Mon. Not. R. astr. Soc. (1991) **253**, 198–206

Accuracy of period determination

A. Schwarzenberg-Czerny★
Department of Astronomy, University of Cape Town, Private Bag, 7700 Rondebosch, South Africa

Accepted 1991 June 13. Received 1991 June 7; in original form 1991 February 11

SUMMARY

Periods of oscillation are frequently found using one of two methods: least-squares (LSQ) fit or power spectrum. Their errors are estimated using the LSQ correlation matrix or the Rayleigh resolution criterion $\delta\nu_R = 1/\Delta T$, respectively. In this paper we demonstrate that both estimates are statistically incorrect. On the one hand the LSQ covariance matrix does not account for correlation of residuals from the fit. Neglect of the correlations may cause large underestimation of the variance. On the other hand the Rayleigh resolution criterion is insensitive to signal-to-noise ratio and thus does not reflect quality of observations. We derive the correct variance estimates for the two methods. In the process we demonstrate that centre of the power spectrum line is a maximum likelihood estimate of frequency of the oscillation and demonstrate it is statistically equivalent to fitting of a sinusoid by LSQ, so the methods are statistically equivalent. Our new and correct variance estimate is quite simple and practicable. It is using the autocorrelation function (ACF) of the residuals to determine their mean correlation length and is valid under certain assumptions. We tested the extent to which the assumptions may be relaxed by numerical simulations.

1 INTRODUCTION

In the first paper of this series (Schwarzenberg-Czerny 1989), attention was focused on detection of coherent oscillations among noise, and testing its statistical significance. In the present paper we assume that the oscillation has already been detected. We discuss critically classical methods for determination of the value of its period. We point out statistical inconsistencies in them. The inconsistencies mainly affect the error estimate and only rarely the period value itself. So we focus our attention on error estimation. Practical experience shows that current error estimates are rarely, if ever, reliable. They are certainly inconsistent with the scatter of measured period values. We pay special attention to methods for determination of periods from a power spectrum line and by LSQ fitting a sinusoid, because of their great practical value. Our interest is also motivated by their apparently contradictory properties. Namely, on the one hand Lomb (1976) and Scargle (1982) found that the two methods use the same statistic. On the other hand, in their currently used form they yield mutually inconsistent accuracy estimates. The LSQ variance, based on the inverse normal equations matrix is sensitive to the signal-to-noise (S/N) ratio. The power spectrum resolution criterion, due to Rayleigh is independent of the S/N ratio. Based on sound statistical principles, we analyse these methods anew and demonstrate that they are indeed statistically equivalent. Thus they yield same period estimate and with the same accuracy. However, the true accuracy turns out to be inconsistent with either of the afore-mentioned classical estimates.

An important role in our analysis is played by correlation of noise in observations. What really matters is the correlation of residuals from the fit of a sinusoid performed either explicitly by LSQ or implicitly by calculation of power spectrum. The correlation of residuals may arise for various reasons. Namely, (i) the physical process under study may produce noise with some degree of correlation (e.g. flickering in Cataclysmic Variables and X-ray Binaries), (ii) imperfect measurements may introduce further correlations (e.g. atmospheric transparency variations) and (iii) the assumed mathematical model is a poor fit (e.g. a sinusoid for Cepheid or pulsar light curves). Most observers are quite weary of the type (ii) effects and strive to avoid them. Since all physical systems have limited frequency response band, none is quite free of the type (i) effect, although its role might be negligible in certain cases. Actually, observations of the effect give an extra information on the involved physics. The type (iii) effects are by far the commonest cause of correlation in the residuals. Also, observers are mostly ignorant of their role.

Most textbooks on statistics explain how to take into account correlation of observations in the least-squares analysis and essentially we could finish the first part of this paper by stating so. However, the author is not aware of observations published in a refereed astronomical journal in which correlation of observations was explicitly taken into account in period determination by LSQ. Evidence exists to

★ On leave from Warsaw University Observatory, Poland.

support the dim suspicion that in no cases is the correlation accounted for. In Section 2 we discuss formulation of the LSQ method for both correlated and uncorrelated observations. We restrict our discussion to observations depending on one independent variable, e.g. time. In Section 3 we discuss effects of the correlation on results of period analysis and consequences of their neglect. For this purpose we use analytic considerations. As a particular case we discuss period determination by LSQ fitting of a sinusoid.

Because of affinity between power spectrum and the LSQ fit of a sinusoid (Lomb 1976; Scargle 1982) we turn our attention to power spectrum analysis too. First we consider as an academic example a maximum likelihood method for period determination from a power spectrum. From properties of this method we draw several conclusions of practical importance. In Section 4 we outline practical methods for estimation of periods and their variance. We tested our ideas by Monte Carlo simulations discussed in Section 5.

1.1 Least-squares analysis for correlated and uncorrelated data

The least-squares analysis is sensitive to correlation of observations noise ('the residuals from the fit'). The most commonly used LSQ algorithm is valid only for Gaussian white noise residuals with the covariance matrix $\mathbf{C_x}=\sigma_x^2\mathbf{I}$.

In order to establish notation we rederive the formulae for the LSQ method. In a general case of correlated data, the parameters $\boldsymbol{y}$ of a function $f_t(\boldsymbol{y})=f(t,\boldsymbol{y})$ are fitted so that the sum of squares of the residuals is minimal (Eadie *et al.* 1971):

$$min.=Q^2=[\boldsymbol{x}-\boldsymbol{f}(\boldsymbol{y})]^\dagger\mathbf{C_x}[\boldsymbol{x}-\boldsymbol{f}(\boldsymbol{y})]. \tag{1}$$

It is convenient to expand the vector function $\boldsymbol{f}(\boldsymbol{y})$ into Taylor series around a point $\boldsymbol{y}_0$ close enough to the solution. Retaining only first and second terms we get

$$\boldsymbol{f}(\boldsymbol{y})=\boldsymbol{f}(\boldsymbol{y}_0)+\frac{\partial\boldsymbol{f}(\boldsymbol{y}_0)}{\partial\boldsymbol{y}}(\boldsymbol{y}-\boldsymbol{y}_0).$$

After translation of origins in $\boldsymbol{x}$ and $\boldsymbol{y}$ spaces to $\boldsymbol{f}(\boldsymbol{y}_0)$ and $\boldsymbol{y}_0$, respectively, we obtain:

$$min.=Q^2=(\boldsymbol{x}-\mathbf{A}\boldsymbol{y})^\dagger\mathbf{C_x}(\boldsymbol{x}-\mathbf{A}\boldsymbol{y}), \tag{2}$$

where matrix

$$A_{ij}=\frac{\partial f_i(\boldsymbol{y}_0)}{\partial y_j},$$

called a design matrix depends on a given problem. The linear expansion is accurate enough as long as $\boldsymbol{f}(\boldsymbol{y})$ is approximately linear for 1σ changes in $\boldsymbol{y}$. When this is not the case, the LSQ estimate is no longer optimum and unbiased. Then the maximum likelihood solution should be used instead of the LSQ method.

