

УДК: 524.331

## ИССЛЕДОВАНИЕ ПЕРЕМЕННОСТИ ДОЛГОПЕРИОДИЧЕСКИХ ПУЛЬСИРУЮЩИХ ЗВЕЗД. I. МЕТОДЫ АНАЛИЗА НАБЛЮДЕНИЙ

И.Л.АНДРОНОВ, В.И.МАРСАКОВА

Поступила 15 апреля 2006

Принята к печати 9 июня 2006

Обсуждаются методы обработки временных рядов долгопериодических переменных звезд по наблюдениям из баз данных наблюдателей переменных звезд. Проведено сравнение методов с точки зрения получения лучшей точности характеристик индивидуальных циклов переменности.

1. *Введение.* Так как долгопериодические переменные звезды характеризуются сильной изменчивостью кривых блеска от цикла к циклу, то для выявления закономерностей в их переменности и проведения их точной классификации необходимы длинные ряды наблюдений, охватывающие десятки и сотни пульсационных циклов.

Мы использовали наблюдения из баз данных Ассоциации Французских Наблюдателей Переменных Звезд AFOEV [1] и Японской Лиги Наблюдателей Переменных Звезд VSOLJ [2]. Они охватывают переменность в интервале до 80 лет, начиная с 1921г. Конечная дата различная для разных звезд, так как базы данных пополнялись с течением времени. Однако в интервале около JD 2431000-39000 часто данные были редки или вообще отсутствовали. Поскольку данные представляют собой смесь наблюдений большого количества любителей, каждый из которых обладает индивидуальной фотометрической системой, они показывают значительный разброс  $0^m.3-0^m.5$ , в то время, как точность оценок блеска наиболее опытных наблюдателей существенно лучше - до  $0^m.07$ .

2. *Первичная обработка наблюдений.* Для нашего анализа из данных по каждой звезде были удалены неуверенные значения (т.е. оценки, отмеченные как "неуверенные", самими наблюдателями) и оценки блеска "слабее чем". Для основного анализа использовались наблюдения, полученные в визуальном диапазоне. Часть наблюдений вблизи одного из минимумов  $R$  Суг показаны на рис.1, вместе с рассматриваемыми аппроксимациями.

Чтобы уменьшить разброс данных, применялась следующая методика.

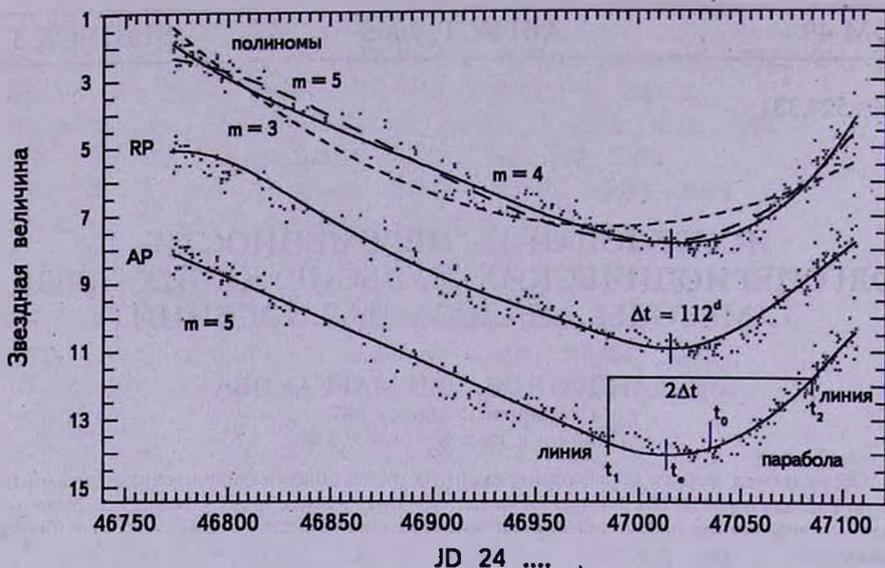


Рис.1. Аппроксимации отдельного экстремума кривой блеска мириды R Суг различными методами с числом определяемых параметров  $m$ : полиномы (вверху), "скользящие" (в середине) и "асимптотические" (внизу) параболы. Исходные данные (253 точки) показаны трижды для различных аппроксимаций со сдвигом  $3^m$ . Вертикальными черточками показаны оценки момента экстремума  $t_e$ . На нижнем рисунке отмечены границы ( $t_1$ ,  $t_2$ ) перехода между линейными асимптотами и переходной параболой при аппроксимации методом "асимптотических парабол", а также середина параболического интервала  $t_0 = (t_1 + t_2)/2$ .

