Bull. Astron. Belgrade № 145 (1992), 17 – 24.

UDC 520:51-72;-73
Original scientific paper

DISCRETE FOURIER TRANSFORMS OF THE SERIES WITH RANDOMLY MISSED DATA: EXPERIMENTAL TESTS

D. Djurović

Department of Astronomy, Studentski trg 16, YU-11000 Belgrade, Yugoslavia

(Received: January 10, 1991)

SUMMARY: The errors of the period (P), the amplitude (A) and the phase (Ph) of the sinusoidal signal, computed by discrete Fourier transforms (DFT) formula, applied to the series with randomly missed data, have been examined.

In the case of the uniform probability distribution of missed data, the error of P is negligible, while the error of A is increasing function of A and number of missed data N. For N given the ratio $q = \sigma_A/A$ (σ_A-st.dev. of A) is constant. In fact $q = q(N)$. Consequently, the phase standard deviation (σ_{Ph}) is indepependent on A.

1. INTRODUCTION

The Fourier transforms (FT) of a function $x = x(t)$ into the function $X = X(f)$, where the variables t and f are usually considered as time and frequency, are described and discussed in many books and scientific papers (see, for example, Bracewall, 1958; Blackman and Tukey, 1958; Deeming, 1975; etc). In this paper, however, it will be supposed that a reader is familiar with the theory of the FT.

In the cases when $x(t)$ is presented in the form of the equally spaced data series $x_j = x(t_j)$, the consequences of the practical restrictions, as a limitation of the interval of t (leakage) or a digitalisation of the function (aliasing), can be theoretically predicted. However, when the series $\{x_j\}$ is composed of unequally spaced data that is more complicated (see, for example, Bloomfield, 1970; Vanicek, 1971; Meisel, 1978; etc), sometimes-impossible.

Let us remember that the leakage limits the frequency resolution of the FT method, the aliasing enlarges the computed amplitude $A(f)$ at the given frequency f. So, instead of the true amplitude $A(f)$ we obtain:

$$A_C(f) = A(f) + \sum_{k=1}^{\infty} A(f \pm 2kf_N) \qquad (1)$$

where f_N is known as Nyquist frequency. Since $f \in [-f_N, f_N]$, the frequencies $f + 2kf_N \geq f_N$ and $f - 2kf_N \leq -f_N$.

Consequently, if the elementary interval $\Delta t = 1/2f_N$ is so small that f_N is larger than the frequency of any component really existing in the polyharmonic process $x(t)$, each $A(f \pm 2kf_N) = 0$. Therefore, their sum in (1) should be zero or the amplitude $A_C(f)$

would be free of aliasing.

When a series $\{x_j\}$ is already given it can be impossible to choose f_N sufficiently large to avoid the aliasing, but it is important to know whether $A_C(f) = A(f)$ or $A_C(f) \neq A(f)$. For that the knowledge of the upper frequency component in $x(t)$ is needed.

In the discussion the above it is imposed that $A_C(f)$ represents a mathematical expectation of the computed amplitude.

In practice, instead of the true Fourier transform $X(f)$ we compute the function $X_n(f)$ defined by the formula:

$$X_n(f) = \sum_{j=1}^{n} x(t_j)e^{-i2\pi f t_j} \tag{2}$$

Formally, $X_n(f)$ can also be represented as:

$$X_n(f) = \int_{-\infty}^{+\infty} g_n(t)x(t)e^{-i2\pi f t}dt, \tag{2a}$$

where :

$$g_n(t) = \delta(t - t_j),$$

$\delta(\tau)$ being well known Dirac's function.

Since the Fourier transform of the product $g(t)x(t)$ in (2a) is the convolution of their Fourier transforms $G_n(f)$ and $X(f)$ (respectively), we can write:

$$X_n(f) = G_n(f) * X(f) \tag{3}$$

where:

$$G_n(f) = \sum_{j=1}^{n} e^{-i2\pi f t_j} \tag{4}$$

$G_n(f)$ is called the spectral window.