The solution of equation (2) is obtained in a standard way be equating gradient of Q^2 to zero. Its covariance matrix follows from the change of variables theorem:

$$\boldsymbol{y}=(\mathbf{A}^\dagger\mathbf{C_x}^{-1}\mathbf{A})^{-1}\mathbf{A}^\dagger\mathbf{C_x}^{-1}\boldsymbol{x} \tag{3}$$

$$\mathbf{C_y}=(\mathbf{A}^\dagger\mathbf{C_x}\mathbf{A})^{-1}. \tag{4}$$

Neglecting the correlation of residuals leads to a solution and its covariance matrix which are both incorrect. They are obtainable from equation (4) by substitution of $\sigma_x^2\mathbf{I}$ for $\mathbf{C_x}$:

$$\boldsymbol{y}^{(1)}=(\mathbf{A}^\dagger\mathbf{A})^{-1}\mathbf{A}^\dagger\boldsymbol{x} \tag{5}$$

$$\mathbf{C_y}^{(2)}=\sigma_x^2(\mathbf{A}^\dagger\mathbf{A})^{-1}. \tag{6}$$

Here, $\sigma_x^2=Var\{x_t\}$ denotes the variance of the noise. The solution and their errors which are used in literature correspond to $\boldsymbol{y}^{(1)}$ and $\mathbf{C_y}^{(2)}$. They do differ from the correct ones given in equation (4). Later we show that the difference between the solutions is of little consequence. However, difference of the correlation matrices is profound. More, $\boldsymbol{y}^{(1)}$ and $\mathbf{C_y}^{(2)}$ are mutually inconsistent. If we treat $\boldsymbol{y}^{(1)}$ as a certain linear function of $\boldsymbol{x}$, then its covariance matrix according to the change of variables theorem is (e.g. Eadie *et al.* 1971):

$$\mathbf{C_y}^{(1)}=(\mathbf{A}^\dagger\mathbf{A})^{-1}\mathbf{A}^\dagger\mathbf{C_x}\mathbf{A}(\mathbf{A}^\dagger\mathbf{A})^{-1}. \tag{7}$$

It is $\mathbf{C_y}^{(1)}$ rather than $\mathbf{C_y}^{(2)}$ which is the correlation matrix of $\boldsymbol{y}^{(1)}$. Note that Toeplitz $\mathbf{C_x}$ does not guarantee that $\mathbf{C_y}$ is Toeplitz. The asymmetry of $\mathbf{C_y}$ reflects the asymmetry of $\mathbf{A}$ and it stems from asymmetry of parameters of the fitted curve.

The case of perfectly correlated groups of D observations each is easy to understand. To account for correlation it is enough to use only one observation from each group, set $\mathbf{C_x}=\sigma_x\mathbf{I}$ and substitute into equation (6). Because the omitted observations contain no additional information, no increase of covariance $\mathbf{C_y}$ is incurred in this way. The used observations are uncorrelated so the results are correct and correspond to $\boldsymbol{y}$ and $\mathbf{C_y}$. Algebraically skipping observations corresponds to division of both $\mathbf{A}^\dagger\mathbf{A}$ and $\mathbf{A}^\dagger\boldsymbol{x}$ by D. Comparing results, what we see is that no error in the solution is produced by neglect of correlation but covariance $\mathbf{C_y}^{(2)}$ is underestimated by a possibly large factor of D:

$$\boldsymbol{y}=\boldsymbol{y}^{(1)}=(\mathbf{A}^\dagger\mathbf{A})^{-1}\mathbf{A}^\dagger\boldsymbol{x} \tag{8}$$

$$\mathbf{C_y}=\mathbf{C_y}^{(1)}=\sigma_x D(\mathbf{A}^\dagger\mathbf{A})^{-1} \tag{9}$$

$$\mathbf{C_y}^{(2)}=\sigma_x(\mathbf{A}^\dagger\mathbf{A})^{-1}. \tag{10}$$

One can expect similar results for a general signal with finite correlation length. In the next section we demonstrate that it is so indeed. Summarizing, both least-squares solutions $\boldsymbol{y}$ and $\boldsymbol{y}^{(1)}$ are equivalent and have the same covariance matrices $\mathbf{C_y}=\mathbf{C_y}^{(1)}$. Thus *neglect of correlation of noise does not affect the values of the least-squares solution*, as long as our assumptions hold. However, the error estimates suffer severely from the neglect. The covariance matrix which is used in literature for period estimation, $\mathbf{C_y}^{(1)}$ is smaller from the true correlation matrix $\mathbf{C_y}$ by a possibly large factor D. Thus *neglect of the correlation causes underestimation of errors by a factor equal to a mean number of observations with correlated noise.*

2 NOISE WITH A SMALL CORRELATION LENGTH

2.1 Evenly spaced observations

Let us for the rest of this paper assume that observations depend only on one independent variable – time. In a more general case of correlations in observations obtained in multi-dimensional space of independent variables, one has to consider number of correlated observations in adjacent

volume. Further, we assume that the noise is Gaussian, i.e. that its component n_t is normal random variable, $E\{n_t\}=0$ $Var\{n_t\}=1$ for any fixed time t. We shall consider here two kinds of noise. The white noise $\boldsymbol{w}$ has identity covariance matrix $\mathbf{C_w}\equiv E\{\boldsymbol{w}^\dagger\boldsymbol{w}\}=\mathbf{I}$. Any stationary noise process has Toeplitz covariance matrix, i.e. a Hermitian matrix homeomorphic to the autocovariance vector $\boldsymbol{a}$: $C_{n\,ij}=a_{i-j}=a^\dagger_{j-i}$. The stationary noise with finite correlation length D has band covariance matrix, with band width D. Each noise can be represented by filtered white noise. Its covariance matrix also depends on the filter. In matrix notation:

$$\boldsymbol{n}=\mathbf{B}\boldsymbol{w} \tag{11}$$

$$\mathbf{C_n}=\mathbf{B}\mathbf{B}^\dagger, \tag{12}$$

where matrix $\mathbf{B}$ defines the filter. The covariance matrix has no more symmetry than its filter matrix, except for its Hermitian property.