Через все наблюдения проводилась сглаживающая кривая методом скользящих парабол [3], который будет более подробно рассмотрен ниже. Для каждого наблюдателя вычислялось среднее значение отклонения блеска от сглаживающей кривой  $O-C$ . Если оно оказывалось отстоящим от общей сглаживающей кривой более, чем на  $3\sigma_1$  ( $\sigma_1$  - среднеквадратичная точность среднего значения для каждого наблюдателя), то ко всем наблюдениям этого наблюдателя прибавлялось это среднее значение. Если впоследствии звездная величина какого-либо наблюдения отстояла более чем на  $3\sigma_1$  от сглаживающей кривой, то данная точка удалялась. Для реализации этого алгоритма использовалась программа [4]. Таким образом, осуществлялось приведение разных наблюдателей к одной фотометрической системе и фильтрация сильно выпадающих точек. При амплитудах, больших  $2^m$ , характерных для исследуемых звезд и достаточно плотном расположении наблюдений по времени, эта методика позволяет получать хорошую точность сглаживающей кривой.

3. *Средняя фазовая кривая.* Система нормальных уравнений, соответствующих методу наименьших квадратов (МНК), может быть записана как

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

$$A_{\alpha\beta} = \sum_{k=1}^n f_{\alpha}(t_k) \cdot f_{\beta}(t_k), \quad (2)$$

$$B_{\beta} = \sum_{k=1}^n x_k \cdot f_{\beta}(t_k), \quad (3)$$

где коэффициенты  $C_{\alpha}$ ,  $\alpha = 1 \dots m$  зависят от наблюдений  $x_k$ , полученных в моменты  $t_k$  и от выбора базисных функций  $f_{\alpha}(t)$ . Здесь  $n$  - количество наблюдений, и  $m$  - число определяемых параметров, иначе называемое "порядком модели". Здесь и далее мы предполагаем, что погрешности отдельных наблюдений статистически независимы, и их погрешности одинаковы для всех точек. Общий случай некоррелированных погрешностей с разными весами и дополнительной функцией "веса" ("фильтра") рассмотрен Андроновым [3]. Статистическая теория обратных задач рассмотрена в монографии Теребижа [5].

Сглаживающая функция  $x_c(t)$  и ее производная  $x_c^{[s]}$  порядка  $s$  по параметру  $t$  равны [6]:

$$x_c(t) = \sum_{\alpha=1}^m C_{\alpha} f_{\alpha}(t), \quad (4)$$

$$x_c^{[s]}(t) = \sum_{\alpha=1}^m C_{\alpha} f_{\alpha}^{[s]}(t). \quad (5)$$

Среднеквадратичная погрешность их значений может быть определена как

$$\sigma^2[x_c^{[s]}(t)] = \sigma^2 \sum_{\alpha\beta=1}^m A_{\alpha\beta}^{-1} f_{\alpha}^{[s]}(t) f_{\beta}^{[s]}(t), \quad (6)$$

где  $\sigma^2$  - погрешность "единицы веса", а  $A_{\alpha\beta}^{-1}$  - матрица, обратная к  $A_{\alpha\beta}$ . Математическое ожидание величины  $\sigma^2$  равно

$$\sigma_m^2 = \frac{1}{n-m} \sum_{k=1}^n [x_k - x_c(t_k)]^2 \quad (7)$$

для всех  $m$ , для которых нет систематических отличий наблюдений от сглаживающей функции. Однако мы используем не генеральную дисперсию  $\sigma^2$ , а ее выборочную оценку  $\sigma_m^2$ . Зависимость тест-функции (7) показана на рис.2. Она показывает быстрое уменьшение, сменяющееся практически постоянным уровнем для  $m > 4$ .

Момент экстремума  $t_e$  является корнем уравнения  $x_c^{[1]}(t_e) = 0$ . Среднеквадратичная оценка погрешности равна

$$\sigma[t_e] = \frac{\sigma[x_c^{[1]}(t_e)]}{|x_c^{[2]}(t_e)|}. \quad (8)$$

Эти выражения могут быть обобщены для случая, когда сглаживающая функция зависит от дополнительных параметров  $D_\gamma$ ,  $\gamma = 1 \dots p$ .

