF NOISE

Let us see how our method works when a Gaussian noise with zero mean value is added to the signal  $f(t)$ . In this case the periodogram will be contaminated not only by the "ghosts" due to gaps but by the random peaks due to the noise as well. In spectral analysis of noisy data the probability distribution of the random variable  $D^{(0)}(\omega)$  for the case when the time series is pure noise is of paramount importance. It is known that for even time series the probability distribution is exponential, but in the general case of irregularly sampled data this is not so (Scargle, 1982). Nevertheless, in many practical cases the deviation is rather small (Scargle, 1982; Terebizh, 1992), and the exponential law is not a bad approximation. Moreover, it is not difficult to show that the distortion of the exponential law occurs only at the frequencies  $\bar{\omega}_k/2$ ,  $k = 1, 2, \dots$ , where the  $\bar{\omega}_k$  are those frequencies at which the spectral window has side peaks (but not side lobes). In particular, for all the points distributions considered in this paper, the strongest distortion corresponding to the first side peak does not exceed the factor two. Neglecting this single frequency feature, we adopted for the rest of them the exponential law. This enabled us to use the standard techniques for detecting signals in noise. In our problem due to two sources of contamination for the  $M$ -point discrete time series it is useful to introduce the following quantities:

$$D_q = -\frac{\langle D^{(0)} \rangle}{1+r} \ln[1 - (1-q)^{\frac{2}{M-2}}], \quad (7.1)$$

$$D_h = D_{\max}^{(0)} W_{\max}^{(0)} / W^{(0)}(0), \quad (7.2)$$

where  $\langle D^{(0)} \rangle$  is the mean value of the periodogram  $D^{(0)}(\omega)$ ,  $D_{\max}^{(0)}$  is the largest value in the periodogram  $D^{(0)}(\omega)$ ,  $W_{\max}^{(0)}$  is the intensity of the strongest side peak in the spectral window  $W^{(0)}(\omega)$ ,  $r$  is the "signal to noise" ratio, and  $q$  is the significance level for detecting a signal in noise ( $0 \leq q \ll 1$ ).

Obviously,  $D_q$  is the detection threshold of a signal in noise, while  $D_h$  is the threshold to separate the true spectral lines from the false peaks (due to gaps in observations). The value  $D_q$  is based on Walker's distribution of the strongest

account in the periodogram (Walker, 1914; Terebizh, 1992). The introduction of the ratio  $r$  in Eq. (7.1) allows us to solve several problems. If a pure signal is considered, then  $r = \infty$  and  $D_q = 0$ . On the contrary, when we have pure noise, we put  $r = 0$ . In general,  $0 \leq r \leq \infty$  and we consider the following situations.

(1)  $D_h \leq D_q$ . This inequality occurs when noise is sufficiently strong, but the gaps in observation are small. In this case all false peaks are hidden in the peaks of noise, since the gaps spoil the spectrum less than noise. Here we have no reason to clean the spectrum, and the detection of a signal is claimed according to the well known rule: the spectral line at frequency  $\omega$  is thought to be a signal with the probability  $(1 - q)$  if

$$D^{(0)}(\omega) \leq D_q. \quad (7.3)$$

Still, if we shall apply the method of iterated correlograms with the convergence condition

$$W_{\max}^{(\nu)}/W^{(\nu)}(0) \ll W_{\max}^{(0)}/W^{(0)}(0), \quad (7.4)$$

then with the same probability we shall obtain the true lines while all "ghosts" and noisy peaks will be suppressed. Thus we see that our method not only cleans the spectra from the features due to missing data, but suppresses the noise as well.

(2)  $D_h > D_q$ . This is the opposite case: the data is slightly noisy, but the gap is long. This time the cleaning of the spectrum is needed, and the iterations must be made until the false peaks are below the detection threshold:

$$W_{\max}^{(\nu)}/W^{(0)} \leq D_q/D_{\max}^{(0)}. \quad (7.5)$$

After the end of this procedure we come to the situation described earlier.

Summing up the results of Sections 6 and 7, one can say that the correlation transform (6.3) of a polyharmonic function given on the interval of time with a gap has a property to reconstruct the function at the gaps. This restoration can be regarded as a specific kind of the synthesis described in Section 5. Indeed, while the averaging over the set of realizations yields directly  $k(\tau)$ , the averaging over the set of time points yields  $k(\tau)$  weighted with the correlation window  $H(\tau)$ . This function remembers the fact that the time series was gapped and produces in the periodogram (6.4) just the same false peaks that would be were the periodogram was computed directly from Eq. (6.2). The second, third, etc., correlation transforms make the gap smaller and smaller, but at the same time they force the spectral window  $H^{(\nu)}(\tau)$  to be more and more like the triangular window (6.11). There are two parameters  $D_q$  and  $D_h$  that control the algorithm. If we want to eliminate completely the effects of gaps, we must stop the iterations according to condition (7.4). Unfortunately, in this case there is a danger that weak lines in the spectrum will be lost. In other words, in the final spectrum we shall find only the lines of those harmonics whose amplitudes obey the condition

$$A_k > A_{\max} \left[ \frac{W_{\max}^{(0)}}{W^{(0)}(0)} \right]^{1/2}, \quad (7.6)$$

where  $A_{\max}$  is the largest amplitude in (6.6).