We shall denote the time interval spanned by observations by T, the characteristic length of observed signal (e.g. period) by L and the noise correlation length by D. We shall consider only an asymptotic case of short correlation length

$$T\gg L\gg D\gg 1. \tag{13}$$

In any realistic case, matrix $\mathbf{A}$ is rectangular and has no symmetry at all. The covariance matrices $\mathbf{C_y}$, $\mathbf{C_y^{(2)}}$ and $\mathbf{C_y^{(1)}}$ differ by powers of $\mathbf{C_x}$. On the one hand, only $\sim D$ terms in each row of $\mathbf{C_x}$ are large, and so only $\sim D$ terms of column of $\mathbf{A}$ contribute to each product. On the other hand, since we assumed $\mathbf{A}$ varies only on length L large compared to D, all elements of $\mathbf{A}$ which contribute are nearly the same:

$$(\mathbf{C_xA})_{ij}\approx\left[\sum_k (C_x)_{ik}\right]A_{ij}. \tag{14}$$

For circular $\mathbf{C_x}$ all sums in the square brackets are the same. For Toeplitz $\mathbf{C_x}$, $\mathbf{C_xA}=\lambda\mathbf{A}$. If all observations are positively correlated $\lambda=D$. Very much the same applies to multiplication by an inverse $\mathbf{C_x^{-1}}$, except that now the factor is $1/D$, so that the product by $\mathbf{C_x^{-1}C_x}$ returns $\mathbf{A}$.

A more strict derivation of these factors is possible. For lack of space we sketch it only. It follows from the derivation of equation (14) that the factors do not depend on $\mathbf{A}$. Also Toeplitz $\mathbf{C_x}$ can be approximated by a circular matrix differing from it only in corners, with no large effects on the matrix products considered. Then, circular matrices $\mathbf{C}$ form a group with matrix multiplication which is homomorphic to the group of corresponding vectors $\mathbf{c}$ with convolution or to the one of their Fourier transforms $\mathcal{F}\boldsymbol{c}$ with common multiplication. Assuming Gaussian filter b of width D for the noise, one can easily perform all calculations in the $\mathcal{F}\boldsymbol{c}$ group and then convert results back to matrices. The results are the same as in equation (9). Some attention must be paid to the normalization of b since the variance of x, σ_x^2, is known to the observer and must be kept fixed. Equation (9) is valid for a general asymmetric and rectangular design matrix obeying restriction on length scale given by equation (13). Although strictly speaking they are valid only for circular $\mathbf{C_x}$, for $T\gg D$ deviations of $\mathbf{C_x}$ from circular symmetry are so small that to a good approximation can be neglected. Thus our conclusions may be extended on a general case of stationary noise since its covariance matrix is Toeplitz.

2.2 Unevenly spaced observations

Under certain conditions the results of the previous section are extendable for the case of observations unevenly distributed in time. They depend on convolutions of rows of $\mathbf{C_x}$, i.e. of shifted ACF, with columns of $\mathbf{A}$. The convolutions can be represented by integrals. The results hold as long as the distribution of observations in time is such that the ACF can be found (Edelson & Krolik 1988) and the products of rows of $\mathbf{C_x}$ times columns of $\mathbf{A}$ are good Monte Carlo approximations of the corresponding integrals. We shall not expand this topic in this paper as it belongs to the Monte Carlo integrals theory. Clearly the conditions are not satisfied, e.g. when the average separation of observations is larger than the correlation length D or when there is a pattern in the distribution of the observations at the length comparable to the characteristic length L of the fitted curve. The correlation length D now corresponds to the mean number of consecutive correlated observations.

3 ON VARIANCE OF PERIOD ESTIMATES

3.1 The least-squares period estimate

Let us now consider a specific case but with important applications, namely that of fitting by LSQ of the periodic function of time $s(\Omega t)$, such that

$$s(\phi+1)\equiv s(\phi)=\sum_{i=2}^{I} y_i\varphi_i(\phi) \tag{15}$$

and all trial functions φ, are orthogonal

$$\int_{-\infty}^{+\infty}\varphi_i(\phi)\varphi_j(\phi)\,d\phi=\delta_{ij}. \tag{16}$$

It is convenient to introduce new notation for the oscillation frequency $\Omega=y_1$ and for the normal equations matrix $\boldsymbol{\Theta}=\mathbf{A}^\dagger\mathbf{A}$. Applying the results of the previous section we find that the variance of Ω is

$$Var\{\Omega\}\equiv C_{y11}=\sigma_x^2 D(\boldsymbol{\Theta}^{-1})_{11}. \tag{17}$$

Since the design matrix $\mathbf{A}$ is built of derivatives of s so we obtain the elements of $\boldsymbol{\Theta}$ by differentiation and summation:

$$\Theta_{11}=\sum_t t^2\left(\frac{\partial s_t}{\partial\phi}\right)^2=\left(\sum_t t^2\right)\left\langle\left(\frac{\partial s_t}{\partial\phi}\right)^2\right\rangle\approx\frac{T^3}{3}\left\langle\left(\frac{\partial s_t}{\partial\phi}\right)^2\right\rangle, \tag{18}$$

$$\Theta_{1i}=\sum_t t\frac{\partial s_t}{\partial\phi}\varphi_{ti}=\left(\sum_t t\right)\left\langle\frac{\partial s_t}{\partial\phi}\varphi_i\right\rangle\approx 0, \tag{19}$$

$$\Theta_{ij}=\sum_t \varphi_{ti}\varphi_{tj}=\langle\varphi_i\varphi_j\rangle\approx\delta_{ij}, \tag{20}$$

$$i,j=2,\ldots,I, \tag{21}$$

where δ is Kronecker symbol and $\langle,\rangle$ indicates time average. Note that we replaced variables of differentiation in the design matrix $ds/d\Omega=t(ds/d\phi)$. Both s and φ are periodic and t does not change much over one period $P\approx L$ because of our assumption $T\gg L$ (equation 13). Therefore we dragged their average values outside the sums over time.