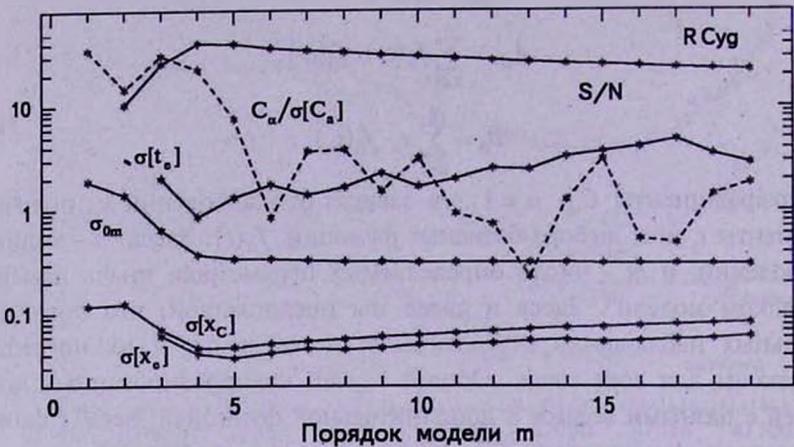


Рис. 2. Зависимость от порядка модели  $m$  тест-функций  $S/N$ ,  $\sigma_{0m}$ ,  $\sigma[x_c]$  для тестового минимума, показанного на рис. 1. Для числа параметров  $m > 2$  (парабола и полиномы высоких степеней) показаны оценки точности определения момента экстремума  $\sigma[t_c]$  и соответствующего значения сигнала  $\sigma[x_c]$ . Для трех критериев (максимум  $S/N$ , минимумы  $\sigma[t_c]$  и  $\sigma[x_c]$ ), оптимальное значение  $m = 4$  (полином степени  $m - 1 = 3$ ).

Можно выбрать начальное приближение и затем определять дифференциальные поправки, используя систему условных уравнений

$$\sum_{\alpha=1}^m f_{\alpha}(x) \cdot \delta C_{\alpha} + \sum_{\gamma=1}^p \frac{\partial x_c}{\partial D_{\gamma}} \delta D_{\gamma} = x_k - x_c(t_k). \quad (9)$$

После нескольких итераций можно определить  $m + p$  параметров  $C_{\alpha}$ ,  $\partial D_{\gamma}$  и соответствующие оценки ошибок. Очевидно, производные

$$f_{m+\gamma}(t) = \frac{\partial x_c(t)}{\partial D_{\gamma}} = \sum_{\alpha=1}^m C_{\alpha} \left( \frac{\partial f_{\alpha}(t)}{\partial D_{\gamma}} \right)_{t=t_k} \quad (10)$$

могут быть определены, используя значения коэффициентов  $C_{\alpha}$ , определенных из  $m$  нормальных уравнений для фиксированных значений  $D_{\gamma}$ .

Другими словами, в линеаризованной модели  $f_{m+\gamma}(t)$  могут рассматриваться как базисные функции с дополнительными номерами от  $m + 1$  до  $m + p$ . Можно формально написать  $\delta C_{m+\gamma} = \delta D_{\gamma}$  и считать число неизвестных равным  $m' = m + p$ . Оценки погрешностей могут быть вычислены по приведенным выше формулам. Важно отметить, что оценки погрешностей коэффициентов  $C_{\alpha}$ , полученные по моделям с  $m$  и  $m + p$  неизвестными, в общем случае различны.

В МНК обычно сравнивают дисперсию отклонений от сглаживающей функции с дисперсией исходных наблюдений. В качестве тест-функции мы использовали статистику

$$S(f) = \frac{\sigma_c^2}{\sigma_o^2} = 1 - \frac{\sigma_{o-c}^2}{\sigma_o^2}, \quad (11)$$

где  $\sigma_o$  - среднеквадратичное отклонение "наблюдений" ( $O$ ) от среднего.

Индекс  $C$  соответствует "расчетным" (или "вычисленным") значениям и  $O-C$  - отклонениям "наблюдаемых" значений от "расчетных".