A more realistic approach which takes into account the presence of noise tells us that a complete suppression of the false peaks is not necessary for a reasonable interpretation of the spectrum. It is quite sufficient to reduce the false features so that they are below some level defined by the detection threshold  $D_q$ . In this case the end of iterations is determined by the condition (7.5).

In reality, the suppression of weak signals is not strong if the contaminations of the spectra by noise and by gaps are comparable. In such cases only 2–3 iterations are needed. In the opposite case more iterations is required and the distortion of the spectral lines in the final spectrum may become noticeable. Still, even distorted, the remaining lines are the true ones, since after the complete synthesis of the correlogram we have no reason to expect the appearance of the “ghosts”.

## 8 NUMERICAL RESULTS

To see how our method works when data are taken from observations, we took a time series of the polar variations of latitude (henceforth, the PVL)

$$\Delta\varphi(t) = x(t) \cos(\lambda) - y(t) \sin(\lambda), \quad (8.1)$$

where the coordinates of the pole  $x(t)$  and  $y(t)$  were taken from the Annual Reports of the BIH for 1967 to 1986 at the interval  $\Delta t = 0.1$  yr. The values  $\Delta\varphi(t)$  were computed for the longitude  $\lambda$  of the Pulkovo observatory. This time series consists of the annual and Chandler harmonics with periods 1.00 and 1.18 yr, respectively. The properties of the PVL are well known and it is the reason to use it as a standard while testing new methods of spectral analysis. The original PVL is slightly noisy, that is why to test our method in the presence of noise we added to the PVL the Gaussian noise with zero mean value in such a way that in all examples the ratio “signal to noise” equals unity. Also, in all numerical runs we fixed the significance level at  $q = 0.05$ . Besides this, to enhance the contrast between weak and strong lines in the spectra we plotted not the periodograms but square roots of them. The resulting correlograms and periodograms are given in the normalized form (with respect to the largest value). Below we show the application of our method to several typical patterns of gaps.

### 8.1 An Even Distribution of Points

Figure 5 illustrates the application of our method to the time series PVL (120 regularly spaced points without gaps). The levels  $D_q$  and  $D_h$  are indicated for the periodograms. In this example,  $D_h$  denotes the level of the first side lobe (not the side peak). As expected, we have  $D_q > D_h$ , and, strictly speaking, no cleaning is needed. Still, we made two iterations and suppressed both the noisy peaks and the structure of the side lobes. In this case the periodogram was contaminated by noise and side lobes (but not by the side peaks).