Here we assume that the distribution of observation is such, that we can replace sums by the corresponding integrals. This is a rather weak assumption, amounting in essence to requiring that moments of observations are not correlated with the phase of oscillations under study. It is not necessary to obtain complete phase coverage during any single cycle of observations. In fact, less than one observation per cycle may be still acceptable (Schwarzenberg-Czerny 1989). Before we proceed to evaluation of the inverse matrix element Θ_{11} let us note that units of elements of $\mathbf{A}$ and $\boldsymbol{\Theta}$ are not homogeneous. So their relative magnitudes depend on the choice of units. Thus we must define our units before proceeding any further. It is also clear from the form of equation (19) that the mean time of observation $\langle t \rangle$ is convenient origin of time, so that $\boldsymbol{\Theta}_{1i}$ vanishes. Mean separation of observations must be taken for the unit of time so that D corresponds to the mean number of correlated observations, as before. It is convenient to adopt the amplitude of the oscillation for the unit of signals s, x and y_i, $i=2,\ldots,I$. Thus the following estimates for the averages hold

$$\left\langle \left(\frac{\partial s}{\partial \phi} \right)^2 \right\rangle \approx 1, \tag{22}$$

$$\langle \varphi_i \varphi_j \rangle \approx \delta_{ij} \quad \text{orthonormality,} \tag{23}$$

$$\left\langle \frac{\partial s}{\partial \phi} \varphi_i \right\rangle \leq \sqrt{\left\langle \left(\frac{\partial s}{\partial \phi} \right)^2 \right\rangle \langle \varphi_i^2 \rangle} \approx 1 \quad \text{Schwarz inequality,} \tag{24}$$

for all ss except pathological ones with steep gradients and many discontinuities.

With our particular choice of units and trial functions, the matrix $\boldsymbol{\Theta}$ turns out to be nearly diagonal, so that its approximate inversion is simple. Thus the parameters y_i are nearly uncorrelated. Then the variance of the frequency estimate is

$$Var\{\Omega_{\mathrm{LSQ}}\} = \frac{3\sigma_x^2 D}{T^3 \langle (\partial s/\partial \phi)^2 \rangle} \approx \frac{3\sigma_x^2 D}{T^3}. \tag{25}$$

The last equality corresponds to our particular choice of signal units and a smooth s. We stress once again importance of symmetric choice of the time origin in keeping period and epoch uncorrelated and their variances small. This is a particular case of orthogonalization of variables.

3.2 The maximum likelihood period estimation from power spectrum

A signal containing periodic oscillation produces a feature in its power spectrum or 'a spectral line'. The half-width of the line $\delta\Omega_{\mathrm{R}} \approx 2\pi/T$ is a measure of power spectrum resolution. It corresponds to the change in frequency producing half period phase shift over the whole interval of observation. Except for a constant factor it is equivalent to Rayleigh resolution criterion. Comparison with equation (25) demonstrates readily that the Rayleigh resolution overestimates period variance by a large factor. In fact it corresponds to a frequency interval in which the value of sum of squares χ^2 is less than half of one for a spurious frequency. More importantly, the Rayleigh criterion or any similar based on fixed phase shift is independent of signal-to-noise ratio and thus does not reflect quality of the data.

Lomb (1976) and Scargle (1982) demonstrated that LSQ and power spectrum methods use equivalent statistics. However, they produced no practical methods for exploiting the equivalence. Henceforth we shall call Scargle (1982) modified power spectrum simply power spectrum. In this section we shall derive the maximum likelihood method (MLM) for period estimate from power spectrum. We shall demonstrate that this method is equivalent to LSQ fitting of a sinusoid.

We apply the MLM to estimation of period from power spectrum $\boldsymbol{p}$, treated as 'observations'. It is convenient for the present purpose to consider $\boldsymbol{p}$ as a vector of components $p(\omega)$. The components are numbered by frequency ω, a possibly continuous independent (i.e. non-random) variable. Let $f(p,\boldsymbol{y})$ be the probability distribution of p and $\boldsymbol{y}$ its parameters, such as $\omega, t; \sigma^2, \Omega, (S/N), \ldots$ Some of the parameters are not random variables, being either known in advance or specified as arguments, e.g. times of observation t or current frequency ω. They are not parameters in the sense used in the estimation theory and we shall call them independent variables. To stress their role we shall indicate them explicitly, if desirable. The aim of estimation is to find values of other dependent parameters treated as random variables, e.g. power of noise σ^2 or frequency and S/N ratio of the observed oscillation $\Omega, (S/N)$. We are particularly concerned with estimation of Ω.

The likelihood function for N independent observations is usually defined as

$$\ln L(\boldsymbol{p},\boldsymbol{y}) = \sum_{i=1}^{N} \ln f(p_i,\boldsymbol{y}). \tag{26}$$

In the present case the 'observations', i.e. the power spectrum $p(\omega)$, are continuous and possibly their values at different frequencies are correlated. We shall use the following likelihood function:

$$\ln L(\boldsymbol{p},\boldsymbol{y}) = \int_{-\infty}^{\infty} \Gamma \ln f(p(\omega),\boldsymbol{y})\, d\omega, \tag{27}$$

where $\Gamma(\omega)$ is an as yet unspecified weighting function, taking care of the degree of correlation of observations and normalized so that its integral over a certain frequency band corresponds to a number of independent power spectrum observations in the band. Integration in equation (27) is extended over the whole band of power spectrum affected by a given oscillation, i.e. over its whole window pattern.

In case of discrete and evenly spaced observations covering a time interval T, independent observations of power spectrum are too evenly spaced by the frequency interval of $\omega_0 = 1/T$, for all kinds of signals (e.g. Scargle 1982). Thus the number of independent observations of power spectrum per unit frequency interval is

$$\Gamma^{(1)} = \frac{1}{\omega_0} \equiv \text{const.} \tag{28}$$

To estimate this number in a general case of unevenly sampled observations we define a 'calibration signal', consisting of the sum of a large number of unit amplitude sinusoids

with random (i.e. non-coherent) phases and spaced evenly in frequency by ω_0. It is a well-known fact, that the power spectrum of a sinusoid of frequency Ω is the sum of the sampling window functions $W(\omega)$ shifted to $\pm\Omega$. Strictly speaking, this result is valid for the classical power spectrum only. For Scargle's (1982) modified power spectrum one has to use power spectra of high-frequency unit test signals instead of W, however, this affects little our argument – unless a high-frequency pattern is present in the distribution of observation times. Non-coherence of the sinusoids in the calibration function ensures that their positive and negative interferences are equally likely. Thus the expected value of power observed in the calibration signal is simply sum of powers of components:

$$E\{p(\omega)\}=\sum_{n=-\infty}^{\infty} W(\omega+n\omega_0). \tag{29}$$

For sufficiently dense distribution of calibration sinusoids this sum is proportional to an integral of the window function. Equation (29) has a curious property: observed power grows proportionally to the number of aliases and sidelobes present in the window function. However, power of the input calibration signal per unit frequency interval is fixed at the value of $1/\omega_0$. Thus all growth of the observed power is spurious and related to the fact that the power spectrum contains more and more correlated frequencies as number of aliases grows. To account for that and for proper normalization the weighting function for the general case must take the following form

$$\Gamma=\frac{1}{\omega_0}\frac{\int W_E(\omega)\,d\omega}{\int W(\omega)\,d\omega}. \tag{30}$$

