

Advanced Time Series Analysis of Generally Irregularly Spaced Signals: Beyond the Oversimplified Methods

IVAN L. ANDRONOV, DSC, PROF

11.1 INTRODUCTION

Time series analysis is the tool to study the dynamics and evolution of processes and objects of any nature. Its main aim is to extract information with a minimum number of parameters needed for approximating the data within a given interval, or to forecast it outside, with a required (or best) accuracy. In other words, the famous “Occam’s Razor” principle is “*Entia non sunt multiplicanda praeter necessitatem*,” or “Entities are not to be multiplied beyond necessity” (Schaffer, 2015).

Many of the well-known methods were inspired by the data being used in astro- and geo-informatics. “Ideally,” as in mathematical analysis, one has to analyze an infinite continuous function with continuous derivatives of any order. “Back to reality,” and the “ideal” is a finite number n of data points $(t_k, x_k, k = 1..n)$ with the additional “bonus” of a regular spacing of the arguments: $t_k = t_0 + \delta \cdot k$, where δ is called “time resolution,” “time step,” etc. The regularity is a necessary condition for the Fourier transform (FT), fast Fourier transform (FFT), and analysis using the autocorrelation function (ACF) and cross-correlation function (CCF) (Fisher, 1954; Scheffe, 1959; Anderson, 1958, 2003; Press et al., 2007). A review of the history of the methods was presented by Wermuth (2011). Combinations of methods are used in neural networks and machine learning (Haykin, 1999).

The data are often irregularly spaced, having gaps and irregular argument distributions either in time, or (if periodic) in the phase domain. Anyway, often methods have been applied, which contain a sequence of few simple steps and simplified formulas. Even for test data with well-defined properties, this may lead to significant differences between the parameters of the test model and the values obtained using simplified methods. This type of partial modeling is often called “de-trending” (or “trend removal”) and “pre-whitening.” Mikulášek

(2007a) called such type of modeling “Matrix-Phoebia,” as, mathematically, this is similar to ignoring the nondiagonal values of the matrix of normal equations in the least-squares (LS) method.

Below we present references to the correct (“non-simplified”) methods and show in which cases they are significantly better than the common “simplified” ones. Obviously, “simplified” and “non-simplified” methods should produce the same corresponding parameters and approximations, if the conditions used for the “simplified” methods are satisfied. If not, an improved method should be used instead of one of the simplified ones.

The variety of methods reflects the variety of types of variability. If we could have an infinite number of observations, theoretically one could get an infinite number of values of the functions describing the FT. Practically, in astro- and geo-time series, the data are not only finite in the number of observations, but also often have gaps between single observations (e.g., photographs from Harvard, Sonneberg, Odessa, Moscow, Asiago plate collections and others, CCD photometric surveys [NSVS, Wozniak et al., 2004; ASAS, Pojmanski, 2002; OGLE, Paczynski, 1986; Udalski et al., 1997; CRTS, Drake et al., 2009; MASTER, Lipunov et al., 2010; WASP, Butters et al., 2010; Street et al., 2003, ZTF; etc.], or space observations [Hipparcos/Tycho, Høg et al., 2000; KEPLER, 2019; TESS, 2019; WISE, Wright et al., 2010; GAIA, 2019]).

Many of these signals contain different contributions, for the extraction of which one has to use complicated models, rather than a sequence of simple ones. Such an approach significantly improves the accuracy, which allows us not only to determine the model parameters, but also to more surely estimate their statistical significance and to avoid wrong detection/discovery and thus fake interpretations.

The elaboration of advanced algorithms and programs was inspired by long-term collaboration with

the “Inter-Longitude Astronomy” (ILA) team. This is a joint name of a series of smaller temporarily projects on observation and interpretation of concrete variable stars or groups of stars. It has no special funds; professional and amateur astronomers take part based on their own resources. This “ILA” project is in some way similar and complementary to other projects like WET, CBA, VSOLJ, BRNO, MEDUZA, and currently UkrVO (Vavilova et al., 2012) and Astrominformatics (Vavilova et al., 2017), where many of us take part. Previous reviews on the ILA project were published by Andronov et al. (2003a, 2014, 2017a).

11.2 STATISTICAL PROPERTIES OF THE FUNCTIONS OF (CORRELATED) PARAMETERS OF THE LS FITS

11.2.1 Least-Squares Method: Test Functions

The most common method for the determination of the statistically optimal approximation with a corresponding set of parameters is called the least-squares (LS) method and was proposed about two centuries ago by Carl Friedrich Gauss (1777–1855). It has been described in thousands of textbooks and monographs (e.g., Anderson, 1958, 2003; Forsythe et al., 1977; Press et al., 2007). LS problems with restrictions were discussed by Lawson and Hanson (1974). For a more complete description, we present an extended set of formulas, which are typically omitted.

Typically, it may be written in a form, with minimizing the weighted sum of the residuals of the observational points x_k from the calculated values x_{Ck} at the arguments t_k , i.e.,

$$\Phi = \sum_{k=1}^n (x_k - x_{Ck})^2. \quad (11.1)$$

However, in many cases, one should take into account the weights of the observations $w_k = \sigma_0^2 / \sigma_k^2$, where σ_k is the standard error of the observation x_k and σ_0 is called “the unit weight error.” Then we have

$$\Phi = \sum_{k=1}^n w_k \cdot (x_k - x_{Ck})^2. \quad (11.2)$$

For the pure Gaussian noise, the random value $U = \Phi / \sigma_0^2$ is distributed according to the χ^2 distribution with $(n - m)$ degrees of freedom, where m is the number of the independent parameters C_α ($\alpha = 1..m$) used for computation of the approximation $x_C(t)$.

More complicated, but statistically justified, is the use of the weight matrix w_{jk} , thus leading to

$$\Phi = \sum_{kj=1}^n w_{kj} \cdot (x_k - x_{Ck}) \cdot (x_j - x_{Cj}). \quad (11.3)$$

For the “correct” evaluation of the matrix w_{ij} , one should use

$$w_{ij} = \sigma_0^2 \mu_{ij}^{-1}, \quad (11.4)$$

where μ_{ij} is the covariation matrix of the errors of the observations, μ_{ij}^{-1} is an inverse matrix, and σ_0 is any positive constant. Generally, we do not mention the precise values of μ_{ij} , but, by setting the coefficients w_{ij} , we automatically assume the matrix $\mu_{ij} = \sigma_0^2 \cdot w_{ij}^{-1}$.

This expression resembles the metrics in the tensor analysis and shows a squared “distance” between the observations and the approximation.

Next improvement contains the filter (weight) function $p(z_k, z_j)$, $z_k = (t_k - t_0) / \Delta t$, which is dependent on the times of observations t_k, t_j and on the “shift” t_0 and “scale” Δt , as typically defined in the wavelet analysis. The general expressions for this case are presented by Andronov (1997). For practical purposes, the replacement of w_k in Eq. (11.2) by $p(z_k) \cdot w_k$ may be recommended, so neglecting possible correlations between the statistical deviations of the data from the “true values.”

All these four approaches may be written using a generalized form of the scalar product of vectors, i.e.,

$$(\vec{a} \cdot \vec{b}) = \sum_{kj=1}^n p(z_k, z_j) \cdot w_{kj} \cdot a_k \cdot b_j. \quad (11.5)$$

The vectors \vec{a} and \vec{b} are called “orthogonal” if $(\vec{a} \cdot \vec{b}) = 0$. The “squared” vector may be defined in a usual way:

$$\vec{a}^2 = (\vec{a} \cdot \vec{a}) = \sum_{kj=1}^n p(z_k, z_j) \cdot w_{kj} \cdot a_k \cdot a_j. \quad (11.6)$$

Then the test function may be generally written as

$$\begin{aligned} \Phi &= (\vec{x} - \vec{x}_C)^2 \\ &= \sum_{kj=1}^n p(z_k, z_j) \cdot w_{kj} \cdot (x_k - x_{Ck}) \cdot (x_j - x_{Cj}) \end{aligned} \quad (11.7)$$

and one should determine the values of the coefficients (often called “parameters”) C_α , $\alpha = 1..m$, of the approximation $x_C(t, C_\alpha)$.

In some cases, the coefficients are determined using the principle of the maximum likelihood. However, under a common assumption of normal probability distribution of observational errors, the parameters of the maximum likelihood are exactly at the minimum of the function Φ .

11.2.2 Linear Least Squares

The simplest case of the approximation $x_C(t, C_\alpha)$ is a linear combination of so-called basic functions $f_\alpha(t)$, i.e.,

$$x_C(t, C_\alpha) = \sum_{\alpha=1}^m C_\alpha f_\alpha(t). \quad (11.8)$$

In this case, the minimization of the test function is reduced to the solution of the following system of normal equations:

$$\sum_{\alpha=1}^m A_{\alpha\beta} C_\alpha = B_\beta, \quad (11.9)$$

$$C_\alpha = \sum_{\alpha=1}^m A_{\alpha\beta}^{-1} B_\beta, \quad (11.10)$$

where $A_{\alpha\beta} = (\vec{f}_\alpha \cdot \vec{f}_\beta)$, $B_\beta = (\vec{x} \cdot \vec{f}_\beta)$, and $A_{\alpha\beta}^{-1}$ is the matrix, inverse to $A_{\alpha\beta}$. The matrix $A_{\alpha\beta} = A_{\beta\alpha}$ is symmetrical.

If the set of the basic functions is orthogonal ($A_{\alpha\beta} = (\vec{f}_\alpha \cdot \vec{f}_\beta) = A_{\alpha\alpha} \delta_{\alpha\beta}$), the inverse matrix is diagonal ($A_{\alpha\beta}^{-1} = (1/A_{\alpha\alpha}) \cdot \delta_{\alpha\beta}$), and $C_\alpha = B_\alpha / A_{\alpha\alpha}$.

The following relations are valid: $((\vec{x} - \vec{x}_C) \cdot \vec{f}_\alpha) = 0$, $((\vec{x} - \vec{x}_C) \cdot \vec{x}_C) = 0$,

$$((\vec{x} - \vec{x}_C) \cdot (\vec{x} - \vec{x}_C)) = (\vec{x} \cdot \vec{x}) - (\vec{x}_C \cdot \vec{x}_C). \quad (11.11)$$

Introducing the “reference” vector

$$\vec{x}_0 = \sum_{\alpha=1}^m C_{0\alpha} \vec{f}_\alpha, \quad (11.12)$$

$$(\vec{x} - \vec{x}_C)^2 = (\vec{x} - \vec{x}_0)^2 - (\vec{x}_C - \vec{x}_0)^2. \quad (11.13)$$

This extended relation is valid for any set of constant coefficients $C_{0\alpha}$.

Typically, if it is used, an abbreviated form $\vec{x}_0 = \vec{x}_1 = (C_{11}, C_{11}, \dots, C_{11})$ is applied, where C_{11} is a solution of the one-parameter fit $x_c(t) = C_{11}$, (so $f_1(t) = 1$), i.e.,

a weighted sample mean of x_k . In a general case of nonorthogonal basic functions, the coefficients C_α are dependent on m , so it might be recommended to write a complete form $C_{m\alpha}$ (as we did for C_{11}). So, generally, $C_{m\alpha} \neq C_{L\alpha}$ for different number of parameters. However, after this remark, we will still use C_α as a short designation of $C_{m\alpha}$ for current m .

11.2.3 Influence of Deviations of Coefficients

In the “linear combination” case (Eq. (11.8)), the test function for any coefficients $\tilde{C}_\alpha = C_\alpha + D_\alpha$ may be rewritten as

$$\begin{aligned} \Phi(\tilde{C}_\alpha) &= \Phi_m + \sum_{\alpha,\beta=1}^m A_{\alpha\beta} \cdot (\tilde{C}_\alpha - C_\alpha) \cdot (\tilde{C}_\beta - C_\beta), \\ \Phi_m &= \Phi(C_\alpha) = \Phi_0 - \sum_{\alpha,\beta=1}^m A_{\alpha\beta} \cdot C_\alpha \cdot C_\beta, \quad (11.14) \\ \Phi_0 &= (\vec{x} \cdot \vec{x}). \end{aligned}$$

Here D_α are some coefficients. Let

$$x_D(t) = \sum_{\alpha=1}^m D_\alpha \cdot f_\alpha(t). \quad (11.15)$$

Similarly to common designations $\vec{x} = \vec{O}$ (observed), $\vec{x}_C = \vec{C}$ (calculated), let $\vec{x}_D = \vec{D}$ (development). Then Eq. (11.14) may be rewritten for the squares (scalar products) of vectors, and we have

$$\begin{aligned} &(\vec{O} - \vec{C} - \vec{D})^2 \\ &= (\vec{O} - \vec{C})^2 + \vec{D}^2 = \vec{O}^2 - \vec{C}^2 + \vec{D}^2 = \Phi(C_\alpha + D_\alpha) \\ &= \Phi_0 - \sum_{\alpha,\beta=1}^m A_{\alpha\beta} \cdot C_\alpha \cdot C_\beta + \sum_{\alpha,\beta=1}^m A_{\alpha\beta} \cdot D_\alpha \cdot D_\beta. \end{aligned} \quad (11.16)$$

A two-dimensional simple geometrical interpretation of these equations is shown in Fig. 11.1. As the vectors \vec{C} and $\vec{O} - \vec{C}$ are orthogonal, the squares of their length are related by the Pythagoras theorem.

For the model (11.8), the test function $\Phi(\tilde{C}_\alpha)$ is an m -dimensional paraboloid with a single minimum at $\tilde{C}_\alpha = C_\alpha$. This relatively rare equation may be useful either to see the influence of the rounding of the determined coefficients on the test function, or to determine “confidence intervals” for the parameters (cf. Cherepashchuk, 1993).

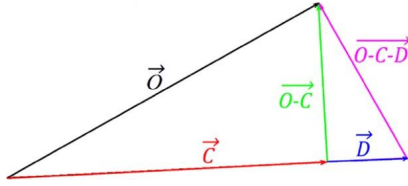


FIG. 11.1 Two-dimensional illustration of the m -dimensional equation (11.16).

11.2.4 Linear Approximation

The mostly used formula is the linear approximation

$$x_C(t) = C_1 + C_2 \cdot t = \tag{11.17}$$

$$= \bar{x} + C_2 \cdot (t - \bar{t}), \tag{11.18}$$

$$\begin{aligned} \sigma^2[x_C(t)] &= \sigma_{\bar{x}}^2 \left(1 + \frac{(t - \bar{t})^2}{\sigma_t^2} \right) \\ &= \sigma_{\bar{x}}^2 \left(1 + \frac{t^2 - 2t\bar{t} + \bar{t}^2}{\sigma_t^2} \right), \end{aligned} \tag{11.19}$$

$$\begin{aligned} \sigma_{\text{wrong}}^2[x_C(t)] &= \sigma^2[C_1] + \sigma^2[C_2] \cdot t^2 \\ &= \sigma_{\bar{x}}^2 \left(1 + \frac{\bar{t}^2 + t^2}{\sigma_t^2} \right). \end{aligned} \tag{11.20}$$

Here the mean values $\bar{x} = B_1/A_{11}$, $\bar{t} = A_{12}/A_{11}$, the variance $\sigma_{\bar{x}}^2 = \sigma_0^2/n$, and $\sigma_{\text{wrong}} = \sigma_w$ is “wrong” (or “oversimplified”) estimate corresponding to Eq. (11.20). The mean values are here defined in a general case, not restricting to the case of equal weights (i.e., assuming that $\sigma_k = \text{const}$). These expressions are valid for any of the forms of the matrix w_{ij} with a definition of the sam-

ple mean value of any function as

$$\begin{aligned} \sigma_{\text{wrong}}^2[x_C(t)] &= \sum_{\alpha}^m R_{\alpha\alpha} \cdot (f_{\alpha}(t))^2 \\ &= \sum_{\alpha}^m (f_{\alpha}(t) \cdot \sigma[C_{\alpha}])^2. \end{aligned} \tag{11.21}$$

The difference in the error estimates using the correct formula and the incomplete one is shown in Fig. 11.2.

For the uniform distribution of times, starting from time zero, and large n , the accuracy of the zero point is $\sigma[C_1] = 2\sigma_{\bar{x}}$, and this is exactly twice larger than the accuracy of the zero point $\sigma_{\bar{x}}$ in Eq. (11.20).

In Fig. 11.2, the illustrative dependence of $x = \Delta(R - I)$ on $t = B - V$ is shown based on Table 2 from Kim et al. (2004). The best fit line is $x_C(t)$. The difference between the error estimates is due to nonorthogonality of the set of basic functions $f_{\alpha}(t_k)$, which leads to an incorrect formula. Thus it should be recommended to use a slightly complicated expression $x_C(t) = \bar{x} + C_2 \cdot (t - \bar{t})$ with an error estimate.

Often the data are regular and the first moment is set to zero, i.e., the time is converted to $t_k = t_{0k} - t_{01}$. Here t_{0k} are times with an arbitrary zero point in this case $\bar{t} = (n - 1)/2\delta$, $\sigma_t = ((n^2 - 1)/12)^{1/2}\delta$. Thus for any n , $\sigma[x_C(0)] = \sigma[C_1] = 2\sigma_{\bar{x}}(1 - 1.5/(n + 1))^{1/2}$. For large n , $\sigma[C_1] = 2\sigma_{\bar{x}}$. Using a simplified formula (11.21), one gets $\sigma[x_C(\bar{t})] = \sqrt{7}\sigma_{\bar{x}}$ instead of the correct value of $\sigma_{\bar{x}}$.

Besides the linear regression (Eq. (11.17)) with coefficients C_2 determined using the LS, more often are used other lines with different slopes. They all have

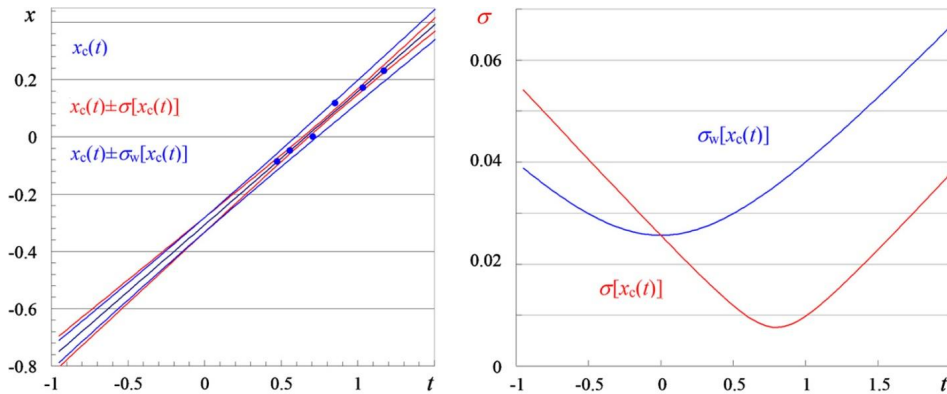


FIG. 11.2 The dependence of the smoothing linear function $x_C(t)$ and its “ $\pm 1\sigma$ ” error corridors for the correct equation (11.19) and wrong (simplified) equation (11.20). For the illustration, we have used the dependence of $\Delta(V - R)$ on the color index $(B - V)$ (Table 2 from Kim et al. (2004)).

a cross-point at (\bar{t}, \bar{x}) , so $x_\epsilon = \bar{x} + b_\epsilon \cdot (t - \bar{t})$. Introducing the second-order moments $\mu_{xx} = \bar{x}^2/A_{11} - \bar{x}^2$, $\mu_{tx} = B_2/A_{11} - \bar{t}\bar{x}$, $\mu_{tt} = A_{22}/A_{11} - \bar{t}^2$, the coefficients are $b_1 = C_2$ (the best fit approximation for $x_C(t)$ as a function of t , i.e., minimizing $(\bar{x} - \bar{x}_C)^2$); $b_2 = \mu_{tx}/\mu_{xx}$ for minimizing $(\bar{t} - \bar{t}_C)^2$; $b_3 = \mu_{tx}/\mu_{xx}$ for minimizing $(\bar{t} - \bar{t}_C)^2$; $b_3 = -q + (q^2 + 1)^{1/2}$, $b_4 = -q - (q^2 + 1)^{1/2}$, $q = (\mu_{tt} - \mu_{xx})/(2\mu_{tx})$ are slopes of the “orthogonal” regression (the coefficient of the main line has the same sign as μ_{tx}). Two-dimensional orthogonal regression is a kind of principal component analysis (PCA), which is discussed below. One may also apply PCA to normalized data, with a corresponding slope $b_5 = \text{sign}(\mu_{tx}) \cdot (\mu_{xx}/\mu_{tt})^{1/2}$, which is dependent on the sign of μ_{tx} , but not on its value.