The formulae (2)-(4) are valuable for equally as well as for unequally spaced t.

In the cases when t_j are equally spaced, the sum $G_n(f)$ does not depend on t_j:

$$G_n(f) = \frac{\sin(n\pi f\Delta t)}{\sin(\pi f\Delta t)} e^{-i(n+1)\pi f\Delta t} \tag{5}$$

where $t_j = j\Delta t$. So, for f given the amplitude and phase modulations can be predicted.

In the cases when t_j are unequally spaced that is impossible because $G_n(f)$ depends on the t_j distribution. In practice, however, we are often obliged to decide between the reconstruction of the missed data (by the interpolation and smoothing techniques), with a risk to create the pseudo-harmonic components or to modify the amplitudes and the phases of the components really existing in the process $x = x(t)$, and the computation of the periodograms with the given original data. The last procedure introduces the errors whose estimation represents the task of our present work.

In this paper we have examined the errors of the computed period, amplitude and phase of the sinusoidal signal only for the case of a uniform probability distribution of missed data.

2. DFT PERIODGRAM DISTORSIONS FREE OF ALIASING

Let the series $\{x_j\}, j = 1, 2, ..., 1000$, be computed by the formula:

$$x_j = A\sin(2\pi fj + \frac{\pi}{3}) \tag{6}$$

for all combinations of $A = 1, 10, 100$ and $P = 1/f = 5, 10, 20, 25, 50, 100$. Beside $\{x_j\}$, the series of N different random integers $I_k \in [1, 1000], k = 1, 2, ..., N$, with a uniform probability distribution have also been computed.

From a given series $\{x_j\}$ the data having an index $j = I_k (k = 1, 2, ..., N)$ are missed. So, the series with $1000 - N$ unequally spaced data (hereafter: incomplete series) are obtained. These series are transformed by discrete Fourier transforms (DFT) formula.

Let P_C, A_C and Ph_C be the period, the amplitude and the phase corresponding to the primary maximum in the periodogram.

The computation programme,tested on the complete series, gives P_C, A_C and Ph_C equal to P, A and Ph, respectively.

To study the stability of P_C, A_C and Ph_C when a composition of N missed data changes, for fixed A, P and N, 30 different series of random integers have been generated and the periodograms of incomplete series have been computed. So for A, P and N given we disposed the set of 30 values P_C, A_C and Ph_C. These data are used for the computation of the standard deviations σ_P, σ_A and σ_{Ph}, respectively.

The above computations have been performed for each combination P, A, and $N = 100, 200, ..., 900$.

From the results obtained we have remarked the following:

1. In the limits of the computation accuracy (± 0.05 of the corresponding units) the periods P_C are equal to the corresponding exact values ($\sigma_P = 0$). In the other words, the bias $\Delta P = P - P_C$ and random fluctuations of P_C are negligible.

2. The amplitudes A_C are biased with respect to A (Table 1). The bias $\Delta A = A - A_C$ is independent neither on P nor on N. Small differences of ΔA for three given amplitudes are now unsufficient for any valuable discussion on the mutual dependence between ΔA and A, but as it will be seen later on this dependence is not assumed.

Table 1
The difference $\Delta A = A - A_C$
for $A = 1, 10, 100$: mean values over the set of $N = 100, 200, ..., 900$.

A	1	10	100
P			
5	0.4	0.7	0.8
10	0.5	0.6	0.9
20	0.4	0.6	0.7
25	0.4	0.6	0.6
50	0.4	0.6	0.9
100	0.4	0.6	0.4

3 . Random fluctuations of A_C of individual series with respect to the mean value over the mentioned set of 30 series (P, A, N fixed) are larger for A and N larger (Table 2). Since they do not depend on P, in the Table 2 mean σ_A for six P analyzed are presented. For $A = 1$ $\sigma_A \leq 0.05$. Therefore, it is negligible with respect to the data accuracy and not presented in the *Table*2. For $A = 100$ the standard deviation σ_A is also presented in Fig.1 as a function of N.