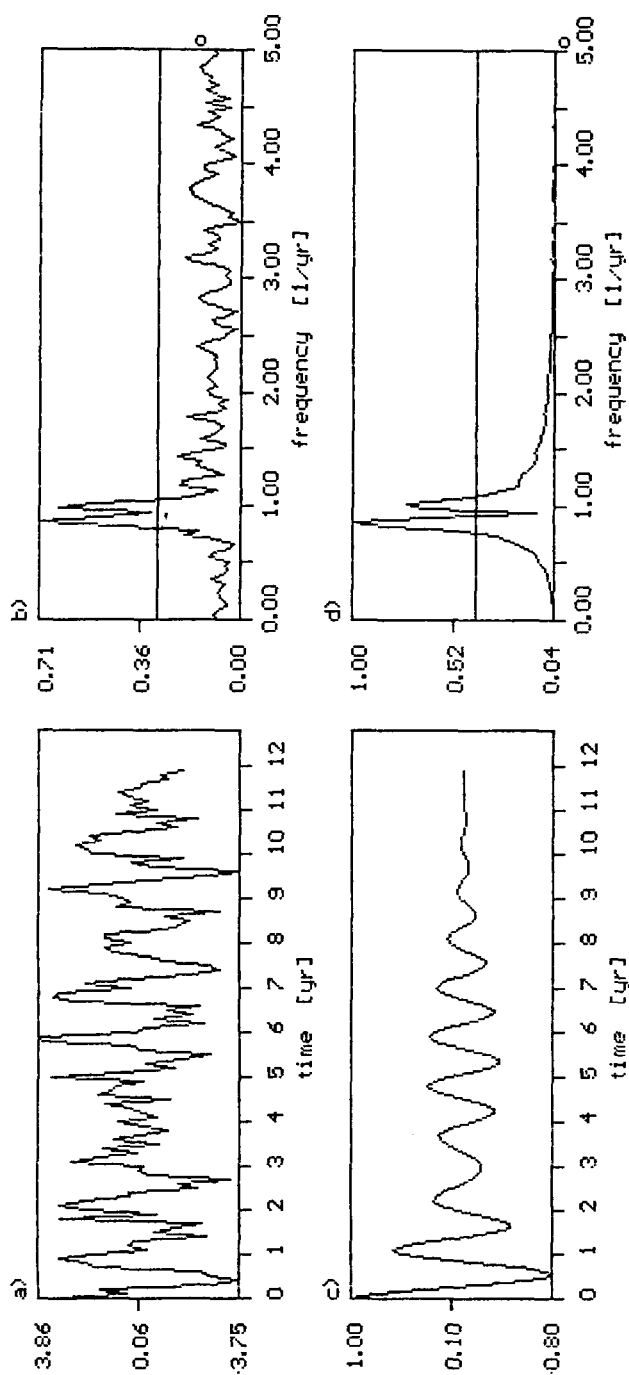


Figure 5 Spectral analysis of the PVL by the method of iterated correlograms. The time series consists of 120 regularly spaced points sampled over the interval 0.1 yr. The "signal to noise" ratio equals unity. The horizontal line corresponds to the detection threshold  $D_q$  with  $q = 0.05$ . The circle marks the level of the first side lobe. (a), the initial time series; (b), the periodogram  $D^{(0)}(\omega)$ ; (c), the second - order correlogram; (d), the second order periodogram  $D^{(2)}(\omega)$ .

### 8.2 *A Long Gap*

In this case of 120 regularly spaced points we omitted 54 points starting from the 34-th one. The results of spectral analysis are shown in Fig. 6. The periodogram of initial time series with the long gap turned out to be contaminated by false lines and noisy peaks. The detection thresholds are now in the relation  $D_q < D_h$ . For this reason we made 3 iterations and obtained the 3-d order correlogram without gaps and the 3-d order periodogram in which all the false peaks have become insignificant.

### 8.3 *Irregularly Missed Points*

Figure 7 illustrates the application of our algorithm to the spectral analysis of the PVL given at 120 points of which 57 points were randomly omitted. We see that the initial periodogram contains many features. In this case we have  $D_q > D_h$ , so the spectrum can be interpreted without confusion. Still, we made one iteration and suppressed the level of noisy peaks and that of false lines.

### *Periodic Gaps*

Of all the patterns of gaps the periodic gaps are the most typical of astronomical data. We tested our method for this case too, making 0.5 yr gaps following each other with 1 yr period. Figure 8 shows that our method worked successfully: two iterations were enough to separate true lines from noise and to suppress all ghosts below the detection threshold.

## 9 ANALYSIS OF TIME SERIES WITH UNEVEN VARIANCE OF NOISE

The practice of long-time acquisition of data shows that the resulting time series are seldom uniform with respect to the intrinsic accuracy of the measurements. This is caused by various obstacles: aging of instruments, change of observers, periodical changes of weather, etc. The variations in accuracy are clearly seen in the compiled time series, the fragments of which have been obtained by observers at different observatories.

It is a curiosity, but in the vast scope of special literature dedicated to spectral analysis, treating of time series with variable noise component has not yet been considered. The first step smoothing procedures are of little importance in this case, since they redistribute the noise and spoil those fragments where the accuracy was initially good. Still, there is one simple way to make the time series uniform with respect to noise. It implies that all the measurements should be properly weighted. This approach is widely used in the least squares technique. One can hope that the weights will suppress the noise component, but at the same time it is clear that the weights may substantially change the spectral content of the systematic component.