Here, by W_E and W we denoted the window functions for equidistant and non-equidistant observation of the same number and spanning the same time interval T. The analytical form of W_E is known so the integral in the numerator can be calculated by elementary methods. Its value is $C\omega_0$, where $C=\mathcal{O}(1)=$ constant. Thus equation (30) can be rewritten as:

$$C=\int\Gamma W\,d\omega. \tag{31}$$

As usual in the MLM, parameters $\boldsymbol{y}$ are found by looking for the maximum of L or as solutions of the following equations:

$$\frac{\partial \ln L(\boldsymbol{p},\boldsymbol{y})}{\partial \boldsymbol{y}}=0. \tag{32}$$

Their covariance matrix is (Eadie *et al.* 1971, section 8.3):

$$Cov\{\boldsymbol{y}\}=E\left\{\left(\frac{\partial \ln L}{\partial \boldsymbol{y}}\right)^2\right\}^{-1}=-\left(\frac{\partial^2 \ln L}{\partial \boldsymbol{y}^2}\right)^{-1}. \tag{33}$$

The conditions under which second equality holds are satisfied here. In order to proceed any further, we must know explicitly the probability distribution f.

The probability distribution of power spectrum for white noise is exponential. It depends on one parameter, variance of the noise (Scargle 1982). However, in the present case we have to cope with a more complex signal composed of a noise and oscillation. Additional parameters concerning of the frequency and shape of the oscillation may have to be considered. In this paper we are interested in sinusoidal oscillations only. So the relevant parameters are the frequency of the oscillation $\Omega\equiv y_1$ and power signal-to-noise ratio $(S^2/N^2)\equiv y_2$. By S^2 and N^2 we indicate power spectra of the deterministic and noise components. The moments of the distribution in such a case are considered in Appendix A. The expected value and variance are $E\{p\}=S^2+N^2$ and $Var\{p\}=\sigma_p^2=2N^2(2S^2+N^2)$, respectively. As long as $S^2\gg N^2$ the standard deviation is small $\sqrt{Var\{p\}}\ll p$ and we may approximate the expected value by the observed one $E\{p\}\approx p$. The exact equality holds in an asymptotic case of strong oscillation $S^2/N^2\to\infty$. Then we may use a linear expansion of the power spectrum as a function (statistics) of observations $\boldsymbol{x}$. As such the power spectrum has also Gaussian probability distribution. The parameters of the distribution $E\{p\}$ and $Var\{p\}$ are already known.

Generally, in the MLM we compare the likelihood of the power spectrum $p(\boldsymbol{y}_0)$ for the true parameter values $\boldsymbol{y}_0$ with the observed one $p(\boldsymbol{y})$. Substituting Gaussians for f in equation (27) and expanding $p(\omega;\boldsymbol{y})$ into linear function of $\boldsymbol{y}$ we obtain the likelihood function in the following form:

$$\ln L(p,\boldsymbol{y})=-\int\Gamma\frac{(\boldsymbol{y}-\boldsymbol{y}_0)^{\dagger}[\partial p(\omega)^{\dagger}/\partial\boldsymbol{y}_0][\partial p(\omega)/\partial\boldsymbol{y}_0](\boldsymbol{y}-\boldsymbol{y}_0)}{8N^2(2S^2+N^2)}\,d\omega + \text{const.} \tag{34}$$

Differentiating and substituting into equation (33) we obtain

$$Cov\{y\}=\left[\int\frac{\Gamma\,d\omega}{4N^2p[1-N^2/2p]}\frac{\partial p(\omega)^{\dagger}}{\partial\boldsymbol{y}_0}\frac{\partial p(\omega)}{\partial\boldsymbol{y}_0}\right]^{-1}. \tag{35}$$

The integrated function is of order $\mathcal{O}(p)$ so it contributes most within the spectral lines. It follows from equation (35) that integration should be extended over the whole interval where line pattern is strong in order to minimize the variance. In what follows we shall consider a strong line ($S/N\gg 1$), whose positive-frequency window pattern does not interfere with the negative one and with other lines. So the profile of the line is $p(\omega)\approx p_\Omega W(\omega-\Omega)$ and depends on two parameters only: $y_1=\Omega$ and $y_2=p_\Omega$ (power in the line centre). Covariance of the two parameters vanishes due to symmetry of the window function, so that $\partial p/\partial p_\Omega$ and $\partial p/\partial\Omega$ are even and odd functions with respect to the line centre. Thus we demonstrated that the MLM power and frequency estimates are uncorrelated.

We shall derive a simple analytical estimate of the variance of oscillation frequency Ω. Substituting $p(\omega)$ into equation (35) we obtain

$$Var\{\Omega\}=\frac{8N^2}{p\Omega}\left[\int\Gamma\,d\omega\left(\frac{dW}{d\omega}\right)^2W^{-1}\sum_{k=0}^{\infty}(N^2/2p)^k\right]^{-1}. \tag{36}$$

For a strong line and near its centre, where most contribution to the integral comes from, $N^2/2p\ll 1/2$, so that the expansion converges fast and it suffices to retain only first term, i.e. unity,

$$Var\{\Omega\}=\frac{8\sigma_\omega^2N^2}{p\Omega}\left[\int\Gamma W\,d\omega\left(\frac{d\ln W}{d\omega}\right)^2\right]^{-1}. \tag{37}$$

Replacing the derivative $d \ln W/d\omega$ in equation (37) by its finite difference approximation $1/\sigma_\omega$ and recalling equation (31) we obtain

$$Var\{\Omega\} = \frac{8\sigma_\omega^2 N^2}{Cp_\Omega}. \tag{38}$$

Relations between signal and noise power density functions p_Ω and N^2 and their amplitudes A and σ_x follow in a simple way from Parseval's Theorem: $p_\Omega \sigma_\omega \sim A^2 \equiv 1$ and $N^2\omega_N \equiv N^2 = \sigma_x^2$. We denoted widths of the signal and noise bands with $\sigma_\omega = 1/T$ and ω_N. The $\equiv$ signs pertain to our particular choice of units.