The correlation coefficient $r = \mu_{tx}/(\mu_{xx}\mu_{tt})^{1/2}$. Its accuracy $\sigma[r] = (1 - r^2)/\sqrt{n-1}$, but, for the significance test, the value of $t = r/\sigma[r]$ is used, where $\sigma[r] = \sqrt{1 - r^2}/\sqrt{n-1}$. Another test of the “null hypothesis” is for $t = \sqrt{n-3} \cdot \ln \sqrt{(1+r)/(1-r)}$ (Fisher, 1954). More details on application of correlation analysis are discussed by Isobe et al. (1990). The case of unequal independent statistical errors in *both* coordinates was reviewed by Press et al. (2007).

11.2.5 Linearization

There are also some approximations, the parameters of which may be more easily estimated by replacing initial functions with other ones. For example, for a power dependence model $x = a \cdot t^b$, $\log x = \log a + b \cdot \log t$, or $\bar{x} = \bar{a} + b \cdot \bar{t}$. Obviously, it is acceptable only for positive values of all x, t . Similarly, for an often exponential model $x = a \cdot \exp(bt)$, $\log x = \log a + b \cdot t$, or, again, $\bar{x} = \bar{a} + b \cdot \bar{t}$. In this case, the expressions become linear, and the parameters are determined without using slightly more complicated nonlinear optimization. However, the coefficients obtained using “linearized” and initial equations are generally different. They coincide only if the residuals are zero, containing no observational errors. For small $\sigma_k \ll x_k$, $\sigma[\log_z(x_k)] \approx \sigma_k/x_k/\ln z$. So it should be recommended to use the parameters from the “linearized” model only as initial values for further iterations in a nonlinear model.

11.2.6 Statistical Properties of Functions of Coefficients

In the general case of nonorthogonal basic functions, the statistical errors are highly correlated. Thus it is important to define a complete covariance matrix for the

errors of the coefficients $R_{\alpha\beta} = \langle C_{d\alpha} C_{d\beta} \rangle$. Generally,

$$R_{\alpha\beta} = \sum_{\gamma \in 1}^m \sum_{i,j,k,L=1}^n A_{\alpha\gamma}^{-1} A_{\epsilon\beta}^{-1} p(z_i, z_j) \cdot w_{ij} p(z_k, z_L) \cdot w_{kL} f_\gamma(t_i) f_\epsilon(t_k) \mu_{jL}. \quad (11.22)$$

This is a statistically correct full version.

Under two assumptions ($p(z_i, z_j) = 1$ and Eq. (11.4)), this long equation significantly shortens to

$$R_{\alpha\beta} = \sigma_0^2 \cdot A_{\alpha\beta}^{-1}. \quad (11.23)$$

The statistical errors (accuracy) of the coefficients are

$$\sigma[C_\alpha] = \sqrt{R_{\alpha\alpha}}. \quad (11.24)$$

The variance (the squared error estimate of $\sigma[G]$) of the general function $G(C_\alpha)$ of coefficients is

$$\sigma^2[G] = \sum_{\alpha\beta=1}^m R_{\alpha\beta} \cdot \frac{\partial G}{\partial C_\alpha} \cdot \frac{\partial G}{\partial C_\beta}. \quad (11.25)$$

One may note a common “simplified” (generally wrong) relation

$$\sigma^2[G] = \sum_{\alpha}^m R_{\alpha\alpha} \cdot \left(\frac{\partial G}{\partial C_\alpha} \right)^2 \quad (11.26)$$

$$= \sum_{\alpha}^m \left(\frac{\partial G}{\partial C_\alpha} \cdot \sigma[C_\alpha] \right)^2. \quad (11.27)$$

This is what Mikulášek (2007a) called the “Matrix-Phoebia.” Both relations coincide only if the matrix $R_{\alpha\beta}$ (and so $A_{\alpha\beta}$) are diagonal. For a simplest comparison, this is like using $a^2 + b^2$ for a result of $(a + b)^2 = a^2 + 2ab + b^2$.

Particularly, for the smoothing function (Eq. (11.8)) itself,

$$\sigma^2[x_C(t)] = \sum_{\alpha\beta=1}^m R_{\alpha\beta} \cdot f_\alpha(t) \cdot f_\beta(t), \quad (11.28)$$

with a “simplified” expression

$$\sigma^2[x_C(t)] = \sum_{\alpha}^m R_{\alpha\alpha} \cdot (f_\alpha(t))^2 = \sum_{\alpha}^m (f_\alpha(t) \cdot \sigma[C_\alpha])^2. \quad (11.29)$$

This equation is valid *only* if the matrix $R_{\alpha\beta}$ is diagonal, so the matrix of normal equations $A_{\alpha\beta}$ is diagonal.

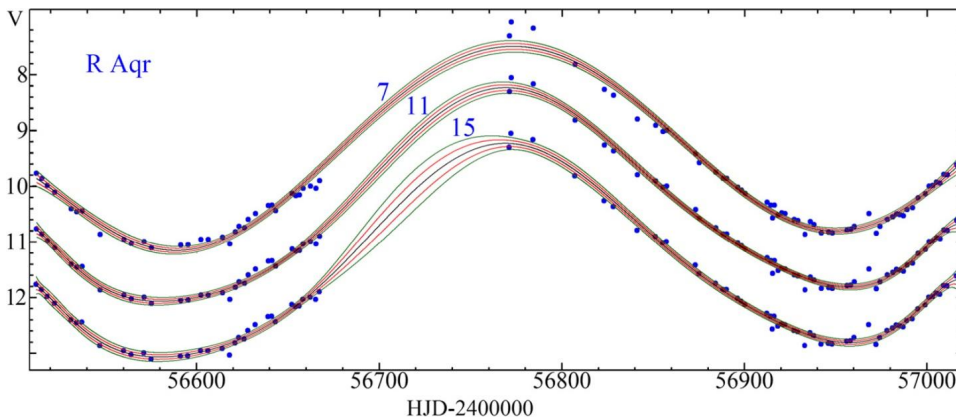


FIG. 11.3 Part of the light curve of R Aqr, the symbiotic binary star with a pulsating component. The observations were taken from the AAVSO international database in the filter V. The approximations $x_c(t)$ and the error corridors $x_c(t) \pm \sigma[x_c(t)]$ and $x_c(t) \pm 2\sigma[x_c(t)]$ for different numbers of parameters m correspond to different criteria for the determination of m .

In other words, the system of basic vectors (vectors of values of the basic functions) is orthogonal.

The “Matrix-Phoebia” (Mikulášek, 2007a) in using this equation (11.29) is in replacing the matrix $R_{\alpha\beta}$ by a diagonal matrix, arbitrarily setting nondiagonal values to zero.

In the popular electronic tables (Microsoft Excel, Open/LibreOffice Calc, GNUmeric), there is a possibility of approximation the data using polynomial and some linearized approximations. However, no error estimates of the parameters and smoothing function are available, as well as the use of the weights. With the function *LINEST*, it is possible to determine parameters and their estimates for the simplest case $w_{ij} = \text{const} \cdot \delta_{ij}$. Even the diagonal form $w_{ij} = \text{const} \cdot \delta_{ij}$ and the covariation matrix $R_{\alpha\beta}$ are not available.

An illustration of the approximation of the real data is shown in Fig. 11.3. The methods to determine an optimal number of parameters is discussed below.

11.2.7 Accuracy of the Derivative and Moments of Crossings

Similarly, for the derivative of degree q of the smoothing function $x_C^{(q)}$, $f_\alpha^{(q)}(t) = d^q f_\alpha(t)/dt^q$,

$$\sigma^2[x_C^{(q)}(t)] = \sum_{\alpha\beta=1}^m R_{\alpha\beta} \cdot f_\alpha^{(q)}(t) \cdot f_\beta^{(q)}(t). \quad (11.30)$$

For “running approximations” with $p(z_k, z_j) \neq \text{const}$, when an approximation consists of mid-points of

the numerous smoothing functions, the expressions are more complicated, as presented by Andronov (1997).

The moment of crossing t_{cross} of the approximation of the constant level x_{cross} (e.g., the gamma velocity or some constant brightness) is determined as the root of equation $x_C(t_{cross}) = x_{cross}$ and has a standard error of

$$\sigma[t_{cross}] = \frac{\sigma[x_C(t_{cross})]}{|\dot{x}_C(t_{cross})|}, \quad (11.31)$$

where $\dot{x}_C(t_{cross})$ is the derivative dx/dt at the moment t_{cross} .

Alternately, Andronov et al. (2008) and Andronov and Andrych (2014) used the inverse approximation $t_c(x)$, where results may be expressed as $t_{cross} = t_c(x_{cross})$, so $\sigma[t_{cross}] = \sigma[t_C(x_{cross})]$.

Such method is preferred when the duration of the ascending or descending branches is much smaller than the duration of the “outburst” (or “eclipse”). Some algorithms that still use $x_C(t)$ approximations are implemented in the software MAVKA (Andrych and Andronov, 2019).

11.3 STATISTICALLY OPTIMAL NUMBER OF PARAMETERS

11.3.1 “Esthetic” (User-Defined)

There are few methods to determine a number of parameters. Probably the most common one may be called an “esthetic,” as the user chooses himself, looking at the approximation and visually estimating its quality. It is realized in the electronic tables like Microsoft Excel,

GNUmeric, Libre Office, Open Office, Kingsoft Office, etc. The default approximation is a polynomial one, and the user chooses the degree of the polynomial. An example is shown in Fig. 11.3. A similar approach is realized in the number of principal components used for filtering multichannel signals (Golyandina et al., 2001).

11.3.2 Analysis of Variance (ANOVA)

The comparison between the values of Φ_m may be done in a few ways. In many cases, there is a comparison between the smoothing function obtained with $m - q$ and m parameters. Very often, the parameter q is set to 1 (e.g., for algebraic polynomials) or 2 (for trigonometrical polynomials). The difference between them may be scaled easily as $B = (\Phi_{m-q} - \Phi_m) / \Phi_{m-q}$. Assuming that the residual signal $x_k - x_{cm}(t_k)$ is a random Gaussian noise with a theoretical covariation matrix μ_{jk} , the random values of Φ_m / σ_0^2 are expected to obey the χ_{n-m}^2 random distribution. As the value of σ_0 is unknown, one uses the sample value of σ_{0m} . The ratio $\sigma_{0m}^2 / \sigma_0^2$ obeys a “reduced” χ_{n-m}^2 random distribution.

Similarly, the value

$$B = \frac{\Phi_{m-q} - \Phi_m}{\Phi_{m-q}} \quad (11.32)$$

has the B (Beta) distribution with the parameters $q/2$ and $(n - m)/2$. The FAP may be computed in many programs, also in the electronic tables, as the function *BETADIST*.

Traditionally, another related value,

$$F = \frac{n - m}{q} \cdot \frac{\Phi_{m-q} - \Phi_m}{\Phi_m} = \frac{n - m}{q} \cdot \frac{B}{1 - B}, \quad (11.33)$$

is used, instead of the value B . If the observational errors obey the normal (Gaussian) probability distribution, the value of F is a random variable corresponding to the Fisher F distribution with q and $n - m$ degrees of freedom (Fisher, 1954). It also may be computed using the electronic tables. Obviously, the estimate of FAP will be the same, if using either B or F . The inverse relation is $B = qF / (n - m + qF)$.

In the particular case $q = 1$, one may determine the same value of FAP using the Student T distribution for the value $T = C_m / \sigma[C_m]$ for the last coefficient of the approximation ($\alpha = m$).

As the system of basic functions is generally not an orthogonal one, the T distribution is valid only for the last coefficient C_m . For fast estimates, people use the “ 3σ ” criterion, i.e., the coefficient is decided to be statistically significant if $|T| = |C_m / \sigma[C_m]| \geq 3$. The FAP may

be computed in the electronic tables using the function $2 \cdot \text{TDIST}(|T|, n - m)$.

In a frequent case $f_1(t) = 1$, the quality of the fit is measured as the square of the correlation coefficient r between the observed and calculated data,

$$S = r^2 = 1 - \frac{\Phi_m}{\Phi_1}. \quad (11.34)$$

Here S is the same as B (Eq. (11.32)) with special parameter $q = m - 1$.

11.3.3 “Best Accuracy” and Related Estimates

Another method was proposed by Andronov (1994a) (see also Andronov, 2003; Andronov and Marsakova, 2006), in which the number of parameters is determined to get the best accuracy estimate of the chosen phenomenological parameter, e.g., the r.m.s. accuracy of the smoothing function $\sigma_m[x_c]$ at the times of observations,

$$\begin{aligned} \sigma_m^2[x_c] &= \frac{1}{N} \sum_{k=1}^n \sigma^2[x_c(t_k)] \\ &= \frac{\sigma_{0m}^2}{N} \sum_{k=1}^n \sum_{\alpha\beta=1}^m A_{\alpha\beta}^{-1} f_\alpha(t_k) f_\beta(t_k). \end{aligned} \quad (11.35)$$

Only in the case $w_{jk} = \text{const} \cdot \delta_{jk}$, this expression is simplified to

$$\sigma_m^2[x_c] = \frac{m}{N} \sigma_{0m}^2 = \frac{m}{N} \cdot \frac{\Phi_m}{n - m}. \quad (11.36)$$

For the more general case $w_{jk} = w_k \cdot \delta_{jk}$, one may introduce a weighted version

$$\sigma_m^2[x_c] = \frac{1}{W} \sum_{k=1}^n w_k \sigma^2[x_c(t_k)] = \frac{m}{W} \sigma_{0m}^2 = \frac{m}{W} \cdot \frac{\Phi_m}{n - m}, \quad (11.37)$$

where $W = \sum_{k=1}^n w_k$.

Similarly, one may introduce “continuous” versions of the root mean squared accuracy,

$$\sigma_{mxc}^2 = \frac{1}{W_I} \int_{t_{\min}}^{t_{\max}} w_I(t) \sigma^2[x_c(t)] dt, \quad (11.38)$$

$$W_I = \int_{t_{\min}}^{t_{\max}} w_I(t) dt, \quad (11.39)$$

where $w(t)$ is a (user-defined) weight function, which is nonnegative in the interval $[t_{\min}, t_{\max}]$.

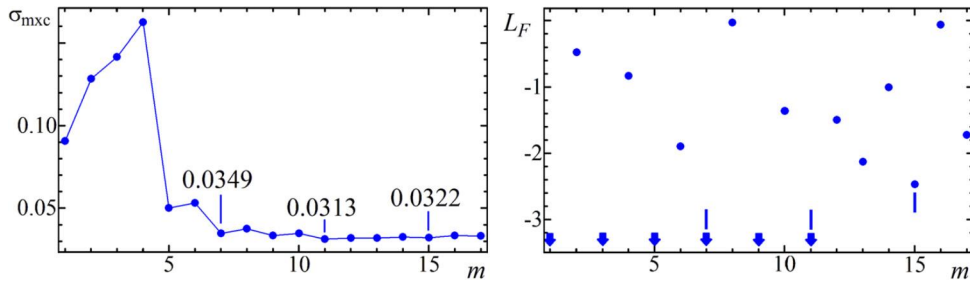


FIG. 11.4 *Left:* Dependence of r.m.s. accuracy of the smoothing function σ_{mxc} at the arguments of observations on the number of parameters m , for the part of the light curve of R Aqr shown in Fig. 11.3. The numbers correspond to the values of σ_{mxc} . *Right:* Dependence of $L_F = \lg \text{FAP}$ on m . The vertical bars correspond to the same three values of m .

Particularly, the function $W_I(t)$ may be a constant in this interval, so making the best r.m.s. accuracy estimate for the interval. Otherwise, it may be a Dirac's $W_I(t) = \delta(t - \tilde{t})$, producing the value for a chosen time \tilde{t} .

Generally, σ_{mxc} initially decreases with increasing m , as Φ_m increases, then increases, as σ_{0m} reaches a “standstill,” and the multiplier m becomes more important. The value of m which corresponds to the minimum of σ_{mxc} , may be recommended to be a statistically optimal one. It may be significantly smaller than the value determined from the FAP.

An additional algorithm to determine the statistically optimal value of m is based on the signal-to-noise ratio (SNR), which is usually defined as $\text{SNR} = P_{\text{signal}}/P_{\text{noise}}$, where P is power, which is proportional to corresponding variances: $\text{SNR} = \sigma_{\text{signal}}^2/\sigma_{\text{noise}}^2$. Sometimes, an “amplitude” SNR is used, i.e., $\text{SNR} = \sigma_{\text{signal}}/\sigma_{\text{noise}}$.

In Fig. 11.4, the dependence of the r.m.s. σ_{mxc} on m is shown. The smallest marked value of $m = 5$ corresponds to the beginning of the wide minimum, which resembles a “standstill.” The largest value $m = 15$ corresponds to the 3σ criterion with a corresponding $\text{FAP} = 0.0034$. The middle value $m = 11$ corresponds to the minimal σ_{mxc} . Current data are of a good CCD accuracy, thus small systematic variations lead to larger optimal values of m than in the case of more noisy (e.g., visual) observations.

There are many weight (often called “window”) functions $p(z)$. The classical “finite-length” (if $-1 \leq z \leq 1$, zero outside) ones are the “rectangular” one $p(z) = 1$, the Bartlett function $p(z) = 1 - |z|$, the “const+cosine” function $p(z) = a_0 + (1 - a_0)\cos(\pi z)$ (for $a_0 = 0.5$ and $a_0 = 0.53836$, they are called the Hann and Hamming filter, respectively), and the infinite Gaussian $p(z) = \exp(-z^2/2)$. There are many “hybrid”

window functions, which are often multiplicative or additive combinations of other window functions (e.g., Prabhu, 2014). Andronov (1997) proposed a simpler function $p(z) = (1 - z^2)^2$, which does not need computation of exponents or trigonometric functions, and so is time consuming for computations.

11.4 NONLINEAR LS METHOD AND DIFFERENTIAL CORRECTIONS

Generally, it is suitable to distinguish between “linear” and “nonlinear” parameters in the approximation. The “linear” parameters are those for which the approximation depends on the corresponding parameter linearly (as in Eq. (11.8)). Instead, “nonlinear” parameters are included in the “nonlinear” basic functions f_α .

The most common example is a sine approximation

$$x_C[t] = C_1 + C_2 \cdot \cos(C_4 \cdot (t - t_0)) + C_3 \cdot \sin(C_4 \cdot (t - t_0)), \quad (11.40)$$

where C_1 , C_2 , C_3 are “linear” parameters and $C_4 = 2\pi/P = 2\pi f$ is a “nonlinear” one. Here $P = 2\pi/C_4 = 1/f$ is the period and f is frequency. The parameter C_4 is “nonlinear,” as the value of the smoothing function $x_C(t)$ changes with C_4 nonlinearly. The parameter t_0 may be arbitrary. It influences the coefficients C_2 and C_3 , but not the sum (Eq. (11.40)).

To determine its value, one may compute a sequence of values of the test function for a set of “nonlinear” parameters, determining the “linear” parameters using the LS described above. Then the set corresponding to the minimum is determined, which corresponds to the minimum of the test function Φ . Then we either decrease the step for the grid of parameters and determine the value of C_4 , or use iterations to determine a more

precise position of the minimum. The parameter t_0 is often set to zero, or to the beginning of observations. However, for faster convergence of the iterations, it is recommended to set it to a sample mean.

For this purpose, there are many methods; among them the most popular are (e.g., Press et al., 2007):

- Gauss–Newton algorithm,
- gradient descent algorithm,
- Levenberg–Marquardt algorithm,
- conjugate gradients algorithm,
- simplex method,
- coordinate descent,
- Monte Carlo (random arguments).