4 . Like the standard deviation σ_A, σ_{Ph} is also independent on P. It's dependence of N is also well pronounced (see the Table 3 and Fig.1, where σ_{Ph} for $A = 100$ is plotted).

Table 2
Standard deviation of $A_C(\sigma_A)$ for $A = 10$ and $A = 100$:
mean values for the set of six P considered. σ_1 - standard deviation of σ_A.

	$A = 10$		$A = 100$	
N	σ_A	σ_1	σ_A	σ_1
100	0.0	0.0	0.7	0.1
200	0.1	0.0	1.1	0.2
300	0.2	0.0	1.4	0.2
400	0.2	0.0	1.7	0.3
500	0.2	0.0	2.2	0.3
600	0.2	0.0	2.8	0.2
700	0.4	0.1	3.6	0.5
800	0.4	0.0	4.1	0.4
900	0.7	0.1	6.3	0.8

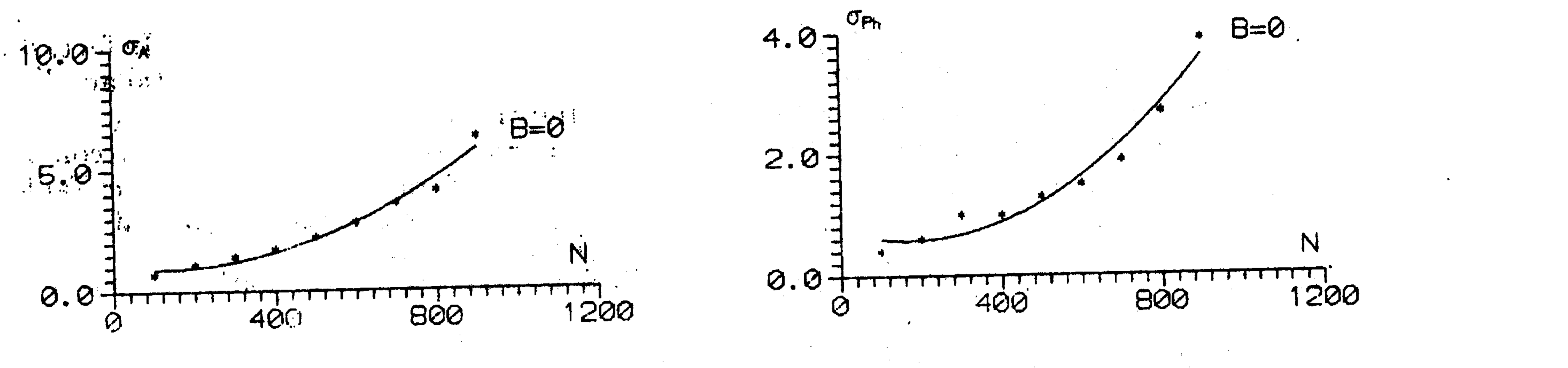

Fig. 1. Standard deviations of the amplitude (σ_A) and the phase (σ_{Ph}) as functions of N. The amplitude A=100.

Table 3

Standard deviation of $Ph_C(\sigma_{Ph})$ for three amplitudes considered above. σ_1 - standard deviation of σ_{Ph}. σ_1 - standard deviation of σ_A.