Если значения  $x_k$  являются некоррелированными случайными значениями, распределенными нормально с одинаковыми средними и дисперсиями, то случайная переменная  $S$  имеет В-распределение

$$\rho(S) = \frac{\Gamma(\mu + \nu)}{\Gamma(\mu)\Gamma(\nu)} S^{\mu-1} (1 - S)^{\nu-1}$$

(например [5]) с параметрами  $\mu = n_1/2$ ,  $\nu = n_2/2$ , где  $n_1$  - число дополнительных степеней свободы, используемых для приближения по сравнению с использованным для определения  $\sigma_o$ . Математическое ожидание среднего значения -  $\langle S \rangle = \mu/(\mu + \nu) = n_1/(n_1 + n_2)$ .

Обсуждение методов поиска периода с применением МНК приведено Андроновым в [6]. В случае периодограммного анализа, математическая модель:

$$x_c(t) = C_1 + C_2 \sin(\omega t) + C_3 \cos(\omega t).$$

Таким образом, значения базисных функций в аргументах  $t_k$  -

$$f_1(t_k) = 1, \quad f_2 = \sin(\omega t_k), \quad f_3 = \cos(\omega t_k).$$

Здесь  $\omega = 2\pi \cdot f$ , где  $f = 1/P$  - пробная частота,  $P$  - пробный период.

"Истинное" приближение гармонической функцией методом наименьших квадратов отличается от широко используемых приближений Диминга [7], Ломба [8] и Скаргля [9]. Подробное описание преобразования Фурье и его модификаций рассмотрено в монографии Теребижа [10].

Для определения  $\sigma_o$  используется  $n-1$  независимая переменная, поскольку из  $n$  значений вычитается среднее. Для одночастотной модели определяются два дополнительных параметра  $C_2$  и  $C_3$ . Таким образом, число степеней свободы -  $n_1 = 2$  и  $n_2 = n - 1 - n_1 = n - 3$ :

$$\rho(S) = \frac{n-3}{2} (1-S)^{(n-3)/2}.$$

Соответствующее среднее значение равно  $\langle S \rangle = 2/(n-1)$ , вероятность  $Pr_1 = \text{Prob}(S > S_o)$  того, что  $S > S_o$  равна

$$Pr_1 = \int_{S_o}^1 \rho(S) dS = (1 - S_o)^{(n-3)/2}. \quad (12)$$

Следует отметить, что это выражение существенно отличается от обычно используемого приближения  $Pr_1 = \exp(-S_o/\langle S \rangle)$  (например, [7]).

Это приближение фактически соответствует распределению  $\chi_2^2$ .

Однако распределение  $\chi^2$  может использоваться, только если известна генеральная дисперсия наблюдений  $\sigma_2$ . Практически вместо нее для определения  $S(f)$  используется выборочная оценка  $\sigma_0^2$ , получаемая из наблюдений.

Таким образом, статистически оправданным является использование В-распределения вместо  $\chi^2$ .

Для большого количества пробных частот необходимо оценить вероятность "ложной тревоги" [8,7]

$$Pr = 1 - (1 - Pr_1)^K, \quad (13)$$

где  $K$  - число "независимых частот", которое может быть оценено для  $n$  равномерно распределенных по времени наблюдений  $K = (f_{max} - f_{min})/\Delta f$ , где  $\Delta f = n/((n-1)(t_n - t_1))$ . Для наблюдений, распределенных по времени примерно равномерно, эта оценка  $K$  не является сильно искаженной.

Спектральное окно вычисляется как

$$S_\omega(f) = \left( \frac{1}{n} \sum_{k=1}^n \cos \omega t_k \right)^2 + \left( \frac{1}{n} \sum_{k=1}^n \sin \omega t_k \right)^2 \quad (14)$$

(например, [8,9]). Эта функция равна единице при  $\omega = 0$  и должна быть близка к нулю для "хороших" (почти равноотстоящих по времени) наблюдений. Если сигнал содержит периодические компоненты с частотами  $f_{0j}$  и спектральное окно имеет пики на частотах  $f_{\omega l}$ , то появляется ряд пиков вблизи "ложных" частот  $|f_{0j} \pm f_{\omega l}|$ .

Для поиска периодичности нами была использована программа FOUR-0 [6], в которой сначала вычисляется тест-функция и спектральное окно для сетки пробных частот с шагом  $\Delta f = \eta(t_n - t_1)$  с  $\eta \approx 0.05 - 0.10$  (например, [11]).