The noise power density function is affected by presence of any correlation. Assuming a Gaussian filter of width D and recalling that the noise can be represented by a convolution $\boldsymbol{b} * \boldsymbol{w}$ of the white noise $\boldsymbol{w}$ with the filter $\boldsymbol{b}$ we obtain

$$N^2 \equiv |\mathcal{F}b|^2 |\mathcal{F}w|^2 = \frac{\sigma_x^2 D}{\sqrt{\pi}} e^{-D^2\omega^2/2}. \tag{39}$$

For our case $P \gg D$ so the exponent function can be neglected. Note our choice of the amplitude of the oscillation and mean separation of observations for units of signals and time. Substituting into equation (38) we obtain for the single oscillation case

$$Var\{\Omega_P\} \sim \frac{\sigma_x^2 D}{T^3}. \tag{40}$$

This equation has the same form as equation (25) except for a constant factor stemming from different normalizations. So the MLM estimate of Ω has Gaussian probability distribution, in the asymptotic case of a strong signal and its expected value and variance are the same as in the LSQ case. Thus the MLM frequency estimate is equivalent to LSQ fitting of a sinusoid, provided that equation (40) is used to estimate variance. Note that as long as $\Omega \ll 1/D$ no explicit knowledge of correlation length is required for power spectrum period estimation.

One can convert Lomb (1976) and Scargle (1982) results into an explicit relation of χ^2 and p: $\chi^2(\omega) = (p_\Omega - p(\omega) + 2D\sigma_x^2)M/\sigma_x^2$, where M, the total number of observations, and p_Ω, D, σ_x^2 are constants. Since χ^2 was computed for a single oscillation so the relation is valid only for the corresponding spectral line. The correspondence of the LSQ and MLM results would correspond to a particular case of a known theorem that LSQ provides MLM estimate in the linear case (Eadie *et al.* 1971). So for strong lines we can entirely dispose with the cumbersome maximum likelihood algorithm and treat power spectrum as reflected χ^2 plot and find the confidence interval accordingly (Fig. 1).

4 NUMERICAL CONSIDERATIONS

4.1 Iterations of the noise covariance matrix

In principle one can assume a trial covariance matrix $\mathbf{C_x}$ and obtain a LSQ fit. The residuals from the fit can be used to find the noise ACF, and its corresponding covariance matrix $\mathbf{C_x}$. Methods exist to estimate the autocorrelation function even for uneven distribution of observations (e.g. Edelson & Krolik 1988). In a general case the input and output matrices $\mathbf{C_x}$ are inconsistent. Given the new correlation matrix $\mathbf{C_x}$, an improved least-squares solution can be computed using equation (4). By iterations the solution can be found, which is consistent with the correlation matrix $\mathbf{C_x}$. This iteration procedure is cumbersome and its convergence is not guaranteed. The iteration procedure may serve as a check and in cases when other methods fail to work.

4.2 The *post mortem* analysis

We propose the following simple method for LSQ variance estimation. Initially proceed as if no correlation was present in the noise and get the LSQ solution as usual. Fit the unknown parameters $y^{(1)}$ and calculate their correlation matrix $\mathbf{C_y^{(2)}}$ (equation 6). In order to find by what factor this matrix underestimates the true correlation matrix $\mathbf{C_y^{(1)}}$, perform a simple *post mortem* analysis. Namely, find the residuals from the fit and calculate their autocorrelation function (ACF). Determine the correlation length D by fitting a Gaussian function centred at 0 lag and width of D. As long as the conditions (13) are met, the true correlation matrix is given by equation (9), or simply

$$\mathbf{C_y^{(1)}} = D\mathbf{C_y^{(2)}}. \tag{41}$$

We discuss below numerical tests which demonstrate how reliable is this method.

This method has several advantages. With no extra computations it can be applied to published results. Provided, that authors published e.g. plots of raw data *and* the fitted curve, D can be roughly estimated by looking how many consecutive observations lay on the same side of the fitted curve, on average. Thus all quoted variances can be scaled readily by D to make them realistic. The *post mortem* analysis can be generalized for the case of observations in n-dimension space of the independent variables. Then the autocorrelation function becomes an n-dimension matrix. It determines the n-dimension correlation ellipsoid. One

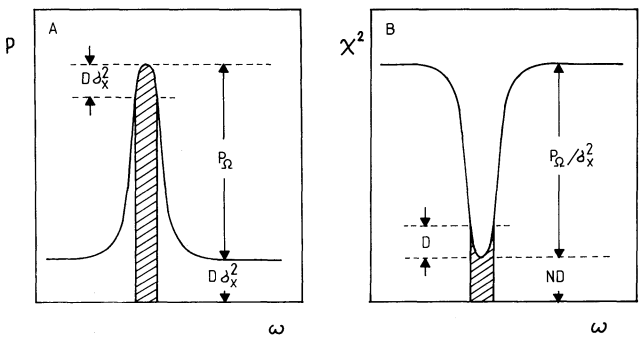

Figure 1. A schematic plot of lines produced by an oscillation detected in power spectrum p (a) and χ^2 periodogram (b). The shaded area corresponds to the maximum likelihood 1σ confidence interval for the frequency of the line and is exactly the same in both plots (see Section 4). The interval corresponds to the width of the line at levels $D\sigma_x^2$ down and D up from its peak for (a) and (b), respectively. Here p_Ω denotes power in the line centre, $N \equiv T$ number of observations, σ_x^2 and D are true noise variance and correlation length and $D\sigma_x^2$ is noise power density, including no window function artefacts.

should multiply variances by the average number of observations within the volume occupied by the ellipsoid.

4.3 Binning of data

An alternative method relies on binning of observations into bins of size D. This method requires performing the least-squares computation twice: first for unbinned data in order to find the solution $\mathbf{y}$ and residuals. The residuals serve to determine the correlation length D. Then obtain second fit for binned data. It yields directly the true covariance matrix $\mathbf{C}_{\mathbf{y}}^{(2)}$ since now $D=1$. Similarly to the previous method, this method can be generalized to n-dimensions.

4.4 Power spectrum linewidth

An equivalent but particularly simple procedure is possible for the strong S/N and simple, non-interfering window patterns. Take power spectrum of an oscillation and find height p_Ω of the corresponding spectral line. Find in the vicinity of the line the mean noise power level $N^2 \equiv D\sigma_x^2$. Then the width of the line at the $p_\Omega - N^2$ level is the 1σ confidence interval of the oscillation period (Fig. 1). Any effect of the correlation of residuals is implicitly included. Finding N^2 may require some care in practice. Many low-power features appearing in power spectra are actually not due to noise but are window patterns of some oscillations and thus should not be taken into account. However, decision may be difficult. This problem may be circumvented if noise variance N^2 is known *a priori*, e.g. from tests of the measuring equipment. Then its power is also known (equation 39). In this case the correlation length D must be known explicitly. This procedure can be performed graphically with no calculation at all.