A:	1		10		100	
N	σ_{Ph}	σ_1	σ_{Ph}	σ_1	σ_{Ph}	σ_1
100	0°.7	0°.1	0°.4	0°.1	0°.4	0°.1
200	1.1	0.1	0.7	0.1	0.6	0.1
300	1.4	0.2	0.8	0.2	1.0	0.2
400	1.6	0.2	1.0	0.2	1.0	0.2
500	2.1	0.2	1.3	0.1	1.3	0.2
600	2.5	0.6	1.4	0.2	1.5	0.3
700	2.7	0.4	1.8	0.2	1.9	0.3
800	4.2	0.5	2.6	0.4	2.7	0.4
900	5.9	0.9	3.8	0.2	3.9	0.3

For $A = 10$ and $A = 100$ practically the same σ_{Ph} are obtained. Under the impression of the known fact the presence of the accidental errors in x_j reflects on the phase accuracy (it decreases with amplitude decreasing) the last results could be seen unexpected. However, it concerns an another matter. Since for N given the ratio σ_A/A is practically the same for both amplitudes (this later is supplementarly argumented by the results presented in the Table 4.), it is not surprising that σ_{Ph} for the corresponding N are equal. Namely, from the DFT formula following relation can easily be deduced:

$$\sigma_{Ph} \approx 57^\circ.296\frac{\sigma_A}{A}\left(2-\frac{\pi}{2}\right)^{-\frac{1}{2}}. \qquad (7)$$

For $A = 1$ coresponding σ_{Ph} is larger for a factor $q = 1.6$.

5 . The phase bias is not identified.

3. DFT PERIODGRAM DISTORSIONS IN THE PRESENCE OF THE ALIASING

To study the aliasing effect in the periodograms of the series with randomly missed data (also the case of uniform distribution) the above computations for the series $\{x_j\}$ have completely been repeated for the series $\{y_j\}$ generated by the formula:

$$y_j = 100\sin(2\pi fj + \frac{\pi}{3}) + B\sin(2\pi f'j), \qquad (8)$$

where B =10, 50, 100, 200, 300, 400; P=1/f=5, 10, 20, 25, 50, 100 (as in the precedent paragraph); $P' = 1/f' = P/(P+1)$. Therefore, the frequency $f' = f+1$ is that of the first aliasing term in (1). However, since $\Delta t_j = t_j - t_{j-1}$ randomly fluctuates,the equation (1) is not appropriate for the estimation of $A(f)$. The other one is not known. Because of that for the practical application of DFT on the series with randomly missed data it could be useful to estimate experimentally the errors of P, A and Ph of the signal $x(t)$ in the presence of the high frequency component.

From the periodogram analysis of the incomplete series $\{y_j\}$ the following is remarked:

1 . The periods P_C are again equal to the corresponding exact values ($\sigma_P = 0$), except for the extreme case when $P = 100$ and $N = 900$. In this case $\Delta P = P - P_C$ fluctuates between ± 2 units: standard deviation $\sigma_A = 0.5$. So, the first conclusion from the precedent paragraph is practically confirmed.

2 . For the amplitude bias $\Delta A = A - A_C$ the next results are obtained:

ΔA:	0.2	0.2	0.4	0.5	0.9	0.9
B:	10	50	100	200	300	400

The tendency of ΔA increasing in function of the total amplitude $A_t = A + B$ now seems real, but having also in mind the results from the Table 1 we do not assume that.

3 . The aliasing evidently enlarges the random fluctuations of A_C (Table 4) and Fig. 2. In fact, standard deviation σ_A increases with the total amplitude $A_t = B + 100$ (in precedent paragraph $A_t = A$), so the ratio $q = \sigma_A/A_t$ remaines practically constant. This later follows from the results in the Table 5, where we can see that the standard deviation of the variable q is very small.

Beside that, σ_A, as previously, is independent on P.

4 . The random fluctuations of the phase depend neither on P nor on B. In general, it can be noted that if $A \geq 10$, then σ_{Ph} depends only on N. To illustrate that σ_{Ph} in the Table 6 is presented in function of N and B, while in the Table 7 it is presented as function of N and P.

5 . The phase bias is not identified.