All these methods achieve more accurate determination of the parameters. But, to estimate statistical properties, the method of differential corrections (Gauss–Newton algorithm) should be finally used to allow estimating the matrix $R_{\alpha\beta}$ and further estimating the accuracy of coefficients and approximation itself. The main idea is to calculate the corrections to the nonlinear coefficients in such a way that the test function Φ should reach its deeper minimum. Let us enumerate the nonlinear parameters from $m + 1$ to $L = m + q$. Then the LS method is applied twice, separately for m initial equations with some input values of $C_{m+1}..C_L$, and later to the system of L equations for the residuals, i.e.,

$$\sum_{\alpha=1}^L A_{\alpha\beta} \cdot \Delta C_{\alpha} = \Delta B_{\beta}, \quad (11.41)$$

$$\Delta B_{\beta} = ((\vec{x} - \vec{x}_C) \cdot \vec{f}_{\beta}). \quad (11.42)$$

Here the basic functions $f_{\beta}(t) = \partial x_C(t) / \partial C_{\beta}$. For $\beta = 1..m$, they contain only “nonlinear” coefficients. However, additional basic functions with $\beta = m + 1..L$ contain both “linear” and “nonlinear” coefficients:

$$f_{\beta}(t) = \sum_{\gamma=1}^m C_{\gamma} \cdot \frac{\partial f_{\gamma}(t)}{\partial C_{\beta}}. \quad (11.43)$$

One may note that the inner part of the matrix $A_{\alpha\beta}$, $\alpha, \beta = 1..m$, is the same for both systems (of m and L equations), and $\Delta B_{\beta} = 0$ (within rounding errors) for $\beta = 1..m$.

The next step is to add differential corrections ΔC_{α} to the input values of C_{α} and replace them: $C_{\alpha} + \Delta C_{\alpha}$. These iterations are repeated while all $|\Delta C_{\alpha}|$ will decrease below some limiting accuracy ε , or the number of iterations will not exceed some limiting value (e.g., 30). If the initial “guess” of the parameters is good, only few iterations are needed to reach the “ ε ” limit. If not,

the Levenberg–Marquardt method is used. It is based on adding to the diagonal elements of the matrix $A_{\alpha\beta}$ values of $\lambda > 0$ (Levenberg, 1944) or $\lambda \cdot A_{\alpha\alpha}$ (Marquardt, 1963). That is, only the diagonal elements of the matrix $A_{\alpha\alpha}$ are multiplied by a factor of $(1 + \lambda)$, so the modified LS equations

$$\sum_{\alpha=1}^L (A_{\alpha\beta} + \lambda \cdot \delta_{\alpha\beta}) \cdot \Delta C_{\alpha} = \Delta B_{\beta} \quad (11.44)$$

are solved (Marquardt, 1963). After moving of the iterations from the “risk zone” with large λ , its value should be decreased to a final value of zero to allow correct values of the matrices $A_{\alpha\beta}$ and $R_{\alpha\beta}$. These methods are similar to the “Tikhonov regularization” (Tikhonov, 1963). A very important point is to choose a correct initial point, as it may lead to, instead of a global minimum, a local minimum, or even a maximum. If the initial values are close to the solution, the iterations converge rapidly. Sometimes, after an iteration, the value of the test function may become larger. In this case, one may use smaller steps and move to a closer point $C_{\alpha} + \lambda_{\text{step}} \cdot \Delta C_{\alpha}$, $0 < \lambda_{\text{step}} \leq 1$. This will slow down convergence of the iterations, but may make the interval of convergence wider. In a simpler (but generally slower) method of “steepest descent,” no inverse matrix is needed, and one may just estimate $\Delta C_{\alpha} = \Delta B_{\alpha} / A_{\alpha\alpha}$. Asymptotically, for large $\lambda \gg A_{\alpha\alpha}$ for all α , the direction of the vector of differential corrections changes from that for “differential corrections” to that of the “steepest descent” with $\lambda_{\text{step}} \approx 1 / (1 + \lambda) \ll 1$.

The Monte Carlo method needs ranges for all available parameters instead of the initial point of iterations. Its realization is the easiest in the computer program, but needs too many test samples N_C to get accuracy δ_C : $N_C \approx \delta_C^{-q/2}$, where q is the number of (naturally, “nonlinear”) parameters, for which the Monte Carlo search is applied (Andronov and Tkachenko, 2013).

11.5 NONUNIQUE MINIMUM OF THE TEST FUNCTION

Sometimes the data do not allow to determine all physical parameters needed, as the information is not sufficient. For example, a visible magnitude m is related to the absolute magnitude M and the distance r in parsec: $m = M - 5 + 5 \cdot \lg r$. The observational parameter is m , and, from one equation, it is not possible to get two unknown parameters M and r . However, to determine the distance r from other measurements, e.g., parallax from ground-based or space

(HIPPARCOS, GAIA) observatories, one may determine an absolute magnitude separately. Similarly, the half duration of the eclipse is related to radii of both stars (assumed to be spherically symmetrical), the distance between them, and the inclination i (Shul'berg, 1971; Andronov and Tkachenko, 2013). Thus one may determine less phenomenological parameters than the physical ones. Or, for the same values of the phenomenological parameters, one may get a region of physical parameters, which are thus poorly defined separately.

Another problem may appear if there are few minima of the test function for different sets of parameters. Sometimes, these minima are of comparable depth, and occasionally the deepest minimum may correspond to a wrong set. This is why the results from discovery papers should be checked and corrected with better accuracy, using additional further observations. These simple examples show necessity of complementary methods.

11.5.1 Bootstrap Method

An alternate method for determination of the accuracy of the model parameters is the so-called "bootstrap" method (Efron, 1979; Efron and Tibshirani, 1993; Shao and Tu, 1996). The parameters are determined for the initial data, and used as a solution. Then new datasets are generated using random numbers $j = \text{int}(\text{random} \cdot n) + 1$, where "random" is a pseudorandom number, which is uniformly distributed in a range (0, 1). So some initial data points are missing, and other ones may be used one or a few times.

Andrych et al. (2020) and Andronov and Kulynska (2020) discussed statistical properties of the approximations of the "bootstrap-generated" data sets in more detail.

Generally, the sample distribution may be asymmetrical. The sample mean of a given parameter may differ from the initial value, so one may use, as an accuracy estimate, the r.m.s. deviation of the generated value from the initial one. Sometimes, instead of single σ , there are asymmetrical positive (σ_+) and negative (σ_-) errors corresponding to, e.g., the 95% confidence interval.

This challenges an usual assumption on Gaussian distribution of statistical errors. Moreover, this type of the confidence interval is not consistent with the definition of weights. For the normal distribution, this interval is 1.96 times larger than the standard error. Approximately, for a sample value, one may just divide by this factor. This factor is larger for the Student distribution for a smaller number of degrees of freedom. Another disadvantage of the bootstrap method is that, due to a decrease of the number of different arguments, and hence much larger gaps, the "best" approximation may

be unrealistically shifted as compared to that for the real sample.

11.5.2 Determination of Times of Minima/Maxima (ToM)

There is a special kind of analysis of the period and its possible changes, based on the "times of minima/maxima" (ToM) (AAVSO) or the "moments of characteristic events" (Tsesevich, 1973; Dumont et al., 1978; Andronov, 1988). Then, from many "near-extremum" observations, the only information that is extracted is the moment of time, which corresponds to a minimum or maximum of the approximation. Some of the methods were discussed by Andronov (2005).

For the moment of extremum t_e (maximum or minimum, what is commonly used to compile international databases) is determined as the root of equation $\dot{x}_C(t_e) = 0$ and has a standard error of

$$\sigma[t_e] = \frac{\sigma[\dot{x}_C(t_e)]}{|\ddot{x}_C(t_e)|}. \quad (11.45)$$

For example, for a simplest parabolic fit

$$x_C[t] = C_1 + C_2 \cdot (t - t_0) + C_3 \cdot (t - t_0)^2, \quad (11.46)$$

$$t_e = t_0 - C_2/(2C_3) \text{ and}$$

$$\sigma^2[t_e] = \frac{R_{22}C_3^2 - 2R_{23}C_2C_3 + R_{33}C_2^2}{4C_3^4}. \quad (11.47)$$

For the polynomial of arbitrary order $s = m - 1$

$$x_C[t] = C_1 + C_2 \cdot (t - t_0) + C_3 \cdot (t - t_0)^2 + \dots + C_m \cdot (t - t_0)^{m-1}, \quad (11.48)$$

the position of the extremum is determined numerically by solving the equation $\dot{x}_C[t_e] = 0$ using the Newton-Raphson method of iterations $t_e := t_e + \delta t_e$, where $\delta t_e = -\dot{x}_C[t_e]/\ddot{x}_C[t_e]$, until $|\delta t_e| \leq \epsilon$, where ϵ may be set to the desired accuracy. In practice, this may be a computer accuracy, when $t_e + \delta t_e - t_e$ is equal to zero because of the rounding errors. The starting point is determined on a regular grid within the given data interval. The type of the extremum corresponds to that from the parabolic approximation ($s = 2$).

The statistical error is estimated similarly to Eq. (11.31):

$$\sigma[t_e] = \frac{\sigma[\dot{x}_C(t_e)]}{|\ddot{x}_C(t_e)|}. \quad (11.49)$$

As the polynomials are most common functions used for approximations (e.g., in the electronic tables),

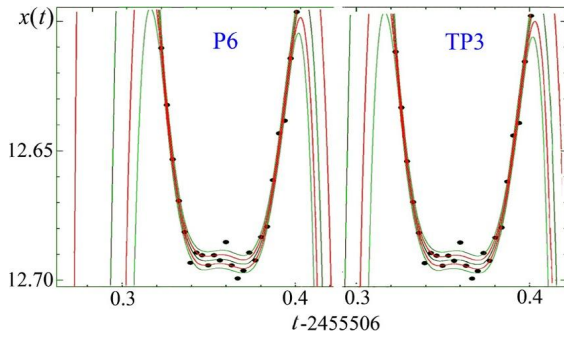


FIG. 11.5 Approximations of the eclipse of the binary system GSC 3692-00624 = 2MASS J01560160+5744488 using a short data sample published by Devlen (2015). *Left:* Algebraic polynomial of degree $m - 1 = 6$. *Right:* Trigonometric polynomial of degree $(m - 1)/2 = 3$. The $\pm 1\sigma$ and $\pm 2\sigma$ “error corridors” are shown.

they were often used for the ToM determination. For example, the catalogue of 6509 extrema of semiregular variables (Chinarova and Andronov, 2000) was used using this method. The number of the parameters $m = s + 1$ was determined separately for each interval of data near extremum. The same algorithm was implemented in other software (Breus, 2007; Andrych et al., 2015). In practice, we test orders s up to eight only. In the popular commercial software PERANSO (Paunzen and Vanmunster, 2016), the degree of the polynomial s is user-defined, so the “esthetic” method is applied.

In Fig. 11.5, the approximations of the short observational run near the minimum are shown, using the same number of parameters $m = 7$ for the sixth-order algebraic polynomial and the third-order trigonometrical polynomial. Remarkable coincidence of both approximations in the interval of observations is observed, but a drastic difference outside the interval. These “common” approximations show a bad approximation at the bottom of eclipse, where a statistical error of one point “makes” the minimum “split.”

To avoid this physically unreal behavior, we have proposed other phenomenological approximations, which were applied to the same data by Andrych et al. (2017). They are shortly described below.

For theoretically symmetrical signals (e.g., eclipses), it is natural to use symmetrical functions, e.g., polynomials of order $s = 2(m - 1)$, i.e.,

$$x_C[t] = C_1 + C_2 \cdot (t - t_e)^2 + C_3 \cdot (t - t_e)^4 + \dots + C_m \cdot (t - t_0)^{2(m-1)}. \quad (11.50)$$

Here $t_e = C_{m+1} = C_L$ is the symmetry point which corresponds to the extremum and is determined numer-

ically using the Newton–Raphson method. Obviously, for $m = 3$, an either “ordinary” or “symmetrical” polynomial is a parabola, and there is a difference in the form of two functions, but not their values or properties. For polynomials of higher orders, the absence of terms with odd power indices keeps the function symmetrical.

Even for a “complete” polynomial with both odd and even power indices, the form

$$x_C[t] = C_1 + C_2 \cdot (t - t_e)^2 + C_3 \cdot (t - t_e)^3 + \dots + C_m \cdot (t - t_0)^m \quad (11.51)$$

is preferable, as t_e is the moment of extremum, even if there may be up to $(m - 1)$ real roots of the equation $\dot{x}_C[t_e] = 0$.

Although polynomials are commonly used for approximations, they are not the best functions because of the Gibbs effect and apparent waves at the approximation. Smaller amplitudes of waves show cubic splines (Andronov, 1987). Marsakova and Andronov (1996) have proposed a spline with varying degree and subintervals – the “asymptotic parabola,” which consists of two straight lines connected with a parabola. This is effective for asymmetric extrema, e.g., in the light curves of pulsating variables of the types RR Lyrae, δ Cephei, and Mira et al.

It is suitable to write the approximation in the form

$$x_C[t] = C_1 + C_2 \cdot G((t - C_3), C_4, \dots, C_{mp}) \quad (11.52)$$

to determine two “linear” parameters C_1, C_2 and $L - 2$ “nonlinear” parameters C_3, \dots, C_L . If choosing the functions G in such a way that $G = 0$ at $t = C_3$, the explanation of the “linear” parameters is simple: C_1 is the smoothed signal value at the extremum. With an additional condition $G \rightarrow 1$ far away from the extremum, C_2 is equal to the amplitude, i.e., the difference between the “extremal” and “quiet” values of the approximation.

If the “quiet” part is present, one may use another definition, $x_C[t] = \tilde{C}_1 + \tilde{C}_2 \cdot \tilde{G}((t - C_3), C_4, \dots, C_{mp})$, with obvious relations between the coefficients $\tilde{C}_1 = C_1 - C_2$, $\tilde{C}_2 = -C_2$ with the same values of other parameters.

For the typical near-extremum observations, there is no “quiet” part, so C_2 may be unrealistically large, even formally “infinite.”

In this case, one should use a “restricted” model – e.g. setting at least one of the “nonlinear” parameters to some limiting value and recomputing the fit for a smaller number of parameters L . In some cases, the “inner interval” is a wide as the observations. Then

the model simplifies to a “singular interval” one, and may be an ordinary parabola, as, e.g. in the method of “asymptotic parabola” (Andrych et al., 2015) and “parabolic spline” (Andrych et al., 2020).

In some cases, we can just use power series (sometimes even noninteger power indices; e.g., Andronov et al., 2017b).

For the “global” approximation with a single analytical function, Andronov (2005) proposed

$$\begin{aligned} \tilde{G}((t - C_3), C_4, C_5) \\ = \frac{2}{\exp(C_4 \cdot (t - C_3)) + \exp(-C_5 \cdot (t - C_3))}. \end{aligned} \quad (11.53)$$

This is an extension of the classical hyperbolic secant function $\text{sech}(z) = 2/(e^z + e^{-z})$ for the case of asymmetrical ascending and descending branches. The connection to physics of the process is that $\tau_- = 1/C_4$ and $\tau_+ = 1/C_5$ are characteristic times of exponential rise/decay at the beginning/end. The accuracy estimate is $\sigma[\tau_-] = \sigma[C_4]/C_4^2$ and, similarly, $\sigma[\tau_+] = \sigma[C_5]/C_5^2$. The position of the extremum is shifted; we now have

$$t_e = C_3 + \ln(C_5/C_4)/(C_4 + C_5). \quad (11.54)$$

The accuracy of this function of three coefficients C_3 , C_4 , and C_5 is estimated using Eq. (11.25).

The following analytical approximation, which is similar to a probability distribution, was proposed by Bódi et al. (2016):

$$\begin{aligned} \tilde{G}((t - C_3), C_4, C_5) \\ = \exp(-\ln 2 \cdot C_5 \cdot (\ln(C_4 \cdot (t - C_3) + 1))^2). \end{aligned} \quad (11.55)$$

These two functions are time consuming, because of numerous computations of exponents and logarithms during a “brute force” determination of three parameters before using the differential corrections. Also, during the iterations, the values should be checked for being in a reasonable interval. For example, for the latter function, $C_4 \rightarrow 0$ for an exactly symmetrical signal, which causes $C_2 \cdot C_5 \rightarrow \infty$. In this case, the function should be changed to a symmetrical polynomial, or to a Gaussian

$$\tilde{G}((t - C_3), C_4) = \exp(-C_4 \cdot (t - C_3)^2). \quad (11.56)$$

This method is also widely applied, as some software uses it to fit spectral lines.

However, for symmetric signals similar to the eclipses of the eclipsing binaries, it is useless, as the eclipses are

of finite length. Moreover, the eclipse duration is an optional parameter to be listed in the “General Catalogue of Variable Stars” (Samus et al., 2017). Thus one should use approximations of a finite length. Andronov (2012) proposed a “New Algol Variable” (NAV) function

$$\tilde{G}((t - C_3), C_4) = (1 - (|t - C_3|/C_4)^{C_5})^{1.5}, \quad (11.57)$$

where the parameter C_5 determines the “flatness” of the shape near the mid-eclipse, i.e., $C_5 = 1$ corresponds to a lower physical limit (when the stars have the same size, and the total eclipse is very short as compared to the eclipse duration); $C_5 = 2$ corresponds to the “classical” mathematical function, which has a nonzero second derivative at the extremum. With increasing C_5 , the shape becomes flatter, and very large values may correspond to short ascending (or descending) branch, which is typical for exoplanet transits. The power 1.5 asymptotically describes the shape of the eclipse of spherical (or even ellipsoidal) stars close to the outer contact. Added to a trigonometrical polynomial, the NAV function is effective not only for the Algols, but also for EB and EW – eclipsing systems with more smooth variations than in EA (Tkachenko et al., 2016).

Mikulášek (2015) introduced few functions improving the Gaussian. At first, the simple parabola in the argument was replaced by a hyperbolic cosine, i.e.,

$$\tilde{G}((t - C_3), C_4) = \exp(1 - \cosh((t - C_3)/C_4)). \quad (11.58)$$

This function tends to zero faster than a Gaussian, but still has a parabolic shape close to mid-eclipse. Thus also a “noninteger” power shape was introduced, i.e.,

$$\begin{aligned} \tilde{G}((t - C_3), C_4) \\ = 1 - (1 - \exp(1 - \cosh((t - C_3)/C_4)))^{C_5}. \end{aligned} \quad (11.59)$$

Andrych et al. (2017) introduced a series of symmetrical functions, where the interval of observations is split into three subintervals (physically, the beginning/ending and middle branches are described using different formulas). They called these functions “wall-supported” (WS) ones. The WS parabola seems the best for describing the exoplanet transits. The WS line is good for total eclipses of stars of comparable sizes. WS “asymptotic parabola” is good for the intermediate cases. For an extreme case of very wide interval beginning close to the previous extremum and ending close to the next one, we have introduced the quadratic spline. In all these functions, the positions of the points splitting the interval are “nonlinear” parameters, which are determined to get the best fit.

For wider intervals, which contain completely the ascending and descending branches, Andronov et al. (2017b) tested almost a half hundred modifications of the shapes (“patterns”).

Obviously, a wide variety of functions needs numerical criteria to choose the best one. Currently, all these 21 methods are implemented in the software MAVKA, which has an option for automatic determination of the method with best accuracy estimate (Andrych and Andronov, 2019) from a list of chosen function(s).

11.6 PERIODOGRAM ANALYSIS: PARAMETRIC VERSUS NONPARAMETRIC METHODS

11.6.1 From Time to Phase

The truly periodic signal satisfies the condition $x(t + k \cdot P) = x(t)$ for any integer k , and P is the period. According to this definition, the values $2P$, $3P$, etc., are also “periods.” So it is usually adopted to use the minimal positive value as the “period.”

In some cases, the physical period is $2P$, e.g., for eclipsing binary stars of the EW-type or for the elliptical variables. Their light curves have two similar waves, which are mirror (reflection) symmetrical in respect to the primary (deeper) or secondary minimum. The difference between these “reflected” parts is usually within observational errors. In this case, the period P is called the “photometric” period, or the “formal” period, while $2P$ is called the “true,” “physical,” or (for binary stars) the “orbital” period.