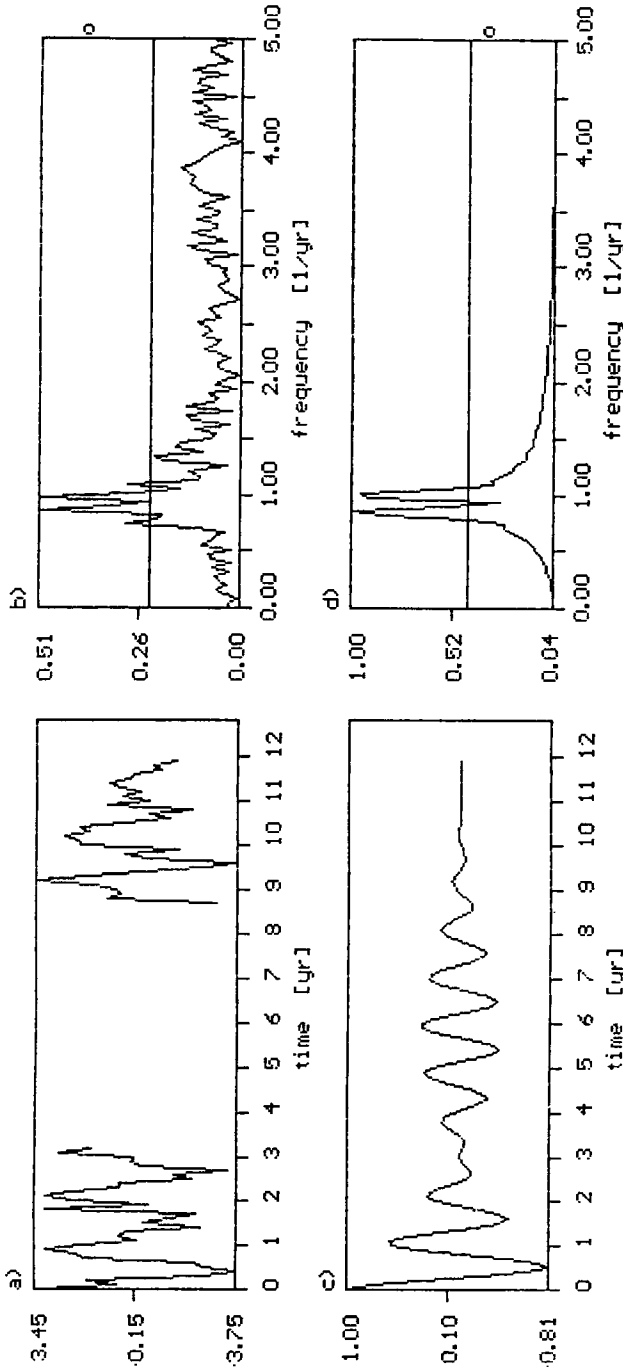


Figure 6 Spectral analysis of the PVL by the method of iterated correlograms. The time series consists of two 33-point segments separated by 54-point gap. The "signal to noise" ratio equals unity. The horizontal line corresponds to the detection threshold  $D_q$  with  $q = 0.05$ . The circle marks the level of the first side peak. (a), the initial time series; (b), the third order correlogram; (c), the dirty spectrum  $D^{(0)}(\omega)$ ; (d), the clean spectrum  $D^{(3)}(\omega)$ .

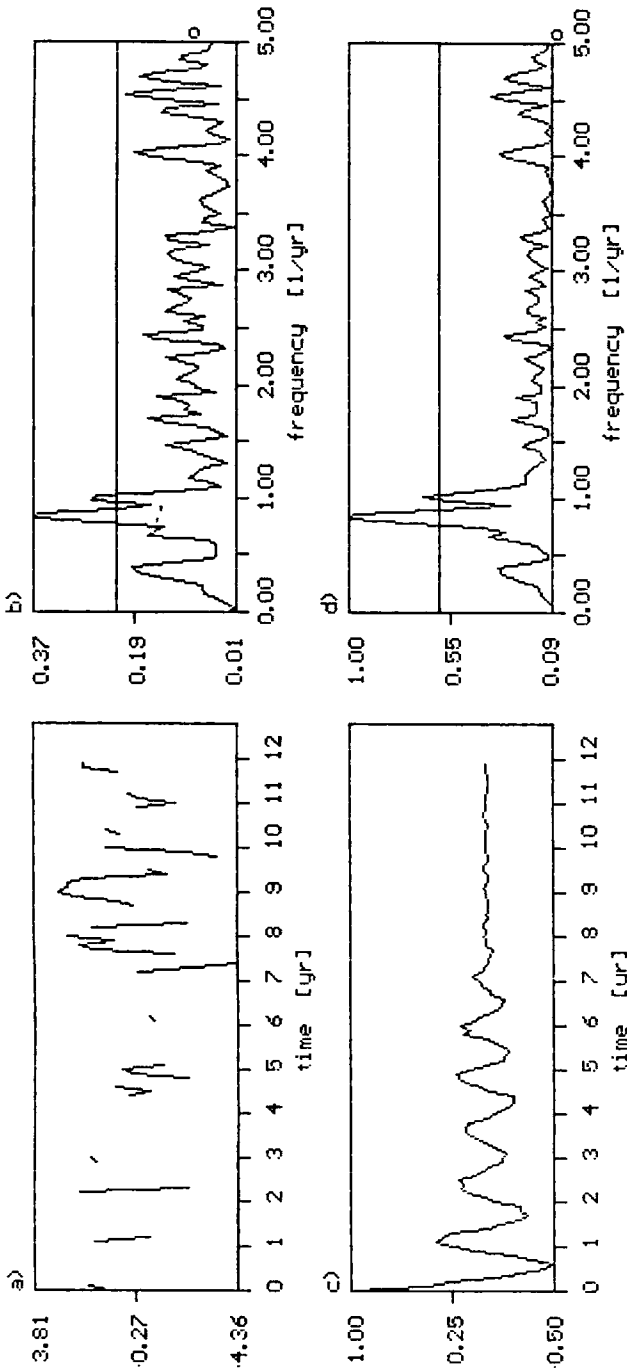


Figure 7 Spectral analysis of the PVL by the method of iterated correlograms. The time series consists of 63 irregularly spaced points. The "signal to noise" ratio equals unity. The horizontal line corresponds to the detection threshold  $D_q$  with  $q = 0.05$ . The circle marks the level of the first side peak. (a), the initial time series; (b), the dirty spectrum  $D^{(0)}(\omega)$ ; (c), the third order correlogram; (d), the clean spectrum  $D^{(1)}(\omega)$ .