5 SIMULATION TESTS

In order to verify how far our assumptions (equation 13) may be relaxed, extensive Monte Carlo simulations were performed. The simulated data consisted of a periodic signal plus the correlated noise. For each simulated data set a period was found by LSQ fitting of a sinusoid. Comparison of the variance of the fitted period with computed values of $\mathbf{C}_{\mathbf{y}}$ (equation 9) served as a test of validity of our assumptions and results.

The signal consisted of a sinusoid of amplitude S/N, constant period P and random phase. Two types of the noise were generated: (i) (MA) a moving average noise obtained by convolving a white noise with a Gaussian filter of width D and (ii) $(S+W)$ a sinusoid of amplitude $1/\sqrt{2}$, period D and random phase plus a white noise of variance 1/2. The MA red noise may be considered a model of flickering in accreting binary stars. The $S+W$ noise imitates a not uncommon situation when another unrecognized oscillation is present in data. In both cases noise had unit variance. The white noise was generated using a Gaussian pseudorandom number generator. In order to avoid switch-on effects the first series of random numbers was discarded. For each set of the parameters D, P and S/N, $NC=10\,000$ data sets were simulated. Each consisted of a time series of length $NT=1000$ sampled at unit time intervals.

A sinusoid was fitted to each time series by the non-linear LSQ method. The calculations started from the true value of the period P and Newton iterations were performed in each case in order to find the best-fit period P_{LSQ}. It was found in practice that six iterations were enough. Those results for which the period difference $\Delta P = P_{\mathrm{LSQ}} - P$ produced a shift in phase in excess of 0.1 on ends were rejected. Their count NM was kept separately in order to estimate the cycle miss probability $p_{\mathrm{m}} \approx NM/NC$. Thus we rejected simulations for which no secure cycle count can be established in the same way as a careful observer would do. For the rest of the simulations the external variance of the period was found:

$$Var\{P_0\} \approx \frac{1}{NC-NM} \sum_{n=1}^{NC-NM} (\Delta P)^2. \qquad (42)$$

This 'observed' variance can be compared with the mean of the LSQ variances computed for each fit (equation 9). The calculations were repeated for a grid of values of the para-

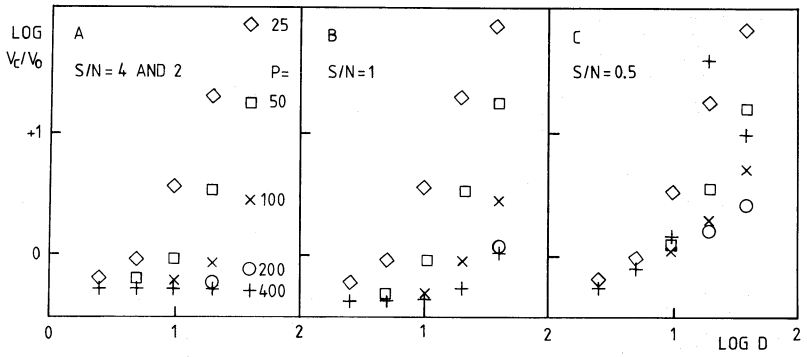

Figure 2. Comparison of two variance estimates of periods obtained by LSQ fitting of sinusoids to simulated oscillations with correlated noise. The calculated variance V_c from LSQ multiplied by the correlation length D accounts for the effect discussed in Section 2. The 'observed' variance was obtained directly from the scatter of fitted period around its value used in simulations. The simulations were obtained for a grid of periods $P=25$–400, each marked with different symbols, noise correlation lengths $D=2.5$–40 and signal-to-noise ratio $(S/N)=4$ and 2 (a), $(S/N)=1$ (b) and $(S/N)=0.5$ (c). To avoid overcrowding, many symbols overlapping with bottom crosses (+) are not drawn. The constant value of V_c/V_0 in (a) and (b) for $D/P<0.2$ demonstrates that as long as our assumptions are satisfied V_c and V_0 are identical except for a constant factor. Note breakdown of their correspondence for small (S/N) (c) for all but the smallest D.

meters $D = 2.5, 5, 10, 20, 40$, $P = 25, 50, 100, 200, 400$; and $S/N = 0.5, 1, 2, 4$. Statistical accuracy of the results was checked by comparison of calculations using different random number series. The corresponding variances differed by no more than 0.01 dex.

Comparison of the observed and computed variances reveals orders of magnitude discrepancies of observed and LSQ variances ignoring correlation length effect. The classical LSQ approach underestimates the variance. Thus the simulations support our first conclusion: the classical LSQ variance estimates are wrong in case of correlated residuals. No clear pattern appears in the $S+W$ noise simulations. A change in the period of the interfering oscillation affects the basic period variance by factor at least of several. The effect is worst for close principal and background periods. At this point we close discussion of the $S+W$ simulations.

The MA results reveal some systematic patterns. In Fig. 2 we plot log ratio of the computed variance to that predicted by equation (9). The latter variance is corrected for noise correlation length D. In Fig. 2 we plot log ratio of the computed and predicted variances against log correlation length D. Points corresponding to the same values of P are marked with the same symbols. Separate graphs are presented for each S/N value. The results for $S/N = 4$ and 2 are the same and thus are not repeated on the plot. They demonstrate that for $D/P < 0.2$ the variances predicted by equation (9) and true variances agree very well, except for a constant scaling factor of $\approx 10^{-0.25} = 0.56$. For $D/P < 0.1$ any differences are no greater than the Monte Carlo errors. No discrepancies occur for correlation length as short as 2.5 time steps. However, for correlation length comparable to the period of oscillation, $D/P > 0.3$ a quite different picture emerges. Instead of being proportional to D/P the true variance is roughly proportional to $(D/P)^{-1}$, so that the plotted ratio of variances grows with $(D/P)^2$. It is possible that this effect saturates for $D/P > 1$. However, no further investigations were carried out since such long D are avoided in practice. S/N ratio does affect our results but only when its value is small, $(S/N) < 1$. For $(S/N) = 1/2$ and $P^3 D > 5 \times 10^6$ the true variance is less than the predicted one. However, the predicted variance at such conditions is large and indicates possibility of wrong cycle count. Indeed, in such cases cycle miss probability was high, reaching 0.4 for $D = 40$ and $P = 400$. So, the variance is artificially decreased by rejection of the outlying period estimates. In such cases the simulated data are simply insufficient to find the period unambiguously. Any random noise fluctuations may be mistakenly taken for oscillation maxima. Still, this situation was correctly predicted by equation (9).

Summarizing, our estimate of variance (equation 9) is reliable for the stochastic noise, as long as our assumptions hold. In fact, our conditions (equation 13) need not be satisfied with large margin. Our variance estimate fails when other periodic signals of comparable magnitude interfere.

6 CONCLUSIONS

Our results may be interpreted in a simple way.