The main idea for the periodic functions is to “pack” all the data into a single interval. Typically, the time is thus shifted by an integer number of cycles, so an age may be introduced, $\tau_P = t - T_0 - P \cdot E$, where T_0 is called the initial epoch and $E = \text{INT}((t - T_0)/P)$ is the cycle number. Here the function INT is defined as the largest integer which does not exceed unity. For example, $\text{int}(-2.7) = -3$. In this case, $E = -3$, $\phi = +0.3$. In some computer languages, the value of this function is set to -2 , so a correction is needed. This should be checked in concrete programming environments.

The “age” is typical for an everyday life situation, e.g., time measured in 12- or 24-hour format (or a decimal part of the Julian Day). It is measured in units of time. To scale the signals with different periods, the dimensionless “phase” $\phi = \tau_P/P$. So $t = T_0 + P \cdot E + \tau_P$, $= T_0 + P \cdot (E + \phi)$.

The astronomical definition of the phase is different from an usual definition in physics and mathematics $\phi = 2\pi\phi$, i.e., in radians.

So $0 \leq \tau_P < P$, $0 \leq \phi < 1$. However, this range may be extended for some methods and for better illustration. For example, the phase $\phi = 0.99$ is the same as $\phi = -0.01$. In Fig. 11.6, the computed nonsinusoidal signal is shown for a random distribution of time. The points should be shifted by an integer number of periods to “be moved” to the main interval.

11.6.2 “Parametric” (“Point-Curve”) Methods

The periodogram analysis is based on estimate of the “quality” of the phase light curve by some parameter, which is called “the test function.” This test function is computed for different trial periods P or, alternately, frequencies $f = 1/P$. Traditionally, the periodogram analysis is divided into two large groups, which are called parametric (or “point-curve”) or nonparametric (“point-point”). In the first group, the phase curve is compared to some approximation (smoothing curve). The test functions are used similarly to Φ in Eq. (11.1), so the position of its minimum (as a function of the period P , rarely also of the initial epoch T_0) is to be found. However, in many cases, the amplitude of the signal is small as compared to the noise, so the relative amplitude of the test function is not too large, so it is far enough from the zero level. So it may be suitable to introduce the test function

$$S(f) = r^2 = 1 - \Phi_{m+q+1}/\Phi_{q+1}. \quad (11.60)$$

Typically, the value $q = 0$ is used, i.e., the approximation by a constant. However, this definition of the periodogram was used by Andronov and Baklanov (2004) for the periodogram analysis with a trend approximated by the polynomial of order q .

The trigonometric polynomial is most often used for the periodogram analysis. However, other periodic functions may be used, e.g., “piecewise constant” splines (Jurkevich, 1971). This method was improved by Stellingwerf (1978), who partially removed the dependence of the test function on the initial epoch. This algorithm is called “phase dispersion minimization” (PDM) and is often used. The test functions are dependent on the number of intervals (bins) m , which is a free parameter. We recommend to use at least $m \geq 3$ for expected near-sinusoidal variability and $m \geq 5$ for double-peaked curves of EW-type stars. However, for light curves with sharp parts (eclipsing binaries, RR Lyrae-type stars), the number m should be increased so that at least two subintervals cover such sharp parts. Generally, according to the Sturges (1926) rule, the number of bins should be $m \approx 1 + 3.32 \cdot \lg n \approx 1 + \log_2 n$.

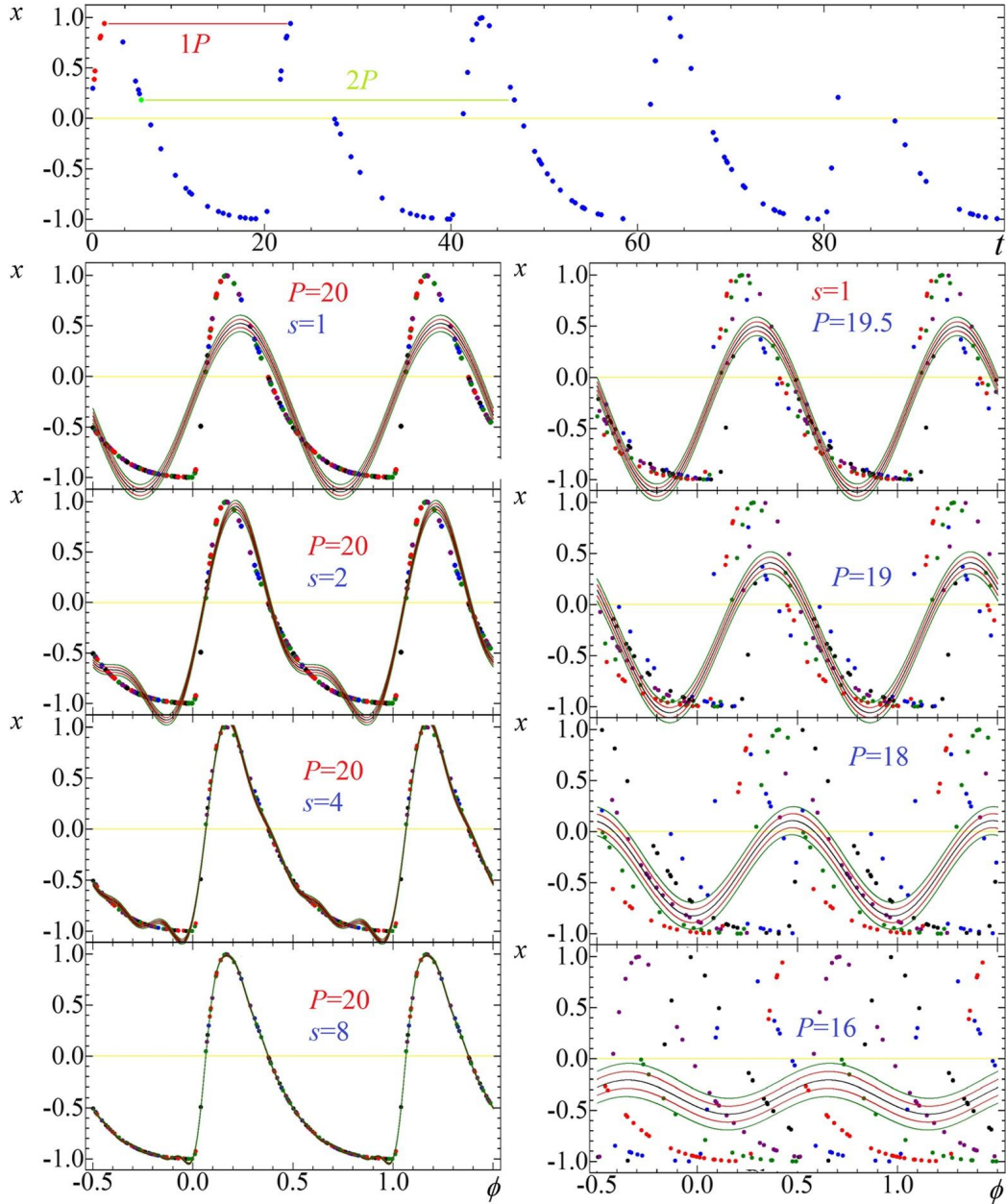


FIG. 11.6 *Top:* The simulated truly periodic signal with a period $P = 20$ with a highly asymmetrical shape. Such shape is similar to periodic flashes (or, in astronomy, to RR Lyr-type stars). The points are shifted by an integer number of periods P inside the preferred interval. *Left:* Trigonometric polynomial approximations of orders $s = 1, 2, 4, 8$ for the true period. *Right:* Phase curves for trial periods, which deviate from the true period, with a cosine (TP1) approximation. Points are shown in different colors, to see that different "seasons" show good curves, but they are shifted between each other. The $\pm 1\sigma$ and $\pm 2\sigma$ "error corridors" are shown. The approximations were made using MCV software (Andronov and Baklanov, 2004).

Andronov (1987) proposed to use more smooth cubic splines, and, additionally, to remove dependence on T_0 either by smoothing the cubic splines, or to determine the best phase shift corresponding to a minimum of the test function.

11.6.3 “Nonparametric” (“Point-Point”) Methods

In another group of methods, which do not assume the basic functions, the distance between the points (different in different methods) is taken into account. The most famous method was published by Lafler and Kinman (1965), typically abbreviated to LK.

The test function is

$$\Theta = C \sum_{k=1}^n \theta(\Delta x_k, \Delta \phi_k), \quad (11.61)$$

where $\Delta x_k = x_k - x_{k-1}$, $\Delta \phi_k = \phi_k - \phi_{k-1}$, and, “formally,” $x_0 = x_n$, $\phi_0 = \phi_n - 1$. Its minimum corresponds to the best period. Here the data are sorted according to the phases ϕ_k for each trial period. The summand $\theta(\Delta x_k, \Delta \phi_k)$ is some kind of “distance” between the two points, which are subsequent in phase. The scaling parameter $C > 0$ is arbitrary. It may be set, e.g., to unity, or to make the mathematical expectation for a pure noise to 1, 2, or any positive value.

In the original LK method, $\theta(\Delta x, \Delta \phi) = (\Delta x)^2$, and the scaling factor $C = \sigma_x^{-2}/n$ is inversely proportional to the variance of the data σ_x^2 .

There were numerous modifications of the method, e.g., $\theta = |\Delta x|$ (see the appendix by Deeming to Bopp et al., 1970), $|\Delta x|^\gamma$ (Pelt, 1975), $(\Delta x)^2/((\Delta \phi)^2 + \epsilon^2)$, $|\Delta x|/(\Delta \phi + \epsilon)$ (Renson, 1978), $\sqrt{((\Delta x)/\epsilon)^2 + (\Delta \phi)^2}$ (Dworetzky, 1983), etc.

The comparative study of these methods was presented by Andronov and Chinarova (1997) with recommendations on values of ϵ .

Pelt (1975) proposed a more general relation, taking into account the distance measure not between the pairs of subsequent in phase points, but with neighbors in some interval of phases, i.e.,

$$\Theta = C \sum_{k=1}^n \sum_{j=1}^{k-1} (x_k - x_j)^2 \cdot \tilde{\theta}(|\phi_k - \phi_j|), \quad (11.62)$$

where the simplest weight function $\tilde{\theta} = 1$ if $|\phi_k - \phi_j| \leq \Delta \phi_{max}/2 \ll 1/2$.

The full width of this interval $\Delta \phi_{max}$ is an additional free parameter. If it is narrow, the fluctuations of the periodogram are large.

There is no single best method among these; otherwise the others could “go to history.” Some modifications called “string/rope length methods” were discussed by Clarke (2002). In practice, the LK method is the most popular one among “point-point” the ones, and it has been implemented in many computer programs, e.g., VSCalc (Breus, 2007), Peranso (Paunzen and Vanmunster, 2016), etc. According to the ADS, this paper has been cited 652 times already in 2019.

Periodograms are discontinuous, so it is often possible to get a local minimum, which is shifted from the “true” position.

The periodograms have the deepest minimum at the main period, whereas there are minima at $2P$, $3P$, kP , etc. The depth gradually decreases with the multiplier k , as the same number of points are distributed less densely at the main wave, so the systematic differences between the points become relatively large. The mathematical expectation of the normalized function Θ for a pure noise (no signal) in the LK method is 2.

For a better apparent contrast of the minimum of the test function Θ to the “noisy continuum” at the periodogram, it should be recommended to use $\lg \Theta$ instead of Θ itself.

This behavior is opposite to that of the one based at the sinusoidal approximation, where the possible peaks appear at multiple frequencies, rather than periods. The width of the peaks is nearly constant.

The periodogram for the two-point distance is discontinuous. So the minimal value of the test function on a grid of frequencies or period s may be shifted from an expected one, even if the signal is an accurate periodic function without noise.

The recommended frequency step is $\Delta f = \Delta \phi / (t_n - t_1)$, where $\Delta \phi$ is the change of the phase difference between the first and the last observations.

For the sinusoidal signals, a value of $\Delta \phi$ between 0.04 and 0.06 is recommended, but not worse than 0.1. For the trigonometric polynomial of order s , the recommendation is $\Delta f = \Delta \phi / (t_n - t_1) / s$. In the FT (see next section), $\Delta \phi = 1 - 1/n$. But there the period is exactly fixed to the duration of the observations $n\delta = (t_n - t_1) / (n - 1) \cdot n = jP$. The periodograms of different kind are shown in Fig. 11.7.

11.7 WHAT IS THE “ORTHODOX” FOURIER TRANSFORM FOR DISCRETE DATA?

The FT is one of the most popular methods for data analysis, as well as for solving tasks in mathematical physics and other directions of mathematics.

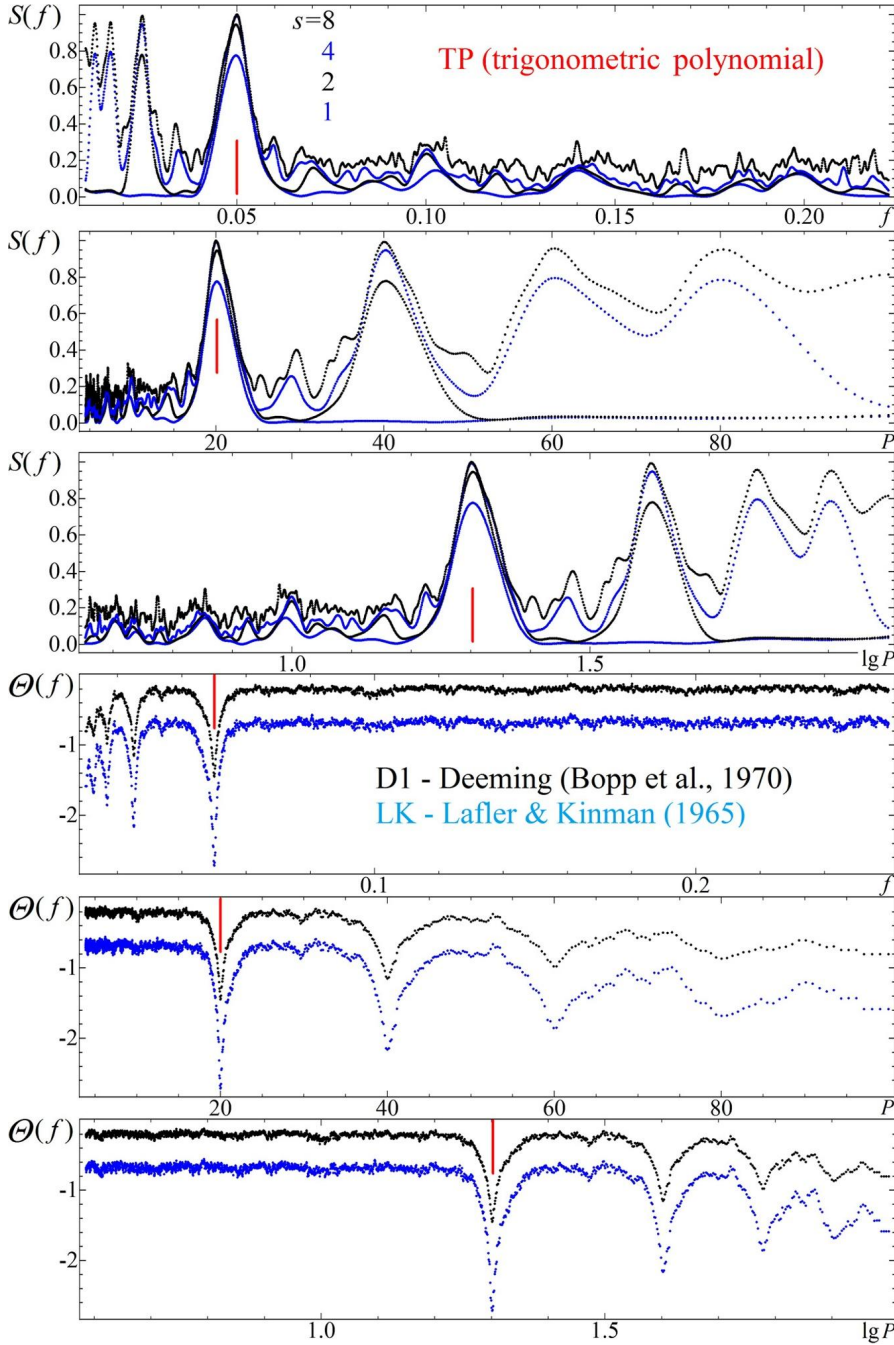


FIG. 11.7 Top: The simulated truly periodic signal with a period $P = 20$ with a highly asymmetrical shape. Such a shape is similar to periodic flashes (or, in astronomy, to RR Lyr-type stars). Periodograms for “parametric” or “nonparametric” are computed using the TP, LK, and D1 methods and are represented as functions of $f = 1/P$, P , and $\lg P$.

A Google search shows almost 38 million links to publications for the “Fourier Transform,” among them thousands of books containing the description (e.g., Anderson, 1958; Bendat and Piersol, 2010; Press et al., 2007). Excellent reviews on FT and the discrete Fourier transform (DFT) are posted on Wikipedia. There are many versions with different designations.

The original work was published by Fourier (1822) almost two centuries ago. One of the classical monographs dedicated solely to FT was published by Tolstov (2012).

The classical FT is typically defined as

$$\hat{x}(f) = \int_{-\infty}^{+\infty} x(t) \cdot e^{-i2\pi ft} dt, \quad (11.63)$$

with an inverse transform

$$x(t) = \int_{-\infty}^{+\infty} \hat{x}(f) \cdot e^{i2\pi ft} df. \quad (11.64)$$

Here t is time and $f = 1/P$ is frequency, P is period, and $i^2 = -1$. As the integrals should be limited, the integral of the function $|x(t)|$ should exist.

The FT is a very powerful tool used in numerous analytical studies, e.g., mathematical physics, statistics, etc. In reality, it has strong limitations due to the absence of infinite information from the observed signals. The most common adaptation of the method to discrete data may be called DFT, which also has some versions.

It is assumed that the signal is defined at a set of discrete points $t_k = t_0 + k\delta$, $k = 0..n_1$, where $n_1 = n - 1$. The inverse FT becomes a sum:

$$\begin{aligned} x_c(t) &= C_1 + \sum_{j=1}^s (C_{2j} \cos(2\pi jft) + C_{2j+1} \sin(2\pi jft)) \\ &= C_1 + \sum_{j=1}^s R_j \cdot \cos(2\pi jf(t - T_{0j})), \end{aligned} \quad (11.65)$$

which is called the “trigonometrical polynomial” of order s .

The relations are listed as

$$\begin{aligned} C_{2j} &= R_j \cdot \cos(2\pi jfT_{0j}), \\ C_{2j+1} &= R_j \cdot \sin(2\pi jfT_{0j}), \\ R_j &= \sqrt{C_{2j}^2 + C_{2j+1}^2}, \\ T_{0j} &= \text{atan2}(C_{2j+1}, C_{2j}/(2\pi) + k) \cdot P/j, \end{aligned} \quad (11.66)$$

where R_j is called “semiamplitude,” i.e., the difference between the maximum deviation of the wave from its mean value, so $2 \cdot R_j$ is a full amplitude between the maximum and minimum of the corresponding wave, and T_{0j} is called the initial epoch (moment of time, which corresponds to the maximum of the wave).

One may choose any integer value of k in this equation – e.g., if we say that the Sun is highest close to noon, “noon” is the initial epoch. Of which day? Of each. Many authors use the earliest initial epoch occurring during the observations. However, the best results for the matrix of normal equations will be obtained choosing T_{0j} closest to the sample (weighted) mean value of times of the observations.

It is also important to note that the brightness in astronomy is measured in “stellar magnitudes,” so the minimum of brightness corresponds to a minimum of the stellar magnitude. This should be clearly written in the papers to avoid misinterpretation by other authors.

Terminologically, the wave j is called the $(j - 1)$ -th harmonic of the main period ($j = 1$). However, some authors use the j -th wave as the j -th harmonic. It is some type of scientific slang, as $j = 1$ is a main wave, and not its harmonic. This also could lead to misinterpretation.