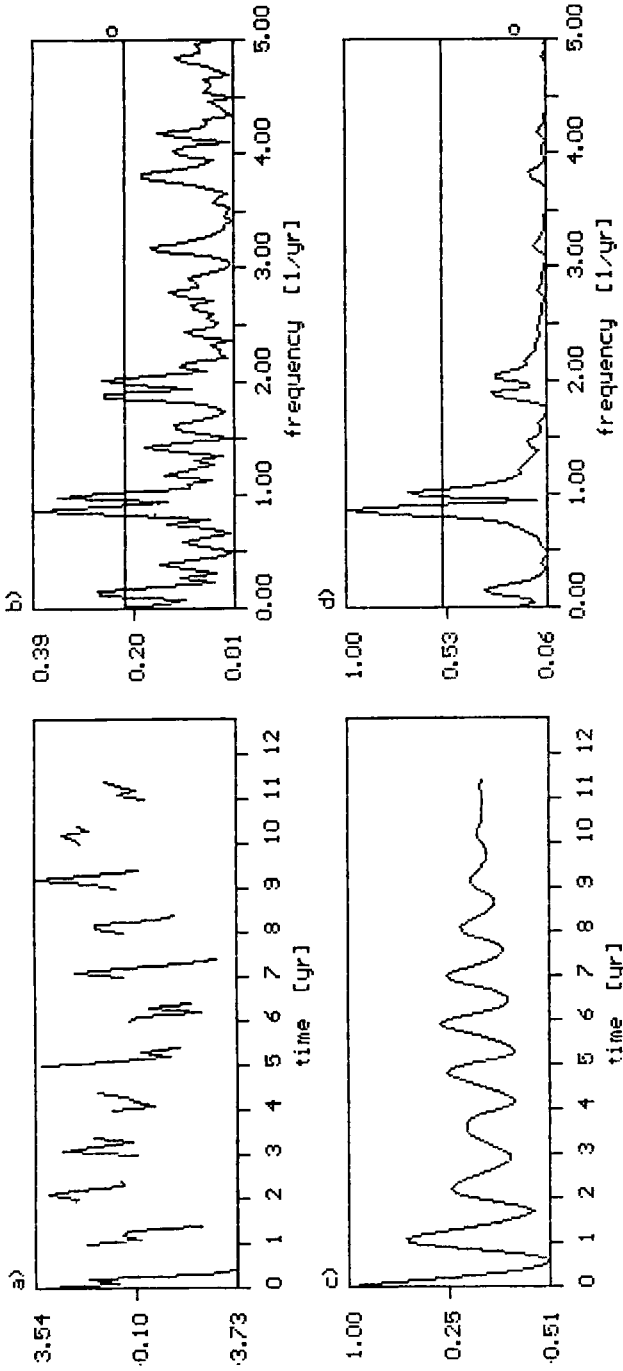


Figure 8 Spectral analysis of the PVL by the method of iterated correlograms. The time series consists of 12 five-point segments separated by five-point gaps. The "signal to noise" ratio equals unity. The horizontal line corresponds to the detection threshold  $D_q$  with  $q = 0.05$ . The circle marks the level of the first side peak. (a), the initial time series; (b), the dirty spectrum  $D^{(0)}(\omega)$ ; (c), the second order correlogram; (d), the clean spectrum  $D^{(2)}(\omega)$ .

If some points of the time series are multiplied by zero weights, then it means that these points are simply omitted. Whence it is clear that the treating of time series with assigned weights is close to the problem of treating gapped time series.

Now we assume that the polyharmonic function (6.6) is observed in such a way that at various time points the root mean square error has two values  $\sigma_1$  and  $\sigma_2$ , ( $\sigma_1 < \sigma_2$ ). For this reason the time window function is determined as follows:

$$w(t) = \begin{cases} 1 & \text{if at point } t \text{ the r.m.s.e. equals } \sigma_1 \\ \sigma_1/\sigma_2 & \text{if at point } t \text{ the r.m.s.e. equals } \sigma_2. \end{cases} \quad (9.1)$$

All remaining definitions and mathematical relations remain unchanged except that the value  $r$  in the definition of  $D_q$  must be replaced by

$$r = \rho_1 \{1 - \lambda(1 - (\sigma_1/\sigma_2)^2)\}, \quad (9.2)$$

where  $\lambda$  is a fraction of weighted information (with respect to the total number of points, for example),  $\rho_1$  is the "signal to noise" ratio at those points where  $w(t) = 1$ .

Now, we again give numerical results to show the application of our approach to the variably noised time series.

### 9.1 A Long Fragment of Measurements with Low Precision

This time the PVL was divided into 3 segments containing 33, 54 and 33 points. It was assumed that the first and the third fragments were measured with high precision ("signal to noise" ratio was 2), while the second segment with low precision ("signal to noise" ratio was 0.2). In Figure 9 we see that the periodogram of the raw time series is too noisy to be properly understood. After weighting the 54-point segment with  $w(t) = 0.316$ , the level of noise in the periodogram substantially decreased, and at a 95 percent confidence level we could claim the detection of four lines. The inequality  $D_h > D_q$  tells us that false peaks are present among these four peaks. One iteration suppressed the noise peaks and forced the ghosts to fall below the detection threshold  $D_q/D_{\max}^0$ . As a result, the two lines – the annual and Chandler's ones are clearly seen far above the level of 0.05 significance in the 3-d order periodogram.

### 9.2 Periodic Changes of Accuracy.

Figure 10 illustrates spectral analysis of the PVL when the "signal to noise" ratio varies from 0.1 to 10 at a 1 yr period. Again, the periodogram of the initial time series turned out to be strongly contaminated by noise. The suitable weighting by  $w(t) = 0.1$  at once gave a periodogram with strong false peaks, which have been eliminated by iterations.