Table 4

Standard deviation σ_A in function of N and B: mean values for the given set of P

B: / N	10	50	100	200	300	400
100	0.8	1.1	1.5	2.2	2.8	3.6
200	1.3	1.7	2.5	3.7	4.8	5.8
300	1.6	2.2	2.8	4.3	6.0	6.7
400	1.9	2.7	3.3	5.3	7.0	8.5
500	2.5	3.5	4.6	7.2	9.1	11.2
600	2.8	3.8	5.2	7.7	10.4	13.8
700	3.7	5.3	6.6	10.9	14.1	17.0
800	5.0	6.7	9.0	13.5	18.8	22.6
900	7.2	10.3	13.1	20.3	26.6	33.6

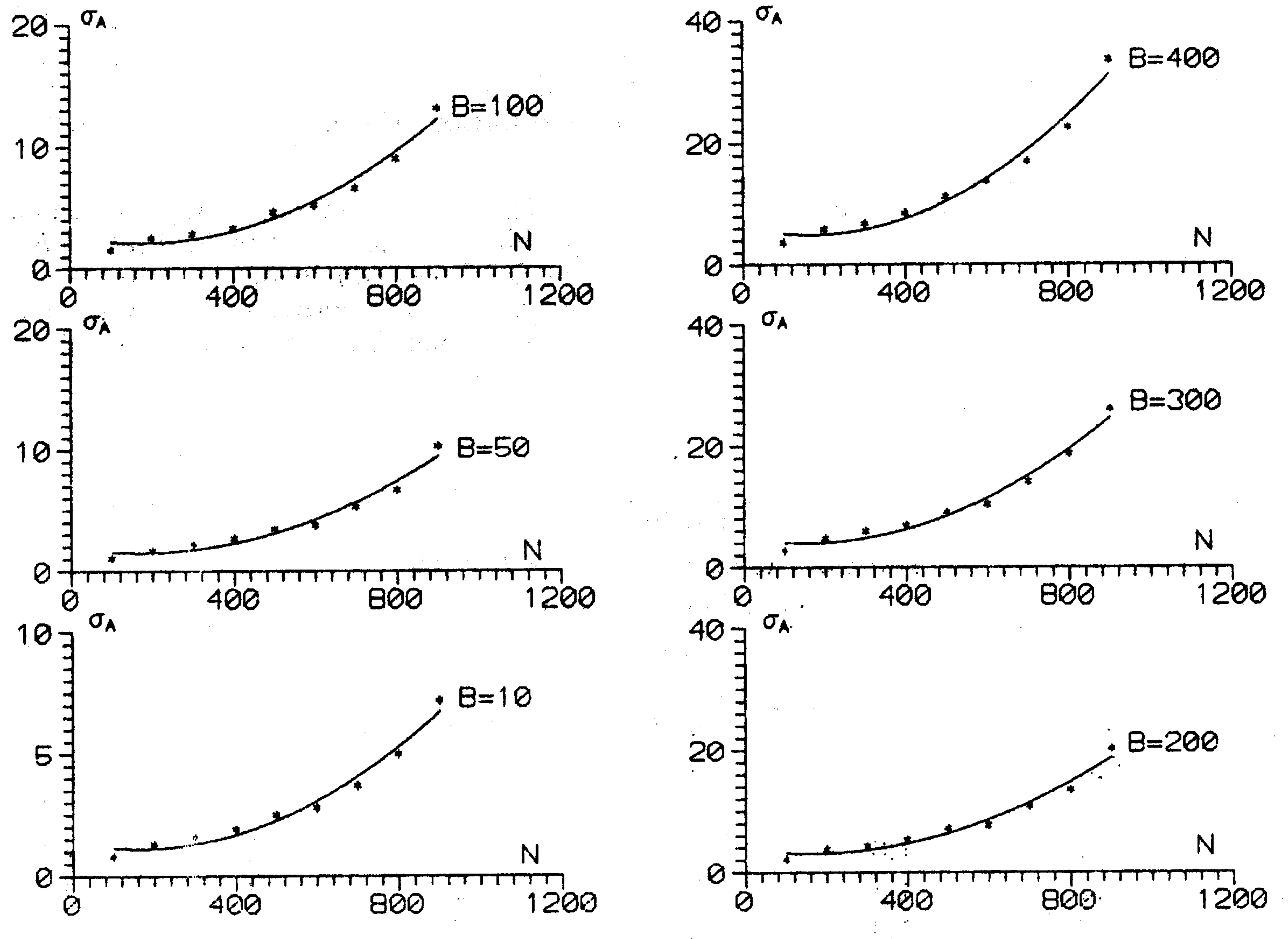

Fig. 2. Standard deviation of the amplitude (σ_A) as functions of N in the presence of the aliasing term whose amplitude is B.