Here it is suggested that the signal repeats from $-\infty$ to $+\infty$ with a period $P = n\delta/j$. Obviously, it is not possible to determine more parameters than there are observations, thus the number of frequencies is limited to $s_{max} = \text{int}(n/2)$. The set of the basic functions is $f_1 = 1$, $f_{2j} = \cos(2\pi jft)$, $f_{2j+1} = \sin(2\pi jft)$.

The set of frequencies for the classical FT is discrete ($f_j = 2\pi j/n$), and the coefficients may be easily determined as

$$\begin{aligned} C_1 &= \frac{1}{n} \sum_{k=1}^n x_k, \\ C_{2j} &= \frac{2}{n} \sum_{k=1}^n x_k \cos(2\pi kj/n), \\ C_{2j+1} &= \frac{2}{n} \sum_{k=1}^n x_k \sin(2\pi kj/n). \end{aligned} \quad (11.67)$$

If n is even, then $j_{max} = s = n/2$, and then the coefficient C_n should be twice smaller than in the equation above, and C_{n+1} is not used, or is set to zero.

An example of inadequate use of the FT is shown in Fig. 11.8. The sums Eq. (11.65) converge to a function, which has a discontinuity. So the point at the border is a mean from the values left and right from this point. There are apparent waves (the Gibbs phenomenon) of

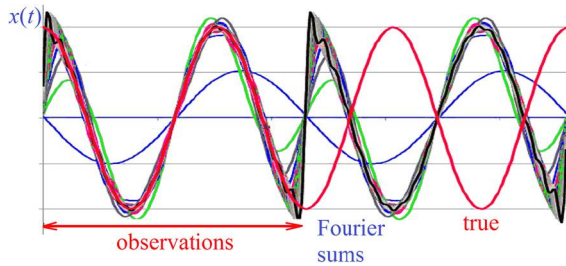


FIG. 11.8 Approximations of a cosine function “observed” during $0.75P$ by Fourier sums (trigonometrical polynomials). For the discrete Fourier transform (DFT), the “period” is assigned to the duration of the data $n\delta$, and not to the true period. The Fourier sums smooth jumps, the slope of which increases with increasing degree of the trigonometrical polynomial s .

no physical meaning, and an abrupt switch between the levels of discontinuous function.

This problem does not arise, if to determine a statistically optimal period from the data, not from the duration of observations.

For irregularly spaced data, the statistically correct decision is to use LS. However, there are some methods, the authors of which still call “Fourier transform,” even though the conditions are not satisfied.

The simplest formula was used by Deeming (1975). He just used

$$\begin{aligned}
 C_1 &= \bar{x}, \\
 C_2 &= \frac{2}{n} \sum_{k=1}^n x_k \cos(2\pi f t_k) = C(f), \\
 C_3 &= \frac{2}{n} \sum_{k=1}^n x_k \sin(2\pi f t_k) = S(f)
 \end{aligned}
 \tag{11.68}$$

as the approximation

$$x_C(t) = C_1 + C_2 \cos(2\pi f t) + C_3 \sin(2\pi f t). \tag{11.69}$$

This coincides with the Fourier coefficients only under conditions of orthogonality of the basic functions, which is generally not fulfilled. Moreover, the coefficients C_2 and C_3 are dependent on the zero point, so later it was recommended to use $(x_k - \bar{x})$ instead of x_k in the corresponding equations.

Lomb (1976) proposed a partial improvement, making a model as in equation (11.69), but fixing $C_1 = \bar{x}$, and using C_2 and C_3 from an LS approximation with $m = 2$ parameters. Scargle (1982) got the same periodogram, but shifting the initial phase for each trial frequency to make orthogonal the basic functions \vec{f}_2 and \vec{f}_3 . He also studied statistical properties of such a periodogram if the signal is pure noise and the accuracy of the data is known. The method was further referred to as the “Lomb–Scargle” method and is the most popular one, having 3851 citations in the papers listed in the ADS in 2019.

The problem in this method is neglecting nonorthogonality of the basic function \vec{f}_1 with two others. It is not so important if the observations cover the phase more or less homogeneously. However, many stars have periods, which are ~ 2 times longer than the typical duration of observations during the night.

In Fig. 11.9, there are examples of approximation of short runs using this most popular method of period search. For the near-extremum symmetrical data, the apparent period is equal to the duration of the data (in this sample, $0.5P$). For the descending branch, if we wish to remove the linear trend, the periodogram shows a peak at $0.36P$. At the same time, the complete three-parameter LS fit (Andronov, 1994a) produces an exact approximation, so a correct value of the period,

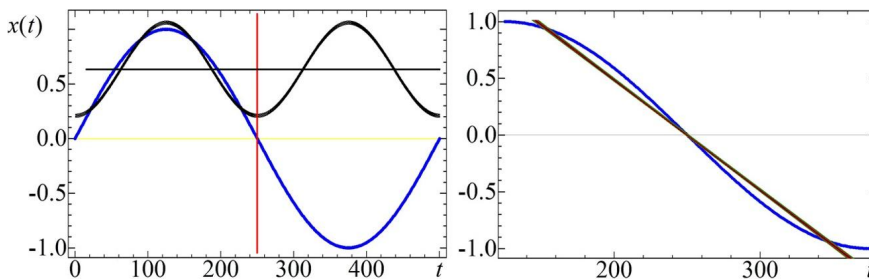


FIG. 11.9 Approximations of a sine function “observed” during $0.5P$ by the Lomb–Scargle method. *Left:* Approximation using the first half of the period. Because the mean value is shifted, the formal period in this method is $0.5P$, and the approximation has two apparent waves during one true period. *Right:* The middle half and the linear trend. The detrended data show an apparent period of $0.36P$.

amplitude, phase, and a zero point. The “sparing” of computer time was important in 1976, but now the computers are fast enough to determine parameters of complete models.

From the non-LS methods, there are “CLEAN” (Roberts et al., 1987) and “CLEANEST” (Foster, 1995). The similarity law in spectral estimation of time series was analyzed by Terebizh (1998).

11.7.1 LS-Based DFT

For a connection between the continuous and discrete FTs, one may refer to Wehlau and Leung (1964), where the discrete signal may be written as a multiplication of functions describing different types of irregularities of the distribution of the arguments of the signal. Among them, there is a smoothing of the data during the exposure time, distribution of data in time.

As the FT of the multiplication of functions is a convolution of their FTs, in the resulting spectrum, there will be biases, i.e., the peaks at the beat frequencies. This is clearly seen in ground-based photometrical surveys. The observations are carried out during a few months in the season. Thus the interval of phases slightly shifts from season to season. An apparent light curve shows an apparent longer period, which does not exist in a reality. Similarly, there may be problems if there are larger gaps between the nights even if a complete light curve is observed.

An example is shown in Fig. 11.10. The model data are a pure sinusoid with a true period ($P_0 = 1.1$), with a period of sampling $P_s = 1$. In the top left figure, the data show a larger “beat” period $P_{beat} = P_s \cdot P_0 / |P - P_s|$ (11, in our example) is present. For this data sample, the periodogram is the dependence of the test function $S(f) = S(|f + j \cdot f_s|)$ on trial frequency, and thus shows equal peaks, which correspond not only to the correct frequency $f_0 (= 1/P_0 = 0.90909)$, but for any frequency $f = |f_0 + j \cdot f_s|$, where j is any integer. Moreover, the periodogram is reflection symmetrical around frequencies $j \cdot f_s/2$. Thus it is recommended to use frequencies in the range $(0, f_s/2]$ if the data are equidistant. The value $f_s/2$ is called the “Nyquist” frequency f_N . In other words, one should have at least two points per period. This “main range” of frequencies does not prevent the periods to be shorter (as in this case). However, one may not distinguish between the peaks of equal height to choose the correct period.

The situation becomes better if there are observations shifted from the “main periodicity.” In Fig. 11.10, the “complete” dataset is with the time interval $\delta = 0.1$, so 10 points per unit time interval. The intermediate values of subsequent points are 2 and 5. It is clearly

seen that the height of the “bias” peaks strongly decreases with increasing number of points per P_s . Obviously, “the best” is the case of “no gaps.” Also, we have compared the periodograms for the same number of data per P_s for equidistant and “random” distribution of the arguments. The random distribution shows very low peaks as compared to the main peak with height $S(f_0) = 1 = 100\%$.

In these samples, we illustrated the influence of the distribution of data in time, with an exact sinusoid in the signal. Naturally, the observational noise of the signal x_k will add noise and spurious peaks at the periodogram.

To describe the peaks, sometimes it is recommended to show a “spectral window” of the observations. The spectral window may be defined as a complex function

$$\begin{aligned} W(f) &= \frac{1}{n} \sum_{k=1}^n e^{2\pi i f t_k} \\ &= \frac{1}{n} \sum_{k=1}^n (\cos(2\pi f t_k) + i \cdot \sin(2\pi f t_k)) = \tilde{C} + i\tilde{S}. \end{aligned} \quad (11.70)$$

For equidistant points $t_k = t_0 + k \cdot \delta$,

$$W(f) = e^{i\pi \cdot f \cdot (2t_0 + (n-1)\delta)} \cdot \frac{\sin(\pi f \delta n)}{n \sin(\pi f \delta)}. \quad (11.71)$$

The first multiplier has a unit absolute value, and the function

$$\tilde{W}(f) = \frac{\sin(\pi f \delta n)}{n \sin(\pi f \delta)} \quad (11.72)$$

is real, symmetric, and periodic with a period $1/\delta$. Some properties are $\tilde{W}(j/\delta) = 1$, $\tilde{W}(j/(n\delta)) = 0$ for noninteger ratios j/n . However, the values between these points are nonzero, biasing the periodogram.

For nonequidistant points, C and S do not cross zero at the same frequency, but one may determine the smallest value of f , which corresponds to the minimum of $|\tilde{W}(f)| = (C^2 + S^2)^{1/2}$. This value may be named Δ_{f_0} and may be used to estimate the “effective number of frequencies” N_{eff} in the periodogram analysis to estimate a false alarm probability (FAP) of a given peak (see Andronov, 1994a for more details).

To decrease the height of the peaks, it should be recommended, whenever possible, to break periodicity of the signal. For ground-based observations, it may be recommended to use telescopes at different longitudes.

The width of the peaks in the periodogram is constant for frequency and is proportional to P^2 for each peak, including the aliases!

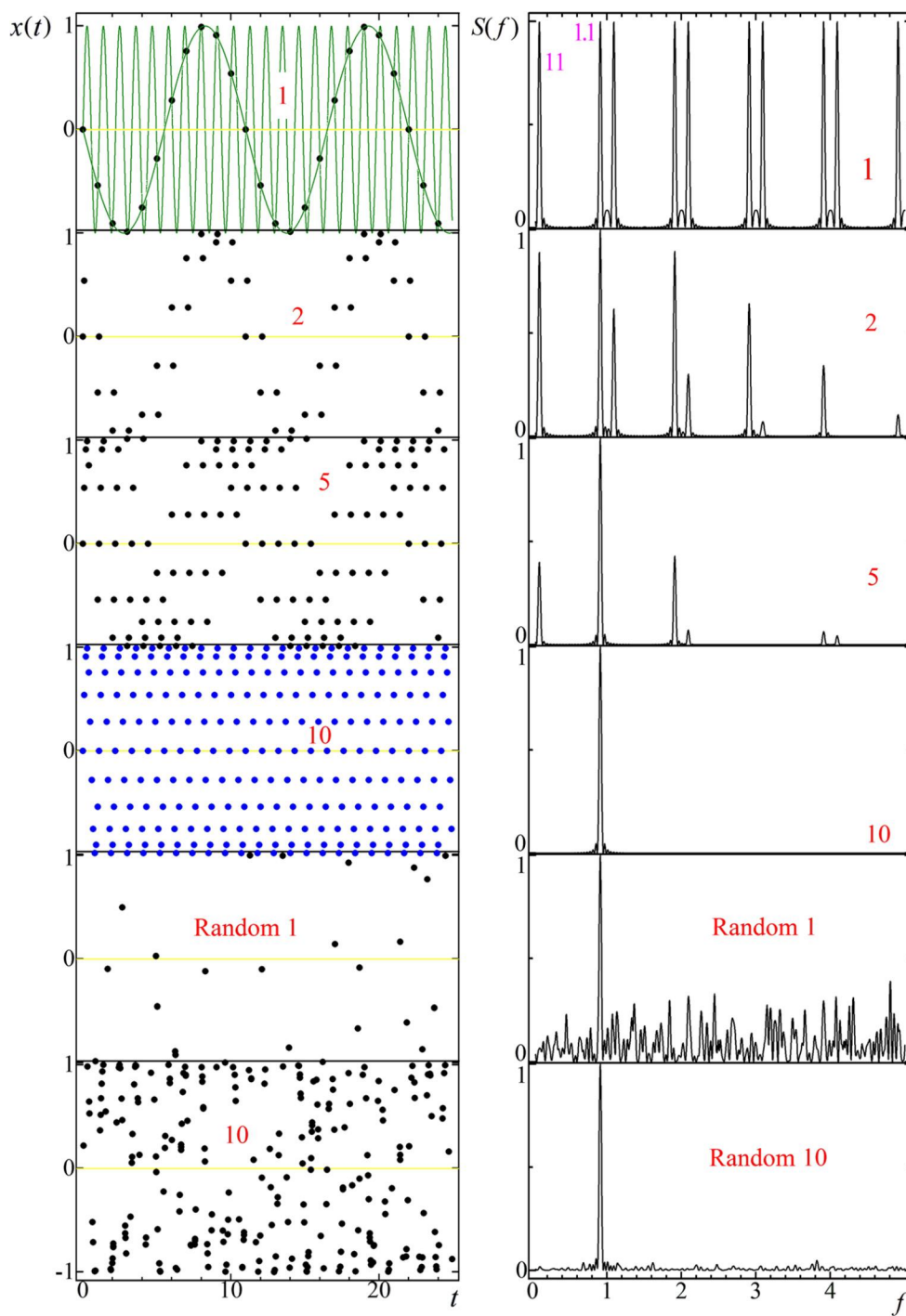


FIG. 11.10 *Left:* The simulated sinusoidal signal with a period $P = 1.1$ with (1, 2, 5, 10) sampling points per unit time. At the top figure, the sampling is 1, for lower figures, the points are added every 0.1. *Right:* The periodograms are computed using the TP1 methods and represented as functions of $f = 1/P$.

11.8 PERIODOGRAM ANALYSIS OF SIGNALS WITH APERIODIC OR PERIODIC TRENDS: WHEN DETRENDING AND PREWHITENING LEAD TO GENERALLY WRONG RESULTS

Often the astro and geosignals contain different contributions with different properties, shape, timescale, etc. These contributions are to be taken into account simultaneously to get statistically optimal results.

The simplest model for a periodic signal with a linear trend is

$$x_c(t) = C_1 + C_2 \cdot \tilde{t} + C_3 \cos(2\pi f \tilde{t}) + C_4 \sin(2\pi f \tilde{t}), \quad (11.73)$$

where $\tilde{t} = t - \bar{t}$. For the fixed value of f , the term \tilde{t} is usually omitted (i.e., set to zero). However, for differential corrections, it is recommended to use a sample mean value of the times of observations. This improves the quality of the matrix of normal equations and thus significantly speeds up convergence of iterations (Andronov, 1994a, 2003). In the software MCV, the user may choose to remove the mean value from times of observations before the analysis.

The periodogram may be defined as $S(f) = 1 - \Phi_4/\Phi_2$, where m in Φ_m shows the number of parameters.

Recently, Olsper et al. (2018) used the same model and called it "generalized Lomb–Scargle periodogram with trend" and applied Bayesian methods for the period estimates.

For example, for a multiharmonic multiperiodic signal superimposed onto a trend approximated with an algebraic polynomial, one may write a model

$$x_c(t) = C_1 + \sum_{\alpha=1}^{s_0} C_{\alpha+1} \tilde{t}^\alpha + \sum_{\beta=1}^{s_P} \sum_{\alpha=1}^{s_\beta} (C_j \cos(2\pi \alpha f_\beta \tilde{t}) + C_{j+1} \sin(2\pi \alpha f_\beta \tilde{t})), \quad (11.74)$$

$$j = 2\alpha + s_0 + 2 \sum_{\gamma=1}^{\beta-1} s_\gamma.$$

Here s_β are degrees of the trigonometrical polynomial corresponding to period $P_\beta = 1/f_\beta$, and s_P is the number of periods.

Such an approximation is used in the software MCV (Andronov and Baklanov, 2004) for up to $s_P = 3$ and

a total number of parameters $m = 21$ in the standard version.

There is a possibility to improve the initial values of the period using differential corrections. This should be done with care, as sometimes the periods taken from a periodogram are biases, so the matrix of normal equations is (nearly) degenerate, and the iterations do not converge to the statistically optimal solution. One may recommend to make corrections to only one period, keeping others fixed.

In many monographs and textbooks, there are recommendations to remove slow trends ("detrending") or remove periodic components ("prewhitening") before further time series analysis. Although we prefer to use complete models using neither detrending nor prewhitening, in the MCV, we have often included this option.

11.9 ANALYSIS OF MULTIPERIODIC, MULTIHARMONIC, AND MULTISHIFT SIGNALS

Similarly to previous expressions, one may write an unsimplified expression taking into account all the following components of variability: multiple periods, nonsinusoidal shape, and possible shifts between the observations of different observers or different runs.

For the periodogram analysis of multiharmonic signals with an algebraic polynomial trend, in the MCV, we propose to use

$$x_c(t) = C_1 + \sum_{\alpha=1}^{s_0} C_{\alpha+1} \tilde{t}^\alpha + \sum_{\alpha=1}^s (C_j \cos(2\pi \alpha f \tilde{t}) + C_{j+1} \sin(2\pi \alpha f \tilde{t})), \quad (11.75)$$

$$j = 2\alpha + s_0.$$

Generally, if there are shifts, a special study is needed to explain their nature. Examples are during multitelescope campaigns, when there are different instrumental systems. Assuming that these differences are not important, one may determine these shifts from this model and then maybe subtract them from the original data to make a joint dataset. In common programs, this option is not available, thus sometimes one may subtract a shift "by eye." In other programs, a sample mean value is subtracted. Before we discussed how such detrending may affect results of the periodogram analysis. Thus we rec-

ommend to use a complete expression, which includes a separate shift C_α for each filter/channel α , i.e.,

$$x_{C_\alpha}(t) = C_\alpha + \sum_{j=1}^s (C_{s_0-1+2j} \cos(2\pi j f \tilde{t}) + C_{s_0+2j} \sin(2\pi j f \tilde{t})) \quad (11.76)$$

11.10 RUNNING APPROXIMATIONS

11.10.1 General Expressions

They are based on the local approximation $x_C(t, t_0, \Delta t)$ in the interval $(t_0 - \Delta t, t_0 + \Delta t)$ and only a central point $t = t_0$ is taken into account, so the smoothing function $\tilde{x}_C(t_0, \Delta t) = x_C(t_0, t_0, \Delta t)$. "Running means" or (a synonym) "moving averages" are most famous local approximations when the points have an equal weight and are approximated in this "running" interval by a sample mean,

$$x_C(t, t_0, \Delta t) = \frac{1}{n(t_0, \Delta t)} \sum_{|t_k - t_0| \leq \Delta t} x_k, \quad (11.77)$$

where $n(t)$ is the number of points inside this interval. This may be easily generalized to a case of different weights of observations w_k , i.e.,

$$x_C(t, t_0, \Delta t) = \frac{1}{n(t_0, \Delta t)} \sum_{|t_k - t_0| \leq \Delta t} w_k \cdot x_k, \quad (11.78)$$

$$n(t_0, \Delta t) = \sum_{|t_k - t_0| \leq \Delta t} w_k. \quad (11.79)$$

Generally, $\tilde{x}_C(t_0, \Delta t)$ is a piecewise constant function, i.e., is a spline of degree 0 and defect 1. Where the times are distributed regularly and all $w_k = 1$, the running approximations are usually defined at the arguments of observations t_k , i.e.,

$$\tilde{x}_C(t_k, \Delta t) = x_C(t_k, t_k, \Delta t) = \sum_i h_i(\Delta t) \cdot x_{k+i}. \quad (11.80)$$

The final value of Δt should be fixed, so the coefficients h_i of the filter are fixed. The summation takes place for all i with $h_i \neq 0$. The statistical accuracy is

$$\sigma[\tilde{x}_C(t_k, \Delta t)] = \sigma \cdot \sqrt{\sum_i h_i^2(\Delta t)}, \quad (11.81)$$

assuming that all $\sigma_k = \sigma$. The filters may be asymmetrical, $h_{-i} = -h_i$, symmetrical, $h_{-i} = h_i$, or "general." There are also the "differentiating filters," e.g., $h_i = (-\frac{1}{2}, 0, +\frac{1}{2})$ for the numerical value of the deriva-

tive, or "integrating" ones, e.g., $h_i = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ for the "running mean," $h_i = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ for the "running mean," $h_i = (\frac{1}{4}, \frac{2}{4}, \frac{1}{4})$ for the central point in the linear approximation and for integration using the trapezium method, and $h_i = (\frac{1}{6}, \frac{4}{6}, \frac{1}{6})$ for the integration using the three-point parabolic approximation for the function. There is a common problem with "border intervals," as there may be missing points needed for this sum. Practically, there may be few approaches. The correct one is to recompute the values of h_i for the asymmetrical case. Simple approaches are to set missing values to "zero" (or the sample mean of the whole series); to set to the nearest "normal" value; or to assume that the series are periodic, so $x_{k+j \cdot n} = x_k$ for all integer j . For a running mean, one may not use missing values and compute a sample mean for a smaller number of points near the border. Also, one may divide the sum by a sum of used coefficients h_i . All these approximations lead to "biased" values close to the borders, so we prefer to use the sets of h_i recomputed for a smaller number of points using the same basic functions and the window function. General principles of design of the digital filters were presented in monographs (e.g., Hamming, 1997).