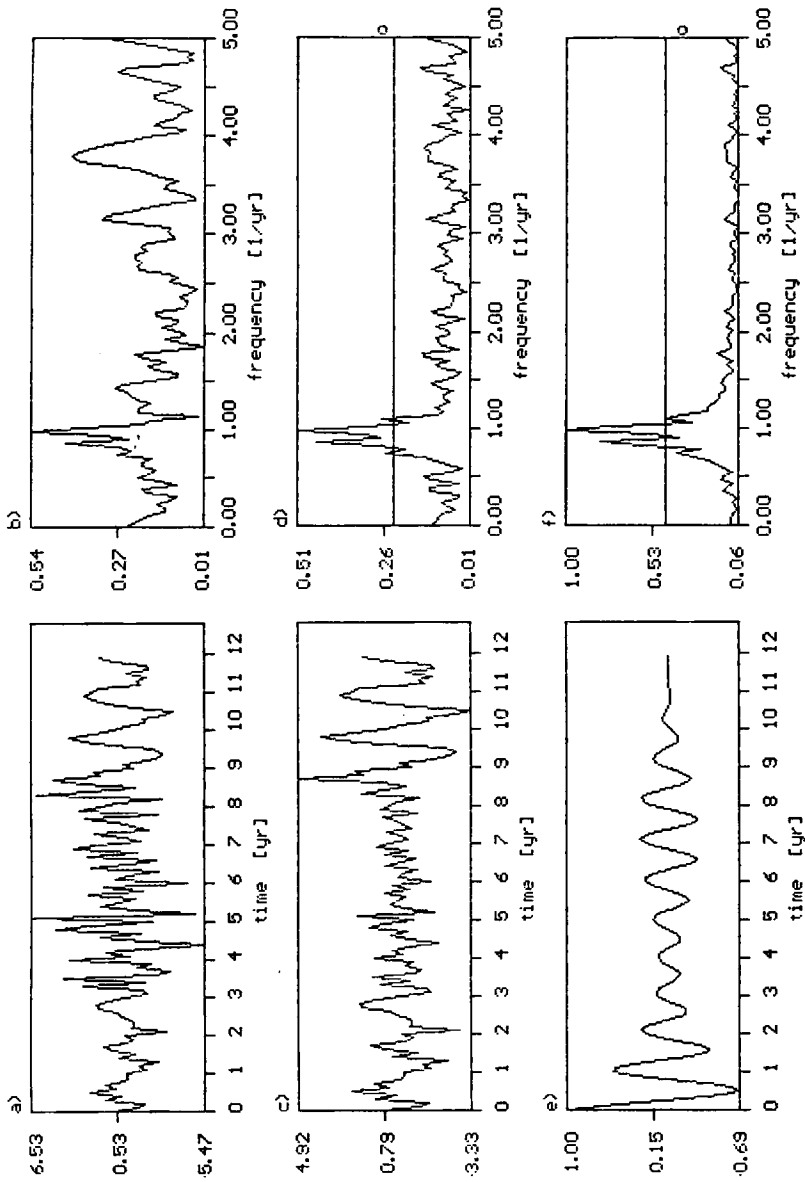


Figure 9 Spectral analysis of the PVL with varying accuracy of measurements. The initial time series consists of three segments containing 33, 54 and 33 points. At the first and at the third segment the "signal to noise" ratio is 2, while at the second segment this is 0.2. The horizontal line corresponds to the detection threshold  $D_q$  with  $q = 0.05$ . The circle marks the level of the first side peak. (a), the initial time series; (b), the periodogram of the initial time series; (c), the initial time series in which the 54-point segment is weighted with  $w(t) = 0.316$ ; (d), the dirty spectrum of weighted time series; (e), the first order correlogram; (f), the clean spectrum  $D^{(1)}(\omega)$ .