Table 5
Mean values of the $q = \sigma_A/A_t$ (in 0.001 units) for the set of $A_t = 10, 100, 110, 150, 200, 300, 400, 500$.
σ_q-standard deviation of q in the same units.

N	q	σ_q
100	7	0.2
200	12	0.2
300	14	0.3
400	17	0.3
500	23	0.3
600	26	0.5
700	35	0.4
800	45	0.7
900	66	0.8

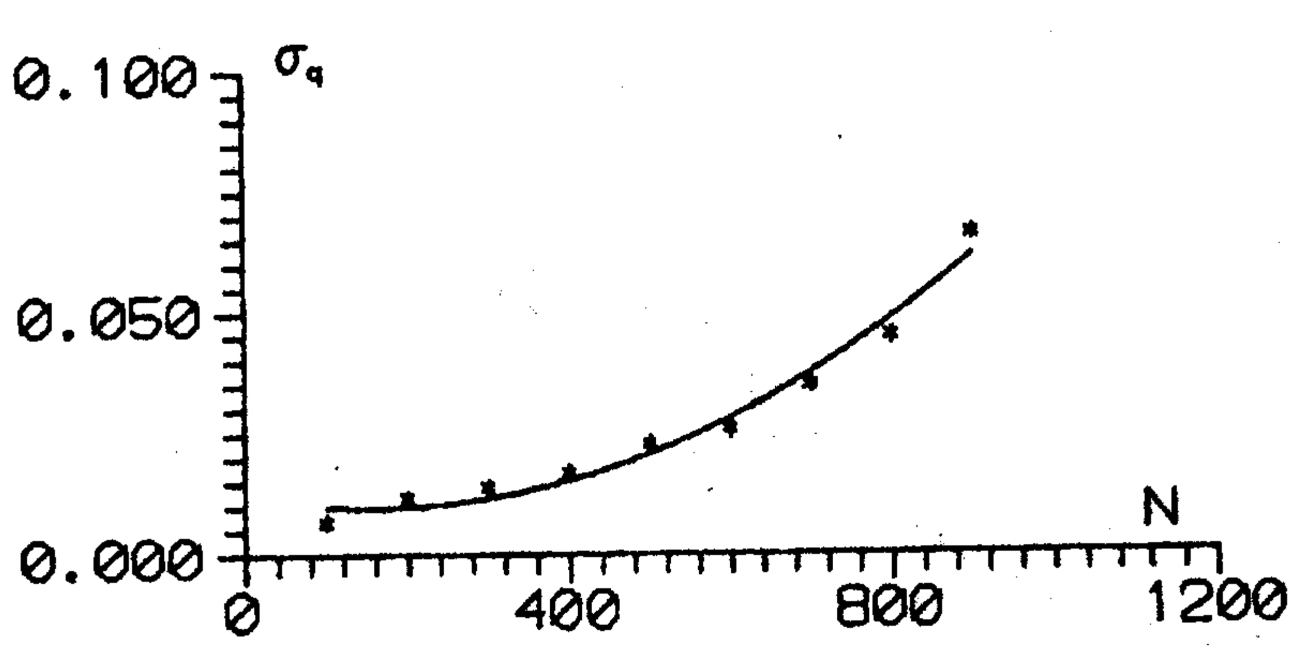

Fig. 3. The ratio $q = \sigma_A/A_t$ as function of N.