11.10.2 Running Approximations and Scalegram Analysis for Irregularly Spaced Data

Obviously, if the variations are periodic, the global cosine approximations are preferred. However, many signals are not harmonic; the individual oscillations may vary in amplitude, shape, cycle length, etc. For such "quasiperiodic" oscillations, local approximations at "running" intervals are more effective. The main idea is to compare approximations with observations at different timescales. For small Δt , the approximation will be dominated by observational errors, and will have apparent waves of low statistical significance. For very large Δt , the approximation will be bad because of large systematic difference from the observations.

Andronov (1997) has studied a general case of running approximations with arbitrary test functions using additional weight functions (= windowing functions = filter functions). In this case, Eq. (11.5) should be replaced by

$$(\vec{a} \cdot \vec{b}) = \sum_{k=1}^n p(z_k) \cdot w_k \cdot a_k \cdot b_k. \quad (11.82)$$

Here $p(z)$ is the weight function dependent on the (dimensionless) parameter $z = (t - t_0)/\Delta t$. The localized function $p(z)$ should be zero outside the interval of smoothing (i.e., for $|z| > 1$). For smoothing, one

may recommend symmetrical functions $p(-z) = p(z)$; otherwise the “forward” and “backward” approximations will not coincide. The simplest weight function is the “rectangular” one, $p(z) = 1$. The disadvantage of this method is that the smoothing function does not smoothly vary with time t_0 at the mid-interval, and has discontinuity when a single point enters/leaves the interval of smoothing. Thus one may recommend to use additional restrictions for derivative $p'(\pm 1) = 0$ and $p(z) \leq 0$. Obviously, there may be an infinite number of such functions, with among them the Hann function $p(z) = \cos^2(\pi z/2) = (1 + \cos(\pi z))/2$ (Hamming, 1997). Andronov (1997) proposes the much simpler function (and thus faster for computations) $p(z) = (1 - z^2)^2$. It has an additional advantage: as the derivative $p'(\pm 1) = 0$, the derivative of the smoothing function $\tilde{x}_C(t_k, \Delta t)$ is continuous everywhere. A global approximations using the cubic splines (Andronov, 1987) has a continuous second derivative. Because the real time series are always limited, and often have irregular gaps, there is no real advantage of using infinite functions like a Gaussian $p(z) = \exp(-c \cdot z^2)$, an exponent $p(z) = \exp(-c \cdot |z|)$, etc. The next user's choice is to define a set of the basic functions $f_\alpha(z)$. Following the tradition of using algebraic polynomials, the parabola is the next case after the constant. Obviously, there may be lines in between. But they may be skipped in a “good” case of equidistant observations, as they produce the same values $\tilde{x}_C(t_k, \Delta t)$ at the moments t_k . In other words, the polynomials of orders (0,1), (2,3), etc., are different for the same data, but the central point is the same. Thus the “running parabola” will exactly fit the cubic polynomial at times t_k . Moreover, the approximation has much better properties as compared with either the “running weighted mean” with this $p(z) = (1 - z^2)^2$, or the “unweighted” ($p(z) = 1$) parabola. The only parameter which remained to be determined, is the filter half-width Δt . For this purpose, one has to compute “scalegrams,” i.e., the dependence of the test functions on Δt . Andronov (1987) used the following characteristics: σ_{O-C} , the unbiased estimate of the r.m.s. deviation of the observations from the fit; σ_{x_C} , r.m.s. accuracy of the smoothing function at the moments of observations; and the SNR. These dependencies are shown in Fig. 11.11. For an illustration, we have used a relatively short part of the light curve of R Aqr in the filter V from the AAVSO database. It was used above for the polynomial approximation in Fig. 11.3.

Briefly, the dependence σ_{O-C} should have two “standstills.” At small Δt , which corresponds to statistical noise (for this sample, the minimum corresponds to 0.070^m), the function has a transition to the “standstill”

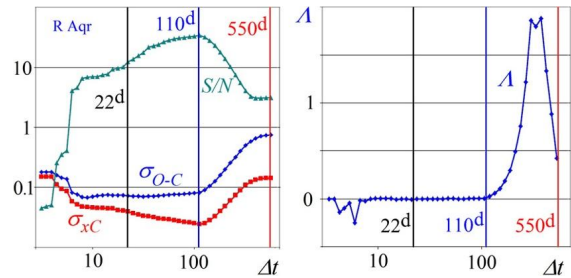


FIG. 11.11 Scalegrams σ_{O-C} , σ_{x_C} , and SNR as functions of Δt (left) and the Λ scalegram (right) for the part of the light curve of R Aqr (same as for the polynomial approximation at Fig. 11.12). The vertical lines correspond to the optimal value $\Delta t = 110^d$ and five times larger and smaller for comparison.

at large Δt due to increasing influence of the systematic deviations of the approximation from the observations.

From the position of the transition, it is possible to estimate the “period,” and the levels of the standstill allow to estimate the effective semi-amplitude of the variations. Andronov and Chinarova (2003) introduced these new effective characteristics of quasiperiodic signals, namely, the effective amplitudes, periods (timescales), and slopes of the scalegram. They have been determined for 173 semiregular variables. Five stars are characterized by outstanding values of at least one of the parameters and were chosen for additional observations to check their peculiar behavior.

The dependence of σ_{x_C} is more complicated, and has a single minimum of 0.024^m at $\Delta t = 110^d$. This is one of the criteria to choose Δt for a final approximation. The “amplitude” SNR is 34. For very noisy observations, the maximum SNR is typically shifted towards smaller Δt as compared to a minimum of σ_{x_C} . Andronov (2003) introduced the “ Λ scalegram,” which more resembles different periodograms, i.e., showing “peaks” instead of “transitions.” From the position and height of the peak, it is possible to determine the effective period P_Λ and semi-amplitude R_Λ . For the data on R Aqr, $P_\Lambda = 345^d$. The period estimate is smaller than $360^d \pm 2^d$ from the cosine fit, with the initial epoch for the maximum brightness (minimum magnitude) $T_0 = 2456769.8 \pm 1.1^d$ and semi-amplitude $R = 1.74^m \pm 0.05^m$. Andronov (1997) has shown that the optimal $\Delta t \approx 0.511P$ for a pure sinusoid with many (nearly regular) observations during a period. A smaller value of P_Λ is explained by the presence of harmonics of the main oscillation. In cases where the variability is present at many timescales with variable cycle lengths, the “transitions” at σ_{O-C} occur at different Δt . Using the σ scalegram, Andronov et al. (1997) discovered a

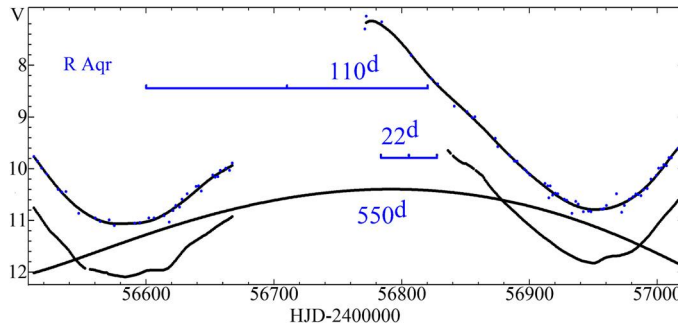


FIG. 11.12 Part of the light curve of R Aqr (same as for the polynomial approximation in Fig. 11.3). The approximations $x_c(t)$ using “running parabola” fit with $\Delta t = 110^d$ (optimal) and five times larger and smaller (the curves are shifted by 1^m). Horizontal bars and numbers show the values of the filter half-width Δt .

fractal type of variability of the magnetic cataclysmic variable AM Her at a very wide range of timescales from 3 seconds to 30 years. The unbiased scatter estimate increases with the filter half-width according to a power law $\sigma_{O-C} \propto (\Delta t)^{0.180}$ from 10^{-4} to 3000 days. The estimate of the fractal dimension is $D = 0.32$.

11.10.3 Running Sines

For nearly sinusoidal signals, it is expected to use a “running approximation”

$$\begin{aligned} \tilde{x}_C(t, \Delta t) = & C_1 + C_2 \cos(\omega \cdot (t - T_0)) \\ & + C_3 \sin(\omega \cdot (t - T_0)), \end{aligned} \quad (11.83)$$

$$\tilde{x}_C(t, \Delta t) = C_1 - R \cos(\omega \cdot (t - T_M)), \quad (11.84)$$

where $\omega = 2\pi/P$, P is the period, and T_0 is the initial epoch. The coefficients C_α are functions of t_0 and Δt , as only the data in the interval $[t_0 - \Delta t, t_0 + \Delta t]$ are used. The following parameters are determined: C_1 , the mean (over the period) value; R , the semiamplitude; T_M , the moment of maximum brightness (minimum stellar magnitude); and brightness at maximum $x_{max} = C_1 - R$ and minimum $x_{min} = C_1 + R$. Obviously, T_M may be easily converted to a phase, using T_0 and P . The recommended value $\Delta t = P/2$, but, for large gaps in the observations, it may be enlarged, e.g., to $\Delta t = P$. This method is effective for studies of many types of stars – semiregular, symbiotic, Mira-type, intermediate polars, RR Lyrae-type with the Blazhko effect. This approximation is close to that used in physics for studies of amplitude and phase modulations. Another example may be temperature changes during a day with a yearly wave. A review of the method was presented by Andronov and Chinarova (2013). An extension of this method may be a shift (in C_1 , T_M) and scaling (in R) of nonsinusoidal shapes (“patterns,” “templates”).

11.10.4 Wavelet Analysis

Wavelet analysis is another extension of the Fourier analysis and of running approximations. It is also related to dynamic spectra, when the Fourier analysis is carried out using the window function. In astronomy, the “seasonal” periodograms are often used, as there are natural borders of the series of more or less dense observations. A similar situation is for time series obtained during a night with gaps between them. In this case, there is no sense for a constant Δt , and the complete data are binned to (unequal in length) subintervals. If the signal has no gaps, these intervals may be chosen also as subsequent nonoverlapping ones. In this case of the nonoverlapping subintervals, the values at the periodogram (and other parameters) are statistically independent.

For overlapping intervals, the “morphing” of the parameters is more smooth. Similarly to the running approximations, one may use not only a rectangular window, but also smooth ones.

If Δt is not dependent on the trial period, the terminology is the “dynamic spectra.” For the wavelet analysis, it is assumed that $\Delta t = \tilde{c} \cdot P$, where \tilde{c} is some constant. For $\tilde{c} \rightarrow \infty$, the approximation becomes “global” rather than “local.”

Andronov (1998, 1999) reviewed the statistical properties of the test functions of the weighted LS improvement of the Morlet-type wavelet. Beyond the typical “wavelet maps,” the additional features are introduced, e.g., the (weighted mean) wavelet periodogram, the wavelet skeleton, and the wavelet approximation.

For irregularly spaced signals, the “noise” at the wavelet periodogram may be decreased by a few dozen percent, and, in some extreme cases, even by a factor of a few times.

For the wavelet periodogram, one may use four different functions, which have different asymptotic behavior at large and small trial periods. They were compared to the scalegrams (e.g., Andronov and Chinarova, 2001).

The analysis of citation of these papers at the site [Researchgate.net](https://www.researchgate.net) shows numerous applications of the method not only for astro- and geodata, but even in medicine (cardiograms).

An example of dependence of the wavelet periodogram on the effective width of the window function and a comparison to the “running sine” analysis was presented by Chinarova (2010). The drastic variations of the semi-amplitude of pulsations of the semiregular variable RU And from nearly constant (0.027^m) to 1.204^m (Mira-type pulsating variable) were detected.

Andronov and Kulynska (2020) discussed a modification of the Morlet-type wavelet by using a compact weight function $p(z) = (1 - z^2)^2$ (initially proposed by Andronov, 1997 and discussed by Andronov, 1999), and have tested statistical properties of the parameters using the traditional matrix form, as well as the “bootstrap”-generated data.

Mann and Haykin (1991) introduced a generalization of Gabor’s Logon Transform, which they called “the chirplet transform” it is pointed to analysis of signals with frequency variations.

11.11 MOMENTS OF CHARACTERISTIC POINTS ($O - C$ ANALYSIS)

11.11.1 Period Determination

One of the methods of analysis is the determination of the moments of the characteristic points t_k only (cf. Tsevech, 1973; Kreiner et al., 2001).

They may be subdivided into two main types: the maxima or minima (extrema) and the moments of crossings by the approximation of some constant level (e.g., the γ velocity [mean] by the curve of the radial velocity). The main expressions were discussed in Section 11.2. AAVSO prefers to use the abbreviation ToM (times of minimum), but it may be extended to times of maximum and also to times of any other characteristic point.

The simplest mathematical model for $O - C$ analysis is

$$t_k = T_0 + P \cdot E_k + \epsilon_k, \quad (11.85)$$

where T_0 is called “the initial epoch,” P is the “period,” E_k is the “cycle number,” and ϵ_k is the residual, which is often called “ $O - C$.”

In the simplest case, $P = t_{k+1} - t_k$ is the ToM corresponding to subsequent cycles, e.g., that shown in Fig. 11.6. For better precision, $P = (t_n - t_1)/(n - 1)$.

If the (integer) cycle numbers E_k are correct, then one may determine corrections ΔT_0 and ΔP using the LS, and then new statistically optimal values $T_0 + \Delta T_0$ and $P + \Delta P$.

If the moments t_k are rare enough, there may be problems with determination of the cycle numbers E_k . In this case, one may apply the greatest common divisor (GCD) algorithm, similar to that proposed by Euclid for integer numbers: $\text{GCD}(a, b) = \text{GCD}(a, b \bmod a)$, swapping the numbers, if needed, for $a < b$, and repeating until $\text{GCD}(a, 0) = a$. For noninteger numbers, this may be replaced by computing the pairs of differences $\Delta_k = t_{k+1} - t_k$, and then sorting in increasing order and trying to find a (generally noninteger) GCD of all these differences, which will be equal to the period (Tsevech, 1973).

A computer algorithm to carry out periodogram analysis was presented by Dumont et al. (1978). Andronov (1988) studies the properties of the test function.

The periodogram analysis for all data points was proposed by Andronov (1991). The initial epoch is set to $T_0 = t_1$. The test periods were chosen as $P_E = (t_n - t_1)/E$, where trial values of E started from $n - 1$ to some reasonably small values, e.g., $P_E = 0.1^d$. Then T_0 and P_E are corrected using the LS, and the phases are recomputed, so the test function is a (weighted) sum of squares of the residuals.

For fast computations, one even may not make the corrections – in this case, the routine may be realized in the electronic tables even without programming.

This method is effective for stable periodic variations and small statistical errors of t_k . However, it may produce a “saw tooth” periodogram in the case of large phase deviations. This is a typical situation for cataclysmic variables, where the intervals between the outbursts vary by few dozen percent (Andronov and Shakun, 1990), or the minima have large shifts (as, e.g., in the cataclysmic variable TT Ari, Kim et al., 2009).

Jetsu and Pelt (1996) had taken into account accuracies σ_k of the individual time points t_k under a common assumption of uncorrelated statistical errors. Then the accuracy of the time differences is

$$\sigma_{kj} = \sigma[t_k - t_j] = \sigma_k^2 + \sigma_j^2. \quad (11.86)$$

11.11.2 Period Changes

To study changes of the period, the ToM are to be determined during a long-time monitoring. The main

mechanisms for the period variations in eclipsing binary stars are the mass transfer (accretion and excretion due to stellar wind); the magnetic stellar wind and gravitational radiation; and the changes in the internal structure in a component. There may be changes of the $O - C$, which are related not to the period variations, but to other mechanisms like the light-time effect due to presence of a third body (a star or a planet) or multiple objects. The ToM may be shifted due to the presence of stellar spots. For binary systems at eccentric orbits, there may be an apsidal motion (cf. Tsevesich, 1973; Kopal, 1959). The discussion on the statistics of the minima compiled in “An Atlas of $O - C$ Diagrams of Eclipsing Binary Stars” (Kreiner et al., 2001) was presented by Kim et al. (2003). A catalogue of 623 systems with eccentric orbits and their classification was presented by Kim et al. (2018). Some system show complicated variations due to few acting mechanisms.

There are few online catalogues of ToM, e.g.,

- <http://var2.astro.cz/ocgate/>,
- <http://www.as.ap.krakow.pl/o-c>,
- <http://www.bav-astro.eu/index.php/veroeffentlichungen/service-for-scientists/lkdb-engl>,
- <http://www.aavso.org/bob-nelsons-o-c-files>,
- <http://www.aavso.org/observed-minima-timings-eclipsing-binaries>

The compilations of ToM are often published in journals, e.g., in the “Open European Journal on Variable Stars” (cf. Paschke, 2018). The relation between period and ToM of variable stars was discussed by Kopal and Kurth (1957). Introducing the functions $t(E)$ ($M(E)$ in their notation) and $P(E)$ as the time and period corresponding to the cycle number E , one may distinguish the “discrete” $P(E) = t(E + 1) - t(E)$ and “continuous” $P(E) = dt(E)/dE$ definitions. They discussed some models. In the “continuous” model,

$$t(E) = T_0 + \int_0^E P(E) dE. \quad (11.87)$$

Assuming the simplest model for the period variations,

$$t(E) = T_0 + P_0 \cdot E + Q \cdot E^2, \quad (11.88)$$

we get $P(E) = P_0 + 2QE$ and $dP/dE = 2Q$, $\dot{P} = dP/dt = (dP/dE)/(dt/dE) = 2Q/(P_0 + 2QE)$. The characteristic timescale of the period variations is defined usually as $\tau = P/|\dot{P}|$, which is equal for this case to $\tau = P^2/(2|Q|)$, and is thus dependent on time. For $\dot{P} = \text{const}$, the period is $P(E) = P_0 \cdot \exp(\dot{P} \cdot E)$, and

$$t(E) = T_0 + P_0 \cdot (\exp(\dot{P} \cdot E) - 1), \quad (11.89)$$

where the index 0 corresponds to the value at $E = 0$ (cf. Andronov and Chinarova, 2013).

Having a model $t(E)$, one may define $E(t)$ as an inverse function. It is suitable for compute phases (which are decimal parts of $E(t)$) for plotting phase curves. For parabolic $O - C$, the option to compute phases is included in the software MCV (Andronov and Baklanov, 2004) and MAVKA (Andrych and Andronov, 2019). However, the phases may be computed also for more complicated models (e.g., Kim et al., 2005).