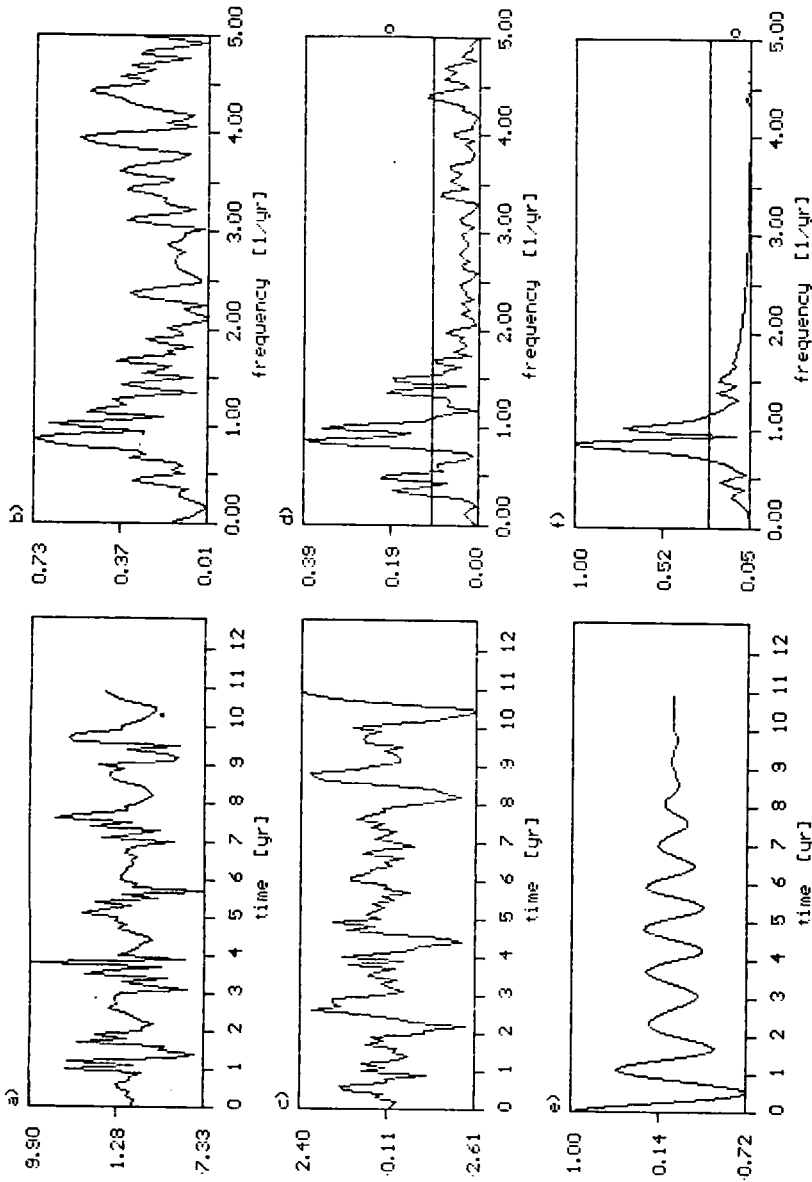


Figure 10 Spectral analysis of the PVL with varying accuracy of measurements. The initial time series consists of eleven ten-point segments. At the first, the third, etc., segments the "signal to noise" ratio is 10, at the rest it is 0.1. The detection threshold  $D_0(q = 0.05)$  is traced by the horizontal line. The circle marks the level of the first side peak. (a), the initial time series; (b), the periodogram of the initial time series; (c), the weighted time series with  $w(t) = 0.1$ ; (d), the dirty spectrum of the weighted time series; (e), the second order correlogram; (f), the clean spectrum  $D^{(2)}(\omega)$ .