6 . Parabolic approximations of σ_A and σ_{Ph} in function of N give a satisfactory results (full lines on Figs 1 and 2).

As a general conclusion, it follows from our work the ratio $q = \sigma_A/A_t$ varies only in function of N. The consequence of that the phase errors also depend only on N. The second order polynomial approximation of q could be sufficiently good (Fig. 3).

The volume of our computations,devoted to the problem of the errors of P, A and Ph, computed by the DFT formula from the series with unequally spaced data, is not sufficient to obtain the more general conclusions, but we assume the estimations of

Table 6
Standard deviation σ_{Ph} for $P = 50$ in function of N and B.

B: N	10	50	100	200	300	400
100	0°.5	0°.5	0°.4	0°.5	0°.5	0°.5
200	0.6	0.6	0.6	0.6	0.6	0.6
300	1.0	1.0	1.0	1.0	1.0	1.0
400	1.0	1.0	1.0	1.0	1.0	1.0
500	1.3	1.3	1.3	1.3	1.3	1.3
600	1.3	1.3	1.3	1.2	1.3	1.3
700	2.2	2.1	2.2	2.2	2.2	2.2
800	2.5	2.5	2.5	2.5	2.4	2.4
900	3.3	3.3	3.3	3.3	3.3	3.3

Table 7
Standard deviation σ_{Ph} in function of N and P: mean values for the given set of B

P: N	5	10	20	25	50	100
100	0°4	0°4	0°4	0°5	0°4	0°4
200	0.6	0.7	0.7	0.7	0.6	0.7
300	0.8	0.7	0.8	0.7	1.0	0.8
400	1.1	1.1	1.1	1.0	1.0	1.0
500	1.4	1.3	1.3	1.3	1.3	1.2
600	1.4	1.3	1.5	1.6	1.5	1.6
700	2.0	1.6	2.1	2.1	2.1	2.0
800	2.4	2.7	2.4	2.4	2.6	2.6
900	3.6	3.8	4.0	3.9	3.4	6.5

the errors we delt above could be helpful when one treats the unequidistant data series.

Acknowledgments – The author would like to express his appreciation to Dr J. Mališić, prof. of the Mathematical Faculty, for some precious suggestions and corrections of the manuscript

This work has been supported by the Republic Fund for science in Serbia through the project "Physic and Motions of Celestial Bodies".

REFERENCES

Blackman, R.B. and Tukey J.W.: 1959, *The Measurement of Power Spectra from the Point of View of Communications Engineering*, Dover, New York

Bloomfield, D.: 1970, *J. R. Stat. Soc., Ser*, **B32**,, 369

Bracewell, R.: 1965, *The Fourier Transform and its Applications*, New York

Deeming, T. J.: 1975, *Astrophys.and Space Sci.*, **36**, 137

Meisel, D. D.: 1978, *The Astron. J.*, **83**, **5**, 538

Vanicek, P.: 1971, *Astrophys. and Space Sci.* **12**, 10

ДИСКРЕТНЕ ФУРИЈЕОВЕ ТРАНСФОРМАЦИЈЕ СЕРИЈА СА СЛУЧАЈНО ИЗОСТАВЉЕНИМ ПОДАЦИМА: ЕКСПЕРИМЕНТАЛНИ ТЕСТОВИ.

Д. Ђуровић

Институт за астрономију, Студентски трг 16, Београд, Југославија

УДК 520: 51-72;-73
Оргинални научни рад

Испитиване су грешке периоде, амплитуде и фазе синусног сигнала рачунате помоћу формула за дискретне Фуријеове трансформације серија са случајно изостављеним подацима.

У случају униформне расподеле вероватноће изостављених података, грешка периоде је занемарљива, док је грешка амплитуде растућа функција од A и броја изостављених података N. За дато Н, количник $q = \sigma_A/A$ (σ_A-станд. дев. A) је константан. Уствари, $q = q(N)$. Последица тога је да стандардна девијација фазе (σ_{Ph}) не зависи од A.