11.12 AUTOCORRELATION AND CROSS-CORRELATION ANALYSIS

11.12.1 Continuous and Discrete Regular Signals

Autocorrelation analysis is one of classical methods for data analysis, which was described in hundreds of textbooks (e.g., Box et al., 2015, Blackman and Tukey, 1959). Astronomical applications were reviewed by Deeming (1970).

Classical ACF is defined as $r = R_u/R_0$, where

$$R_u = \frac{1}{N_u} \sum_{k=1}^{n-u} (x_k - \bar{x})(x_{k+u} - \bar{x}). \quad (11.90)$$

This is applicable to evenly distributed time series $t_k = t_1 + (k - 1)\delta$. Obviously, $r_0 = 1$ for any variable signal. For the simplest case of pure noise with a variance σ_0^2 , the mathematical expectation is 0 for $u \neq 0$. Sample ACFs differ from this value; the statistical properties were studied by Andronov (1994b).

Here we introduced a new variable N_u . For $N_u = N$, the ACF is called the “biased” one. However, the number of summands in the sum is equal to $n - u$, thus the “unbiased” mathematical expectation is for $N_u = n - u$.

Correlation length is defined as the smallest positive root of the equation $r_u = 0$. It may be expressed in units of time, $\tau_0 = \delta \cdot u$. For a pure sinusoidal signal with a period P , $r_u = \cos(2\pi\delta u/P)$ and $\tau_0 = P/4$. The first nonzero maximum corresponds to $u_{max} = P/\delta$. For quasiperiodic variations, it gives an estimate of some characteristic cycle length.

For the shot noise or autoregressive (AR) process of the first degree, $r_u = \psi^u$ and thus it does not cross zero if the number of observations n is large. However, the removal of the sample mean value \bar{x} from the data leads to a distinct crossing of zero by the ACF (Sutherland et al., 1978), thus leading to wrong interpretation of the character of variability, mainly, the shot noise (AR1), or QPO (AR2).

The bias of the ACF increases with decreasing length of data and more complicated shape of the trend. A correct set of equations for arbitrary length and the number of basic functions were presented by Andronov (1994b).

For the shot noise, the common mathematical model is the Markov process, or the AR-1 model. The theoretical ACF is $r_u = \psi^u = \exp(-u/u_d)$, so the decay time may be determined as $\tau_d = u_d \delta = -\delta = \ln \psi$. So, the ACF decreases by a factor of e for the decay time τ_d . For multicomponent signals, assuming their statistical independence,

$$x_k = \sum_{\beta=1}^m x_{\beta k}, \quad (11.91)$$

$$R_u = \sum_{\beta=1}^m R_{\beta u}, \quad (11.92)$$

$$r_u = \frac{n}{N_u} \sum_{\beta=1}^m \tilde{R}_{\beta 0} \cdot r_{\beta u}, \quad (11.93)$$

where the relative contribution to the ACF $\tilde{R}_{\beta 0} = R_{\beta 0}/R_0$. Using this algorithm, Andronov et al. (2005) proposed a four-component model of the ACF of X-Ray variability of AM Her based on a CHANDRA Observation, and the second component in the shot noise variability of AM Her was discovered.

11.12.2 Bias of ACF due to Trend

For astro- and geosignals, the problem arises with taking into account a sample mean (Sutherland et al., 1978) or a possible trend. Complete study of the influence of trend removal to the resulting ACF for any basic functions was presented by Andronov (1994b). In Fig. 11.13, the ACFs for different lengths are shown for the initial data without trend removal, removal of the sample mean, and removal of the cubic approximation. It is clearly seen that the trend removal significantly affects the ACF and may lead to wrong physical conclusions. For example, numerous studies of AM Her have shown a distinct crossing of zero by the ACF, which is typical for quasiperiodic oscillations, or the AR2 model, rather than for the AR1 model, or the shot noise.

11.12.3 Irregularly Spaced Signals

For irregularly spaced signals, the ACF may be computed in two ways.

(1) The data are interpolated, and then the ACF is computed for such an artificial dataset. This is typically

used for single missing points, e.g., deleted because of an outlying value, or an absence of measurement,

If the point x_k is missing, it may be replaced by values using a local approximation – linear or cubic: $\tilde{x}_k = (x_{k-1} + x_{k+1})/2$ and $\tilde{x}_k = (4 \cdot (x_{k-2} + x_{k+2}) - (x_{k-1} + x_{k+1}))/6$. If the accuracy σ is the same for these points, the accuracy of the interpolated value is $\sigma/\sqrt{2}$ and $\sigma\sqrt{17/18}$, respectively. Such an option is included in the software MCV (Andronov and Baklanov, 2004).

Although the first value is more accurate for uncorrelated noise, one may recommend the second one in the case of suggested fast variations of the signal at a timescale of a few δ , where δ is time resolution. Otherwise the maxima and minima should be flattened.

(2) The second approach is to split the data into nonoverlapping intervals and to compute mean values. Then the new time series are equidistant in time, and the analysis may be carried out.

For example, in the series of papers based on the observations from the AAVSO database, there is an approach to make 10^d means, i.e., to split the interval into 10^d pieces and treat them as one point. The weak points of such method may be in a case of significant variations during these intervals, and the point(s) may correspond to an “effective time,” which differs from the mid-interval.

The third method for the general case of distribution of times t_k is not to make time bins for the observations, but to determine individual time differences and round them off to an integer number of shifts. This corresponds to a “rectangular” time window. One may also apply a structure function like in the method for the periodogram analysis.

Another approach is to compute not the sums of deviations from the fit, but the sum of squares of differences between the pairs. So the minimum of Φ is to be found, rather than the maximum of the correlation coefficient r .

The cross-correlation function may be defined in a similar way – typically by using regular time series, or making such series by linear/cubic interpolation or computing a mean in the small intervals.

Contrary to the ACF, the CCF is generally not symmetric, and it may be used for estimates for the time lags between two signals. An example may be the time lag between the daily temperature and the solar heating at some place, or the light echo due to reemission of the light from a Supernova explosion by distant clouds in the vicinities.

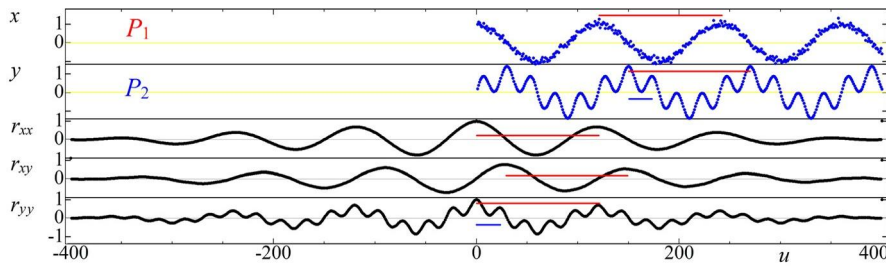


FIG. 11.13 The model signals x_k , y_k : cosine (one period P_1 with some noise) and sine functions (two periods P_1 and P_2) and the corresponding ACF, r_{xx} , r_{yy} , and the CCF r_{xy} . The long and short horizontal lines show periods P_1 and P_2 , respectively.

In Fig. 11.13, the CCF for the model harmonic oscillations with some noise is shown.

The vertical lines correspond to maxima. Definitely, for periodic functions, the peaks at the CCF repeat periodically with the same period. However, due to the different number of the summands $n - u$, the highest peak is expected to be closer to zero.

11.13 PRINCIPAL COMPONENT ANALYSIS AND RELATED METHODS

Singular value decomposition (SVD) and principal component analysis (PCA) are described in many textbooks (e.g., Forsythe et al., 1977; Press et al., 2007). Its applications to astronomical time series were reviewed, e.g., by Andronov (2003); Andronov et al. (2003b); Mikulášek (2007b).

The classical PCA is based on the analysis of the covariation matrix for the data obtained in different channels. It is very closely related to the SVD for the matrix $\epsilon_{\alpha k}$ (typically only a mean value is removed, but generally one may propose to subtract a smoothing value):

$$\epsilon_{\alpha k} = \frac{x_{\alpha k} - x_{C\alpha k}}{\sigma_{C\alpha k}}. \quad (11.94)$$

Typically, in simpler models, $x_{C\alpha k} = \bar{x}_\alpha$, just using a sample mean for each channel α .

The main idea is to apply filtering to the time series by incomplete restoration of the sum, decreasing the number of members of the sum from m to L .

11.13.1 Principal Component Analysis: Multichannel Signals

PCA is used for multichannel signals, when, for the same moment of time t_k , there are multiple values of signal $x_{\alpha k}$.

The main idea of the analysis using SVD or PCA is that the accuracy is equal for all measurements in all channels and all data. For data with very different scatter in different channels, there may be some scaling, e.g., by normalizing the data as $x_{\alpha k} = (x_{\alpha k} - x_{0\alpha k})/\sigma_\alpha$. In the two-dimensional case, this leads to a bisector independently on the value of the correlation coefficient.

The slope of the line is generally different either from the regression line, or the line of the orthogonal regression. For the three-dimensional, the situation becomes less undetermined. One may make a partial restoration of the function and use the values $\sigma_\alpha[O - C]$ for scaling. In this case, some iterations are needed to get a self-consistent solution.

In Fig. 11.14, the part of the light curve of SS Cyg, the prototype dwarf nova, is shown. There are large seasonal gaps, when the star is not visible during the night, but the measurements are nearly simultaneous. Four filters (channels) are used. To prepare the data for the PCA, the observations are to be converted to “pseudo-simultaneous” ones. It may be done by neglecting the time difference between the observations in different channels. Or, alternately, one may use interpolation of the signal from one channel to times of another one, using local cubic polynomial. This is implemented in the software MCV (Andronov and Baklanov, 2004).

In Fig. 11.15, the principal components of variability of the signal shown in Fig. 11.14, are presented. One may note that the components 1 and 2 show distinct variability, despite very different amplitudes. The components 3 and 4 have very low amplitudes, and may be interpreted as an observational noise.

In Fig. 11.16, we compare color indices computed using the PCA, and from the approximations using the “running parabola” approximation with a statistically optimal value of the filter half-width $\Delta t = 0.007^d$, which corresponds to $\text{SNR} = 78$.

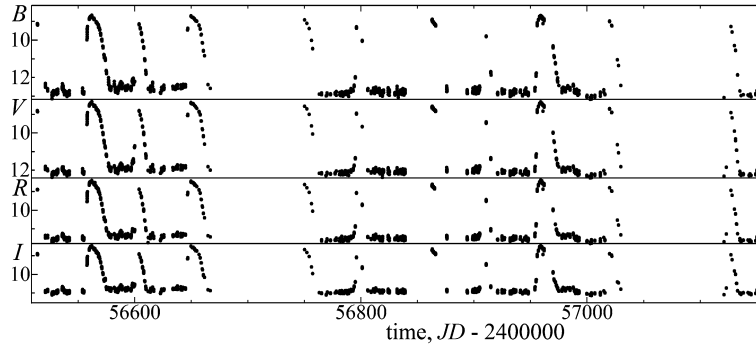


FIG. 11.14 Multicolor BVRI observations of the cataclysmic variable SS Cyg obtained by Robert James (AAVSO code JM).

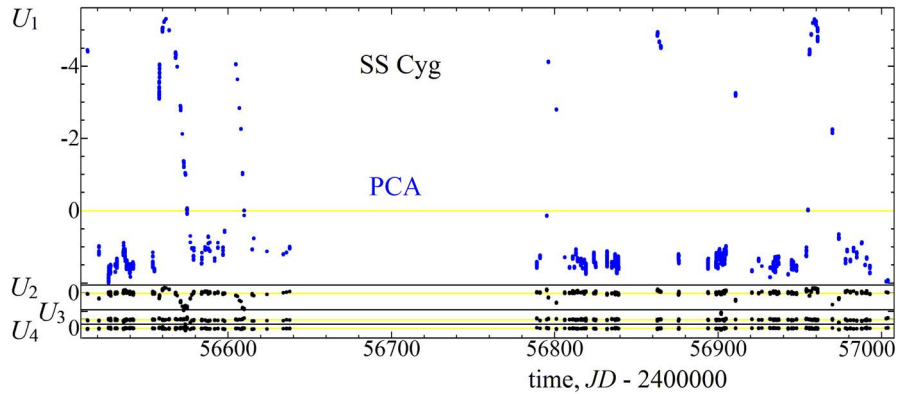


FIG. 11.15 Principal components $U_{\alpha k}$ as a dependence on t_k for the observations of the cataclysmic variable SS Cyg (Fig. 11.14) smoothed using the running parabolae with $\Delta t = 0.07^d$. The right part of the graph was automatically omitted, as there were only single observations in each filter during these nights.

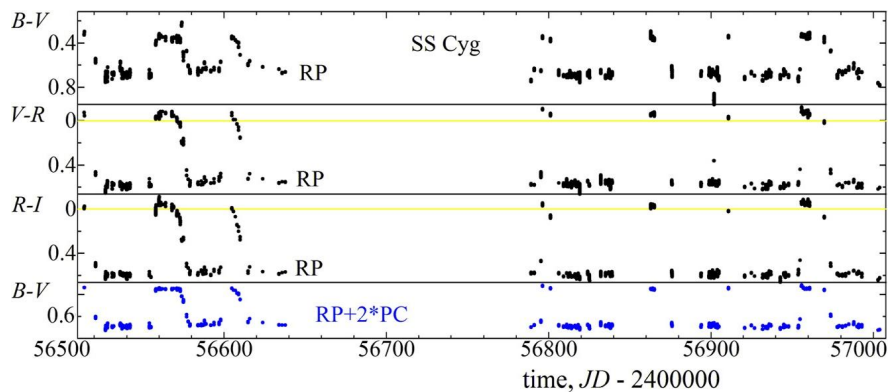


FIG. 11.16 Color indices based on the data smoothed using the running parabolae (up). The bottom graph is the first color index, additionally smoothed using the PCA with the number of principal components $L = 2$. The amplitude of the smoothed curve has decreased by a factor of ≈ 2 , removing two outliers.

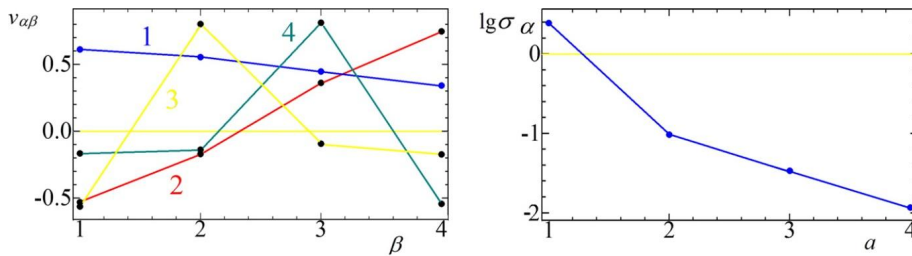


FIG. 11.17 *Left:* The eigenvectors for SS Cyg of the correlation matrix. *Right:* Dependence of the corresponding singular values σ_α on the number of principal components.

In Fig. 11.17, the eigenvectors $v_{\alpha\beta}$ obtained using the PCA, are shown. There is a clear near-linear dependence for the first two components. As the channels are BVRI, the effective wavelength of the signal increases with the number of the component. At the right part of this figure, there is a dependence of the singular value on the number of channel. It shows that the components 3 and 4 have low amplitudes, and may be neglected.

11.13.2 Case of Noisy Channels of Simultaneous Signals

This particular case is applied for measurements of the same signal, which is present in all channels, but the noise is different, as well as the shift, e.g., the changes of the apparent stellar magnitude of a sample of constant stars due to variable extinction and air mass.

Some kind of this approach was introduced by Andronov and Baklanov (2004) in the method of the artificial comparison star, which was implemented in the software MCV. Applications and a description of this method were presented by Kim et al. (2004).

The method of artificial comparison star allows to improve accuracy of the comparison star typically by a few dozen percent. Contrary to the “ensemble photometry” method by Honeycutt (1992), in our method, the weights for the weighted mean are computed after the iterations, and do not depend on the brightness in an approximation.

11.13.3 Singular Spectrum Analysis (SSA)

This method was implemented in the program called Caterpillar (Golyandina et al., 2001), in which L (the number of the principal components for restoration) is defined by the user, so there are no automatic recommendations on what is to be taken into account and what should be neglected as the noise.

The uni-channel (or “one-channel”) signal is used to create a pseudomultichannel dataset with m channels.

Then the principal components are computed, and the user can choose which of the components will be used for partial restoration of the signal. This is some kind of smoothing. The coefficients may also be used for a short-term forecast. It should be noted that the matrix used for PCA has the components from the autocovariation function.

11.13.4 Effective Amplitudes of Low, Fast, and Noise Variability

A related method was proposed by Tremko et al. (1996). The signal is split into three parts – smooth variations with, e.g., orbital phase (periodic or aperiodic dependent on the method of smoothing), faster correlated variability (e.g., quasiperiodic oscillations [QPOs] or flickering), and an uncorrelated noise.

For the residuals of the observations, the covariation matrix may be computed, the mathematical expectation of which is

$$\mu_{\alpha\beta} = \sigma_{c\alpha} \cdot \sigma_{c\beta} + \sigma_{n\alpha}^2 \cdot \delta_{\alpha\beta}, \quad (11.95)$$

where the Kronecker symbol $\delta_{\alpha\beta} = 1$ if $\alpha = \beta$; otherwise it is 0. Here $\sigma_{c\alpha}$ and $\sigma_{n\alpha}$ are standard deviations for correlated contribution and uncorrelated noise, respectively. There are $2n$ independent values of these parameters and $n \cdot (n + 1)/2$ independent values of the matrix, taking into account its symmetry, $\mu_{\beta\alpha} = \mu_{\alpha\beta}$. For $m = 3$, both numbers are equal to 6, thus one may easily estimate all parameters.

For larger m , the parameters may be determined using nonlinear LS (Tremko et al., 1996). This method allows to estimate variances of the correlated signal, so also to estimate the r.m.s. amplitude of the signal $\sigma_{c\alpha}$, as well as of the correlated noise $\sigma_{n\alpha}$.

Such method is an effective tool for cataclysmic variables, where there are aperiodic events like QPO or flickering, and $\sigma_{c\alpha}$ are proportional to the amplitude of the signal in the channel α of any shape. Tremko et al. (1996) used these values to estimate color indices of

different contributions to variability, and thus the temperatures.

11.14 CONCLUSIONS

- To obtain new statistically significant astrophysical results, one needs adequate mathematical models.
- The solution is needed for signals with multiple components, which are of finite length, and the observations generally are irregularly spaced.
- The variety of types of variability needs elaboration of special methods, specific for such signals.
- The careful estimate of the FAP prevents publications of many “beautiful discoveries.”
- The multicomponent character of signals needs complete models instead of “step-by-step” simplified ones.
- Solution can be achieved by development of an expert system of advanced complementary algorithms and programs, which improve existing methods.
- Depending on the stability of the periodic variations, the following methods are recommended: the global fits (for the best stability), local fits of an intermediate filter half-width Δt and periodic shapes (patterns) for “nearly periodic” oscillations (“running sines,” wavelet), and local fits of smaller Δt , which are effective for nonharmonic and chaotic-like variations.
- The observations at different longitudes are very important, as they effectively dump side-bands of the spectral window, i.e., decreasing the peaks with “object-observer” periodic beats.

ACKNOWLEDGMENTS

The data from the AAVSO (<http://aavso.org>) and AFOEV were used for illustration. “The SAO/NASA Astrophysics Data System” (<http://adsabs.harvard.edu>) was used. We acknowledge fruitful recommendations by Petr Škoda, Irina B. Vavilova, Yaroslav S. Yatskiv, Lidia L. Chinarova, Gottfried Schwarz, and an anonymous referee. The monitoring of the first star of our sample – AM Her – was initiated by Prof. V. P. Tsessevich (1907–1983) (Andronov, 2017). Since then, more than 387 papers (cited in the ADS) have been published, with a total number of studied variable stars exceeding 2000. Special thanks go to coauthors of 19+ joint papers: L. L. Chinarova, S. V. Kolesnikov, N. M. Shakhovskoy (1931–2011), L. S. Kudashkina, V. I. Marsakova, V. V. Breus, Yonggi Kim, P. Mason, and S. Zoła.