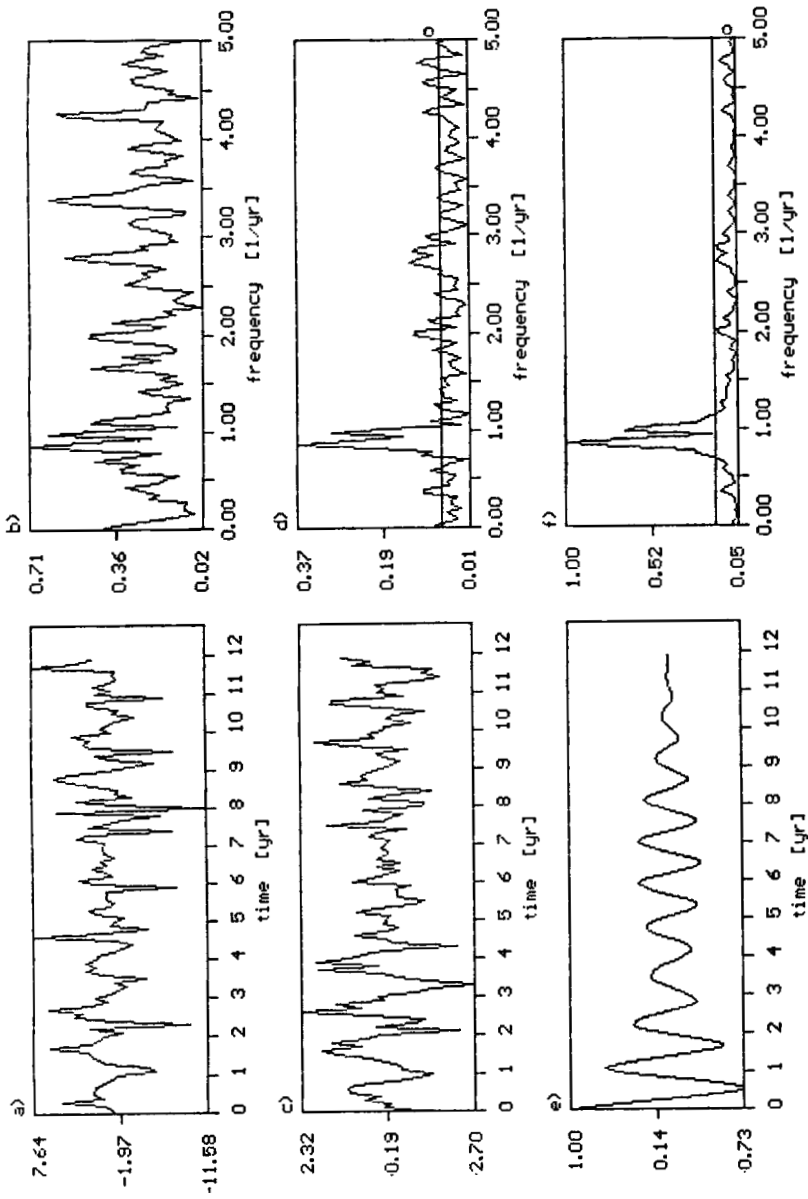


Figure 11 Spectral analysis of the PVL with varying accuracy of measurements. At 57 randomly chosen points the "signal to noise" ratio is 10, while at the rest of 120 points it is 0.1. The horizontal line corresponds to the detection threshold  $D_q$  with  $q = 0.05$ . The circle marks the level of the first side peak. (a), the initial time series; (b), the periodogram of the initial time series; (c), the weighted time series with  $w(t) = 0.1$ ; (d), the dirty spectrum of weighted time series; (e), the first order correlogram; (f), the clean spectrum  $D^{(1)}(\omega)$ .

### 9.3 A Strong Noise at Randomly Distributed Points

In this example it was simulated that the "signal to noise" ratio equals 10 at 57 randomly chosen points while at the rest of 120 points, considered in example 9.1, it is 0.1. Figure 11 shows the work of the method under consideration in this case. We have the same situation: the weighting decreases noise but produces false lines. The cleaning procedure based on the iterated correlograms reduces the "ghosts" below the detection threshold, leaving only two real lines in the final periodogram.

## 10 CONCLUSIONS

The idea of the Time Interferometer is useful for it helps us to adapt some methods of interferometry to problems of the spectral analysis of time series (and, hopefully, vice versa). This concept forced us to make a thorough comparison of fundamentals, method and problems common to both sciences. Partially such study was made by Roberts *et al.* (1987) when they transferred the two-dimensional CLEAN algorithm (Hogbom, 1974) widely used in the radio-astronomy to the spectral analysis of gapped time series.

The basic topic of this paper is an attempt to use the idea of the  $(u - v)$ -plane synthesis to the computation of correlograms. This idea follows from the conceptual identity between the visibility data and the correlogram. Two methods of the correlogram synthesis have been proposed. The first one is basically interferometric and can be realized if the time series is given as a set of realizations. When the only realization is available, the averaging over realizations is replaced by time averaging. In this case one can synthesize only product of correlograms and a certain correlation window. Here the idea of synthesis becomes not so evident, and its further development has led us to the method of iterated correlograms. This method was described in detail and its applications to gapped and weighted astrometric time series was made.

In fact, the method of iterated correlograms solves the so-called restoration problem. The simplest inverse method to solve this problem is based on the deconvolution of Eq. (2.10), but this way is successful only in the absence of noise. In the presence of inevitable noise associated with any measurements, the simple deconvolution method proves to be unsuccessful (Brault and White, 1971). On the contrary, our method is not inverse since it is based essentially on direct correlation transforms which can only suppress noise, not to enhance it. From this follows that our method is stable in the presence of noise. As mentioned above, our method suppresses weak harmonics as well and the more the longer the gap is. For this reason the application of the method should be restricted to time series in which the total length of gaps is no more than 50–70 percent of the total time span.

*References*

- Brault, J. W., White, O. R. (1971) *Astron. Astrophys.* **13**, 169.
- Deeming, T. J., (1975a) *Astrophys. Space Sci.* **36**, 137.
- Deeming, T. J., (1975b) *Astrophys. Space Sci.* **42**, 257 (E).
- Esepkina, N. A., Korolkov, D. N., Pariysky, Yu. N. (1973) *Radiotelesopes and Radiometers*, Moscow, Nauka.
- Hogbom, J. A. (1974) *Astron. Astrophys. Suppl.* **15**, 417.
- Jenkins, G. M. and Watts, D. G. (1968) *Spectral Analysis and Its Applications*, Holden-Day, San Francisco.
- Marple, Jr., S. L. (1987) *Digital Spectral Analysis with Applications*, Prentice-Hall, Englewood Cliffs, New Jersey.
- Otnes, R. K., and Enocson, L. (1978) *Applied Time Series Analysis*, Wiley-Interscience, New York.
- Roberts, D. H., Lehar, J., Dreher, J. W. (1987) *Astrophys. J.* **93**, 968.
- Ryle, M. and Hewish, A. (1960) *Mon. Notices Roy. Astron. Soc.* **120**, 220.
- Scargle, J. D. (1982) *Astrophys. J.* **263**, 835.
- Terebizh, V. Yu. (1992) *Time Series Analysis in Astrophysics*, Moscow, Nauka.
- Thompson, A. R., Moran, J. M., Swenson, G. W. Jr. (1986) *Interferometry and Synthesis in Radio Astronomy*, John Wiley & Sons, New York.
- Vityazev, V. V. (1994) *Astron. Astrophys. Trans.* **5**, 1.
- Walker, G. (1914) *Calcutta Ind. Met. Mem.* **21**, 9.