*Мультигармоническая аппроксимация.* Для периодических сигналов со сложной формой можно использовать базисные функции ( $f_1 = 1$ ,  $f_{2j} = \sin j \omega t$ ,  $f_{2j+1} = \cos j \omega t$ ,  $j = 1 \dots s$ ). Для случайных данных число степеней свободы -  $n_1 = 2s$ ,  $n_2 = n - 1 - 2s$ . После предварительного определения наилучшей частоты (например, методом одночастотного периодограммного анализа), можно использовать метод дифференциальных поправок

$$F_{2s+2}(t) = t \sum_{j=1}^s (C_{2j} \cos j \omega t - C_{2j+1} \sin j \omega t) j. \quad (15)$$

Оценка погрешности  $\sigma[\omega] = \sigma \cdot (A_{2s+2, 2s+2}^{-1})^{1/2}$ . Очевидно,  $\sigma[f] = \sigma[\omega]/(2\pi)$  и  $\sigma[P] = \sigma[f] \cdot f^{-2}$ .

Проблема состоит в том, чтобы выбрать адекватную степень тригонометрического полинома  $s$ , т.е. число гармоник, равное  $s - 1$ . Число определяемых коэффициентов равно  $m = 2s + 1$ . Выбирается максимальное

число  $s_0$  (обычно 4-8) и вычисляются периодограммы для всех  $s \leq s_0$ . В программе FOUR-M [6] выбирается предварительное значение частоты и используется метод дифференциальных поправок. Если построить зависимость  $\sigma_s(s) = \sigma_{s,2s+1}$ , то можно заметить, что значение  $\sigma_s$  уменьшается с  $s$  для небольших  $s$  и затем становится практически постоянным в пределах оценок погрешности. Таким образом, первая часть задачи состоит в том, чтобы определить номер  $s$ , после которого существенное уменьшение  $\sigma_s$  прекращается [6,10].

Введем параметр

$$V_s = \frac{(n-2s+1)\sigma_s^2(s-1)}{(n-2s-1)\sigma_s^2(s)},$$

который (для случайных данных, имеющих нормальное распределение) является случайной величиной, имеющей распределение Фишера с 2 и  $(n-2s-1)$  степенями свободы (например, [12]).

Таким образом, задавая доверительную вероятность, определяем максимальное значение  $s$  в интервале от 1 до  $s_0$ .

Среднеквадратичное значение погрешности  $\sigma_{obs}$  сглаживающей функции в моменты времени наблюдений определяется из уравнения

$$\sigma_{obs}^2 = \frac{\sigma_s^2}{n} \sum_{k=1}^n \sum_{\alpha\beta=1}^m A_{\alpha\beta}^{-1} f_\alpha(t_k) f_\beta(t_k) = \frac{m}{n} \sigma_s^2. \quad (16)$$

Чтобы минимизировать статистическую ошибку сглаживающей кривой, нужно выбрать значение  $m$ , минимизирующее правую сторону этого уравнения. Однако для сильно зашумленных сигналов, значение  $\sigma_s$  уменьшается с  $m$  не очень быстро. Поэтому формально предпочтительно использовать одночастотное приближение или даже постоянное среднее значение.

Такая же проблема возникает при минимизации среднеквадратичной оценки погрешности  $\sigma_{phase}$  определенной интегрированием по всем фазам:

$$\sigma_{phase}^2 = \sigma_s^2 \frac{1}{P} \int_0^P \sum_{\alpha\beta=1}^m A_{\alpha\beta}^{-1} f_\alpha(t) f_\beta(t) dt = \sigma_s^2 \left( A_{11}^{-1} + \frac{1}{2} \sum_{\alpha=2}^m A_{\alpha\alpha}^{-1} \right).$$