(i) Provided, that our new statistically correct procedures are followed, power spectrum and least-squares (LSQ) fit of a sinusoid are equivalent methods for period determination. For this purpose our corrected variance estimate (equation 9) should be used. This estimate is consistent with scatter from different measurements. Old error estimates, based on variance obtained directly from the fit or on the Rayleigh resolution criterion $\delta\nu = 1/T$ are statistically incorrect.

(ii) If on average D consecutive deviations from the LSQ fit are correlated, they give us as much information on noise as a single observation. So in effect we get D times less uncorrelated observations on which our LSQ analysis is based. Thus our actual variance is, by a possibly large factor of D, larger than that indicated by the LSQ fit routine. In other words we could essentially bin each D observation together without increase of the period variance. For small S/N ratio we propose using least-squares fit as before and then following it by our *post mortem* analysis. The analysis requires calculation of the residuals from the fit and then their autocorrelation function. The half width at half intensity of the autocorrelation function maximum around 0 lag is an estimate of the correlation length D. Then all variances and covariances should be multiplied by D expressed in units of mean observation separation (Section 2).

(iii) For the strong S/N ratio an equivalent but simpler procedure using power spectrum is feasible. Take power spectrum of an oscillation and find height p_Ω of the corresponding spectral line. Find in the vicinity of the line the mean noise power level $N^2 \equiv D\sigma_x^2$. Then the width of the line at the $p_\Omega - D\sigma_x^2$ level is the 1σ confidence interval of the oscillation period (Fig. 1). Any effect of the correlation of noise is implicitly included (Section 3.2).

(iv) A particularly pleasing practical property of the new statistically correct variance estimates in (ii) and (iii) is that they can be easily obtained for most already published observations using simple graphic procedure. Namely, to estimate the correlation length D for LSQ (ii), it suffices to count mean number of consecutive observations lying on the same side of the fitted curve. To estimate the noise level for the power spectrum (iii) it suffices to draw its mean level by eye, ignoring all strong lines and their aliases. Then the same height as that from 0 up to the mean noise level should be plotted from the peak of the line *down*. The width of the line at this level provides 1σ confidence interval.

(v) More general and exact numerical procedures are available (Section 4).

ACKNOWLEDGMENTS

I would like to thank Andrzej Kruszewski for drawing my attention to the role of correlations in period estimation. I would like also to thank Louis Balona, Chris Koen, Brian Warner and an anonymous referee who commented upon presentation of this paper. I am grateful to Professor Raul Viollier for granting access to the computing facilities of UCT Department of Theoretical Physics where simulations were performed. It is also a pleasure to acknowledge the use of the computer facilities of Max-Planck-Institute für Extraterrestische Physik in Garching. Finally I would like to extend my thanks to members of University of Cape Town Department of Astronomy and South African Astronomical Observatory, of all ethnic backgrounds, for making my stay in their country both productive and thoroughly enjoyable.

REFERENCES

Eadie, W. T., Drijard, D., James, F. E., Roos, M. & Sadoulet, B., 1971. *Statistical Methods in Experimental Physics*, North-Holland, Amsterdam.
Edelson, R. A. & Krolik, J. H., 1988. *Astrophys. J.*, **333,** 646.
Lomb, N. R., 1976. *Astrophys. Space Sci.*, **39,** 447.
Scargle, J. D., 1982. *Astrophys. J.*, **263,** 835.
Schwarzenberg-Czerny, A., 1989. *Mon. Not. R. astr. Soc.*, **241,** 153.

APPENDIX A

Let us assume that a signal x is a sum of the deterministic and stochastic (noise) processes s and n respectively, $x=s+n$. We indicate their Fourier transforms by $\kappa=\mathcal{F}x$, $\nu=\mathcal{F}n$ and $\delta=\mathcal{F}s$. We assume that noise is independent from signal and that its variance exists and mean vanishes, so that $E\{\nu\}=E\{\mathcal{F}n\}=\mathcal{F}E\{n\}=0$. However, we do not assume here that the noise is white. The expected values and variances of the power spectra of the involved processes are $E\{\bar{\delta}\delta\}=\bar{\delta}\delta\equiv S^2$, $Var\{\bar{\delta}\delta\}=0$, $E\{\bar{\nu}\nu\}\equiv N^2$ $Var\{\bar{\nu}\nu\}\equiv N^4$. These equations define N^2 and S^2. We exploited here known exponential distribution of $\bar{\nu}\nu$ (e.g. Scargle 1982). The power spectrum of the signal is $p\equiv\bar{\kappa}\kappa=\bar{\delta}\delta+\bar{\nu}\nu+\bar{\nu}\delta+\bar{\delta}\nu$. The expected value and variance of the last two terms are $E\{\bar{\nu}\delta+\bar{\delta}\nu\}=\delta E\{\bar{\nu}\}+\bar{\delta}E\{\nu\}=0$ and $Var\{\bar{\nu}\delta+\bar{\delta}\nu\}=E\{(\bar{\nu}\delta+\bar{\delta}\nu)^2\}=E\{\bar{\nu}^2\delta^2+\nu^2\bar{\delta}^2+2\bar{\nu}\nu\bar{\delta}\delta\}=\delta^2E\{\bar{\nu}^2\}+\bar{\delta}^2E\{\nu^2\}+2\bar{\delta}\delta E\{\bar{\nu}\nu\}=2S^2N^2$. We exploited the fact, that because of the time symmetry 0 and $\pi/2$ phases of the noise transform are equally probable, so $E\{\nu^2\}=0$. Using the above auxiliary results it is straightforward to obtain the expected value of the power spectrum of the combined signal:

$$E\{p\}=N^2+S^2. \tag{43}$$

We start calculation of the variance from its definition: $Var\{p\}=E\{(p-N^2-S^2)^2\}=E\{(\bar{\nu}\nu+\bar{\nu}\delta+\nu\bar{\delta}-N^2)^2\}=E\{(\bar{\nu}\nu-N^2)^2\}+Var\{\bar{\nu}\delta+\nu\bar{\delta}\}+2E\{(\bar{\nu}\nu-N^2)(\bar{\nu}\delta+\nu\bar{\delta})\}$. The first term can be evaluated explicitly since the probability distribution of the white noise is exponential (*cf.* Scargle 1982), yielding N^4. We already evaluated the second term. The last term vanishes since it amounts to an expected value of a product of two factors, which signs are independent. Indeed, the sign of the first factor depends solely on the modulus of the noise transforms and the second one on its phase. Collecting all terms together we obtain finally

$$Var\{p\}=N^2(N^2+2S^2). \tag{44}$$