REFERENCES

- AAVSO. <http://aavso.org>.
- Anderson, T.W., 1958. An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, New York, p. 374.
- Anderson, T.W., 2003. An Introduction to Multivariate Statistical Analysis, 3rd Ed. Wiley, 739 p. ISBN: 0471360910, 9780471360919.
- Andronov, I.L., 1987. Smoothing the “smoothing” cubic spline functions. Publications of the Astronomical Institute of the Czechoslovak Academy of Sciences 70, 161. 1987PAICz..70..161A.
- Andronov, I.L., 1988. Period search using the DMRT method – the properties of the test function. *Astronomische Nachrichten* 309, 121.
- Andronov, I.L., 1991. Determination of the period of a variable signal from the times of characteristic events. *Kinematika I Fizika Nebesnyh Tel* 7 (2), 78.
- Andronov, I.L., 1994a. (Multi-)frequency variations of stars. Some methods and results. *Odessa Astronomical Publications* 7, 49.
- Andronov, I.L., 1994b. Autocorrelation function bias owed to a limited number of de-trended observations. Applications to autoregressive models with noise. *Astronomische Nachrichten* 1994 (315), 353–370.
- Andronov, I.L., 1997. Method of running parabolae: spectral and statistical properties of the smoothing function. *Astronomy & Astrophysics. Supplement Series* 125, 207–217.
- Andronov, I.L., 1998. Wavelet analysis of time series by the method of least squares with supplementary weights. *Kinematika I Fizika Nebesnyh Tel* 14, 490.
- Andronov, I.L., 1999. Wavelet analysis of the irregularly spaced time series. In: Priezzhev, V.B., Spiridonov, V.P. (Eds.), *Self-Similar Systems*. Dubna, Russia, 29 July–August, 1998. *Joint Inst. Nucl. Res. Ser. 57; ADS: 1999sss.conf...57A*.
- Andronov, I.L., 2003. Multiperiodic versus noise variations: mathematical methods. In: *ASP Conference Series*, vol. 292, p. 391.
- Andronov, I.L., 2005. Advanced methods for determination of arguments of characteristic events. In: *ASP Conference Series*, vol. 335, p. 37.
- Andronov, I.L., 2012. Phenomenological modeling of the light curves of algol-type eclipsing binary stars. *Astrophysics* 55, 536.
- Andronov, I.L., 2017. Odessa Scientific School of researchers of variable stars: from V. P. Tsessevich (1907–1983) to our days. *Odessa Astronomical Publications* 30, 252.
- Andronov, I.L., Andrych, K.D., 2014. Determination of size of the emitting region in eclipsing cataclysmic variable stars. *Odessa Astronomical Publications* 27, 38.
- Andronov, I.L., Baklanov, A.V., 2004. Algorithm of the artificial comparison star for the CCD photometry. *Astronomical School’s Report* 5, 264.
- Andronov, I.L., Chinarova, L.L., 1997. On statistical properties of test functions in nonparametric methods for periodogram analysis. *Kinematics and Physics of Celestial Bodies* 13 (6), 55–65.

- Andronov, I.L., Chinarova, L.L., 2001. Determination of characteristic time scales in semi-regular stars: comparison of different methods. *Odessa Astronomical Publications* 14, 113–115.
- Andronov, I.L., Chinarova, L.L., 2003. Statistical study of semi-regular variables: scalegram-based characteristics. In: *ASP Conference Series*, vol. 292, p. 401.
- Andronov, I.L., Chinarova, L.L., 2013. Method of running sines: modeling variability in long-period variables. *Czestochowski Kalendarz Astronomiczny* 10, 171.
- Andronov, I.L., Kulynska, V.P., 2020. Computer modeling of irregularly spaced signals. Statistical properties of the wavelet approximation using a compact weight function. *Annales Astronomiae Novae* 1, 167–178. arXiv:1912.13096 [astro-ph.IM].
- Andronov, I.L., Marsakova, V.I., 2006. Variability of long-period pulsating stars. I. Methods for analyzing observations. *Astrophysics* 49, 370.
- Andronov, I.L., Shakun, L.I., 1990. *Astrophysics and Space Science* 169, 237.
- Andronov, I.L., Tkachenko, M.G., 2013. Comparative analysis of numerical methods of determination of parameters of binary stars. Case of spherical components. *Odessa Astronomical Publications* 26, 204.
- Andronov, I.L., et al., 1997. Scalegram analysis of the variability of the polar AM HER. *Odessa Astronomical Publications* 10, 15.
- Andronov, I.L., et al., 2003a. Inter-Longitude Astronomy project: some results and perspectives. *Astronomical and Astrophysical Transactions* 22, 793.
- Andronov, I.L., et al., 2003b. In: *NATO Science Series II*, vol. 105, 325. 2003whdw.conf..325A.
- Andronov, I.L., et al., 2005. Four-component model of the auto-correlation function of AM Her based on a CHANDRA observation. In: *ASP Conference Series*, vol. 330, p. 407.
- Andronov, I.L., et al., 2008. *Astronomy & Astrophysics* 486, 855. 2008A&A...486..855A.
- Andronov, I.L., et al., 2014. Inter-Longitude Astronomy project: long period variable stars. *Advances in Astronomy and Space Physics* 4, 3.
- Andronov, I.L., et al., 2017a. Instabilities in interacting binary stars. *ASP Conference Series* 511, 43.
- Andronov, I.L., et al., 2017b. Comparative analysis of phenomenological approximations for the light curves of eclipsing binary stars with additional parameters. *Astrophysics* 60, 57.
- Andrych, K.D., et al., 2015. “Asymptotic parabola” fits for smoothing generally asymmetric light curves. *Odessa Astronomical Publications* 28, 158.
- Andrych, K.D., et al., 2017. Statistically optimal modeling of flat eclipses and exoplanet transitions. The “Wall-Supported Polynomial” (WSP) algorithms. *Odessa Astronomical Publications* 30, 57.
- Andrych, K.D., et al., 2020. MAVKA: program of statistically optimal determination of phenomenological parameters of extrema. Parabolic spline algorithm and analysis of variability of the semi-regular star Z UMa. *Journal of Physical Studies* 24, 1902. arXiv:1912.07677 [astro-ph.SR], pp. 1–18.
- Andrych, K.D., Andronov, I.L., 2019. MAVKA: software for statistically optimal determination of extrema. *Open European Journal on Variable Stars* 197, 65.
- Bendat, J.S., Piersol, A.G., 2010. *Random Data, Analysis and Measurement Procedures*. John Wiley and Sons, New York.
- Blackman, R.B., Tukey, J.W., 1959. *The Measurement of Power Spectra*. Dover Publications, Inc., NY.
- Bódi, A., Szatmáry, K., Kiss, L.L., 2016. Periodicities of the RV Tauri-type pulsating star DF Cygni: a combination of Kepler data with ground-based observations. *Astronomy & Astrophysics* 596, A24, 8 pp.
- Bopp, B.W., Evans, D.S., Laing, J.D., Deeming, T.J., 1970. Six spectroscopic binary stars. *Monthly Notices of the Royal Astronomical Society* 147, 355–366. <https://doi.org/10.1093/mnras/147.4.355>.
- Box, G.E.P., et al., 2015. *Time Series Analysis: Forecasting and Control*, 5th Edition. Wiley, p. 712.
- Breus, V.V., 2007. Programs for data reduction and optimization of the system work. *Odessa Astronomical Publications* 20, 32.
- Butters, O.W., West, R.G., Anderson, D.R., et al., 2010. The first WASP public data release. *Astronomy & Astrophysics* 520, L10. <https://doi.org/10.1051/0004-6361/201015655>.
- Cherepashchuk, A.M., 1993. Parametric models in inverse problems of astrophysics. *Astronomy Reports* 37 (6), 585–594.
- Chinarova, L.L., 2010. Wavelet analysis of 173 semi-regular variables. *Odessa Astronomical Publications* 23, 25.
- Chinarova, L.L., Andronov, I.L., 2000. Catalogue of main characteristics of pulsations of 173 semi-regular stars. *Odessa Astronomical Publications* 13, 116.
- Clarke, D., 2002. String/Rope length methods using the Laffer-Kinman statistic. *Astronomy & Astrophysics* 386, 763–774. <https://doi.org/10.1051/0004-6361:20020258>.
- Deeming, T.J., 1970. Stochastic variable stars. *The Astronomical Journal* 75, 1027. <https://doi.org/10.1086/111056>.
- Deeming, T.J., 1975. Fourier analysis with unequally-spaced data. *Astrophysics and Space Science* 36, 137.
- Devlen, A., 2015. A new variable star in Perseus: GSC 3692-00624. *Open European Journal on Variable Stars* 171, 1.
- Drake, A.J., et al., 2009. First results from the Catalina Real-Time Transient Survey. *The Astrophysical Journal* 696, 870.
- Dumont, T., Morguleff, N., Rutily, B., Terzan, A., 1978. A search for periodicities in the pulsations of Delta Scuti stars. IV. A new method of computing the period of a complex signal. *Astronomy & Astrophysics* 69, 65.
- Dworetsky, M.M., 1983. A period-finding method for sparse randomly spaced observations or “How long is a piece of string?”. *Monthly Notices of the Royal Astronomical Society* 203, 917–924. <https://doi.org/10.1093/mnras/203.4.917>.
- Efron, B., 1979. Bootstrap methods: another look at the jack-knife. *The Annals of Statistics* 7, 1–26.
- Efron, B., Tibshirani, R.J., 1993. *An Introduction to the Bootstrap*. Chapman & Hall. ISBN 978-0412042317, pp. 436.
- Fisher, R.A., 1954. *Statistical Methods for Research Workers*, 12th Ed. Oliver and Boyd, Edinburgh, pp. xv, 356.

- Forsythe, G.E., Malcolm, M.A., Moler, C.B., 1977. Computer Methods for Mathematical Computations. Prentice Hall, p. 270.
- Foster, G., 1995. The cleanest Fourier spectrum. *The Astronomical Journal* 109, 1889–1902.
- Fourier, J., 1822. *Theorie Analytique de la Chaleur*. Firmin Didot. (Reissued by Cambridge University Press, 2009; ISBN 978-1-108-00180-9).
- GAIA, 2019. <http://gea.esac.esa.int/archive/>.
- Golyandina, N., Nekrutkin, V., Zhigljavsky, A., 2001. Analysis of Time Series Structure. SSA and Related Techniques. Chapman Hall. ISBN 1-58488-194-1, p. 308.
- Hamming, R.W., 1997. *Digital Filters*, Third Edition. Dover Publications, p. 304.
- Haykin, S., 1999. *Neural Networks and Learning Machines*, 3rd Ed. Prentice Hall, NJ. ISBN 978-0-13-147139-9, 906 p.
- Høg, E., et al., 2000. The Tycho-2 catalogue of the 2.5 million brightest stars. *Astronomy & Astrophysics* 355, L27–L30.
- Honeycutt, R.K., 1992. CCD ensemble photometry on an inhomogeneous set of exposures. *Publications of the Astronomical Society of the Pacific* 104, 435.
- Isobe, T., et al., 1990. Linear regression in astronomy. I. *The Astrophysical Journal* 364, 104.
- Jetsu, L., Pelt, J., 1996. Searching for periodicity in weighted time point series. *Astronomy & Astrophysics. Supplement Series* 118, 587.
- Jurkevich, I., 1971. A method of computing periods of cyclic phenomena. *Astrophysics and Space Science* 13, 154.
- KEPLER, 2019. www.nasa.gov/mission_pages/kepler/.
- Kim, C.-H., Kreiner, J.M., Nha, L.-S., 2003. Statistics of times of minimum light of 1140 eclipsing binary stars. *Astrophysics and Space Science Library* 298, 127.
- Kim, Y., et al., 2004. CCD photometry using multiple comparison stars. *Journal of the Astronomy and Space Sciences* 21 (3), 191–200.
- Kim, Y., et al., 2005. Orbital and spin variability of the intermediate polar BG CMi. *Astronomy & Astrophysics* 441, 663.
- Kim, Y., et al., 2009. Nova-like cataclysmic variable TT Arietis. QPO behaviour coming back from positive superhumps. *Astronomy & Astrophysics* 496, 765.
- Kim, Y., et al., 2018. A comprehensive catalog of galactic eclipsing binary stars with eccentric orbits based on eclipse timing diagrams. *The Astrophysical Journal. Supplement Series* 235, 41.
- Kopal, Z., 1959. *Close Binary Systems*. Chapman & Hall, London. Bibcode: 1959cbs..book....K.
- Kopal, Z., Kurth, R., 1957. The relation between period and times of the maxima or minima of variable stars. *Zeitschrift für Astrophysik* 42, 90.
- Kreiner, J.M., Kim, Chun-Hwey, Nha, Il-Seong, 2001. *An Atlas of O – C Diagrams of Eclipsing Binary Stars*. Wydawnictwo Naukowe Akademii Pedagogicznej, Cracow, Poland.
- Lafler, J., Kinman, T.D., 1965. An RR Lyrae star survey with the Lick 20-inch astrograph II. The calculation of RR Lyrae periods by electronic computer. *The Astrophysical Journal. Supplement Series* 11, 216–222. <https://doi.org/10.1086/190116>.
- Lawson, C.L., Hanson, R.J., 1974. *Solving Least Squares Problem*. Prentice - Hall Inc., Englewood Cliffs, New Jersey.
- Levenberg, K., 1944. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics* 2 (2), 164–168. <https://doi.org/10.1090/qam/10666>.
- Lipunov, V.M., et al., 2010. Master robotic net. *Advances in Astronomy*, 349171, 6 pp.
- Lomb, N.R., 1976. Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science* 39, 447–462. <https://doi.org/10.1007/BF00648343>.
- Mann, S., Haykin, S., 1991. The chirplet transform: a generalization of Gabor's Logon Transform. In: *Proc. Vision Interface*, pp. 205–212.
- Marquardt, D., 1963. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics* 11 (2), 431–441. <https://doi.org/10.1137/0111030>.
- Marsakova, V.I., Andronov, I.L., 1996. Local fits of signals with asymptotic branches. *Odessa Astronomical Publications* 9, 127.
- Mikulášek, Z., 2007a. The benefits of the orthogonal LSM models. *Odessa Astronomical Publications* 20, 138.
- Mikulášek, Z., 2007b. Principal component analysis – an efficient tool for variable stars diagnostics. *Astronomical and Astrophysical Transactions* 26, 63.
- Mikulášek, Z., 2015. Phenomenological modelling of eclipsing system light curves. *Astronomy & Astrophysics* 584, A8, 13 pp.
- Olspert, N., Pelt, J., Käpylä, M.J., Lehtinen, J., 2018. Estimating activity cycles with probabilistic methods. I. Bayesian generalised Lomb-Scargle periodogram with trend. *Astronomy & Astrophysics* 615, A111. <https://doi.org/10.1051/0004-6361/201732524>.
- Paczynski, B., 1986. Gravitational microlensing by the galactic halo. *The Astrophysical Journal* 304, 1. ADS:1986ApJ...304....1P.
- Paschke, A., 2018. A list of minima and maxima timings. *Open European Journal on Variable Stars* 191, 1.
- Paunzen, E., Vanmunster, T., 2016. Peranso – light curve and period analysis software. *Astronomische Nachrichten* 337, 239–246.
- Pelt, J., 1975. Methods for Search of Variable Star Periods. *Tartu Astrofüüs. Obs. Teated, Nr. 52, 24 p.*, = Preprint No. 5.
- Pojmanski, G., 2002. The All Sky Automated Survey. Catalog of variable stars. I. 0h - 6h quarter of the southern hemisphere. *Acta Astronomica* 52, 397–427. 2002AcA....52..397P.
- Prabhu, K.M.M., 2014. *Window Functions and Their Applications in Signal Processing*. CRC Press, Boca Raton, FL. ISBN 978-1-4665-1583-3.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2007. *Numerical Recipes. The Art of Scientific Computing*. Cambridge University Press, p. 1262.
- Renson, P., 1978. Method for finding the periods of variable stars. *Astronomy & Astrophysics* 63, 125–129.
- Roberts, D.H., Lehar, J., Dreher, J.W., 1987. Time series analysis with clean – part one – derivation of a spectrum. *The Astronomical Journal* 93 (4), 968–989.

- Samus, N.N., Kazarovets, E.V., Durlevich, O.V., Kireeva, N.N., Pastukhova, E.N., 2017. General catalogue of variable stars: version GCVS 5.1. *Astronomy Reports* 61, 80.
- Scargle, J.D., 1982. Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal* 263, 835–853. <https://doi.org/10.1086/160554>.
- Schaffer, J., 2015. What not to multiply without necessity. *Australasian Journal of Philosophy* 93 (4), 644–664.
- Scheffe, H., 1959. *The Analysis of Variance*. Wiley, p. 477.
- Shao, J., Tu, D., 1996. *The Jackknife and Bootstrap*, 2nd corrected printing. Springer. ISBN 978-1461269038, pp. 517.
- Shul'berg, A.M., 1971. Close Binary Systems with Spherical Components. Nauka, Moskva, p. 246.
- Stellingwerf, R.F., 1978. Period determination using phase dispersion minimization. *The Astrophysical Journal* 224, 953.
- Street, R.A., et al., 2003. SuperWASP: wide angle search for planets. In: *ASP Conference Series*, vol. 294, pp. 405–408.
- Sturges, H.A., 1926. The choice of a class interval. *Journal of the American Statistical Association* 21 (153), 65–66.
- Sutherland, P.G., et al., 1978. Short-term time variability of Cygnus X-1. II. *The Astrophysical Journal* 219, 1029.
- Terebizh, V.Yu., 1998. Similarity law in spectral estimation of a time series. *V. Astrophysics* 41, 198–201.
- TESS, 2019. heasarc.gsfc.nasa.gov/docs/tess/.
- Tikhonov, A.N., 1963. On the solution of ill-posed problems and the method of regularization. *Doklady Akademii Nauk SSSR* 151, 501.
- Tkachenko, M.G., Andronov, I.L., Chinarova, L.L., 2016. Phenomenological parameters of the prototype eclipsing binaries Algol, β Lyrae and W UMa. *Journal of Physical Studies* 20, 4902.
- Tolstov, G.P., 2012. *Fourier Series*. Courier Corporation, p. 352.
- Tremko, J., et al., 1996. Periodic and aperiodic variations in TT Arietis. Results from an international campaign. *Astronomy & Astrophysics* 312, 121.
- Tesevich, V.P. (Ed.), 1973. *Eclipsing Variable Stars*. J. Wiley, New York.
- Udalski, A., Kubiak, M., Szymanski, M., 1997. Optical gravitational lensing experiment. OGLE-2 – the second phase of the OGLE project. *Acta Astronomica* 47, 319–344.
- Vavilova, I.B., et al., 2012. Astroinformation resource of the Ukrainian virtual observatory: joint observational data archive, scientific tasks, and software. *Kinematics and Physics of Celestial Bodies* 28, 85.
- Vavilova, I.B., et al., 2017. UkrVO astroinformatics software and web-services. In: *Astroinformatics, Proceedings of the International Astronomical Union, IAU Symposium*, vol. 325, p. 361.
- Wehlau, William, Leung, Kam-Ching, 1964. The multiple periodicity of Delta Delphini. *The Astrophysical Journal* 139, 843.
- Wermuth, N., 2011. Multivariate statistical analysis. In: Lovric, M. (Ed.), *International Encyclopedia of Statistical Science*, Part 13. Springer, New York, pp. 915–920.
- Wozniak, P.R., et al., 2004. Northern sky variability survey: public data release. *The Astronomical Journal* 127, 2436–2449. ADS:2004AJ....127.2436W.
- Wright, E.L., et al., 2010. The Wide-field Infrared Survey Explorer (WISE): mission description and initial on-orbit performance. *The Astronomical Journal* 140, 1868.
- ZTF. <https://irsa.ipac.caltech.edu/Missions/ztf.html>.