В отличие от приведенной выше оценки погрешности момента экстремума, оценка погрешности  $\sigma[U]$  асимметрии  $U = (t_{max} - t_{min})/P$  может быть вычислена, используя более сложное выражение [6]

$$\sigma^2[U] = P^{-2} \sigma_s^2 \sum_{\alpha\beta=1}^m Z_\alpha Z_\beta.$$

Здесь  $Z_\alpha = z(t_{max}) - z(t_{min})$  и

$$Z_\alpha(t) = \frac{\dot{f}_\alpha(t)}{\sum_{s=1}^m C_s \ddot{f}_s(t)}.$$

$$\begin{aligned} \text{Для мультигармонического приближения, } z_1(t) &= 0, \\ Z_{2j}(t) &= -j \cos j \omega t / (\omega Y), \quad z_{2j+1}(t) = -j \sin j \omega t / (\omega Y), \end{aligned}$$

$$Y = \sum_{j=1}^s j^2 (C_{2j} \sin j \omega t + C_{2j+1} \cos j \omega t).$$

С целью уточнения среднего периода, среднего момента экстремума и других характеристик средней фазовой кривой блеска и ее формы, для исследуемых звезд в данных промежутках времени была использована программа FOUR-M [6], реализующая аппроксимацию тригонометрическим полиномом порядка  $s$ .

Частота начального приближения, обратная периоду из "Общего Каталога Переменных Звезд" [13], была исправлена методом дифференциальных поправок. Чтобы определить статистически значимое значение порядка модели  $s$ , мы использовали критерий Фишера с уровнем вероятности ложной тревоги  $Pr = 10^{-2}$  [12].

Недавно возрос интерес к сравнению коэффициентов аппроксимации тригонометрическим полиномом кривых блеска звезд типа  $\delta$  Сер с теоретическими моделями пульсаций, как видно, например, из работ [14,15].

4. *Определение характеристик экстремумов.* Для определения моментов экстремумов использовались два метода: скользящих парабол (running parabola, RP) [3] и "асимптотических парабол" (asymptotic parabola, AP) [16].

*Скользящие параболы.* В основе всех методов скользящего сглаживания лежит аппроксимация данных в некотором интервале  $[t_0 - \Delta t, t_0 + \Delta t]$  при помощи МНК с выбранными базисными и весовыми функциями. Сглаживающая функция состоит из вычисленных значений, взятых в средней точке каждого интервала сглаживания  $[t_0 - \Delta t, t_0 + \Delta t]$ , где аргумент  $t_0$  выбирается с произвольным шагом.

Полуширина фильтра  $\Delta t$  является свободным параметром, изменяя который необходимо найти компромисс между систематическим отклонением сглаживающей кривой от формы сигнала и статистическим весом сглаживающей величины.

Так же, как и другие методы, скользящие аппроксимации очень чувствительны к промежуткам в наблюдениях. Поэтому в практической реализации методов скользящих парабол и синусов аппроксимация для данного пробного момента принимается, если число наблюдательных точек, попадающих в интервал от  $t_0 - 0.5\Delta t$  до  $t_0 + 0.5\Delta t$  превышает 3; сглаживающие значения вычисляются для  $t_0$  в интервале от  $(1 + \alpha)x_{min} - \alpha x_{max}$  до  $(1 + \alpha)x_{max} - \alpha x_{min}$  (мы использовали  $\alpha = 0.05$ ); оценка погрешности сглаженных значений не превышает  $\sigma_0$ . Если любое из этих условий

не выполняется, то данные не записываются в файл. Это позволяет уменьшить число плохих значений автоматически.

Пусть задан дискретный набор наблюдательных данных  $(t_k, x_k)$ ,  $k=1 \dots n$ . Необходимо вычислить значение сглаживающей функции  $x_c(t_0)$  в некоторой точке  $t_0$ , не обязательно совпадающей с одним из аргументов наблюдений  $t_k$ . Для этого можно воспользоваться разложением функции  $x_c(t)$  вблизи точки  $t_0$  по базисным функциям, для полиномиальной аппроксимации, имеющим вид  $f_\alpha(t) = t_\alpha - 1$ ,  $\alpha=1 \dots m$ :

$$x_c(t) = \sum_{\alpha=1}^m C_\alpha(t_0) f_\alpha(t-t_0). \quad (17)$$

Система нормальных уравнений, соответствующих МНК (1-3), может быть записана как

$$\sum_{\alpha=1}^m A_{\alpha\beta} C_\alpha = B_\beta, \quad (18)$$

$$A_{\alpha\beta} = \sum_{k=1}^n p_k w_k f_\alpha(t_k - t_0) f_\beta(t_k - t_0), \quad (19)$$

$$B_\beta = \sum_{k=1}^n p_k w_k x_k f_\beta(t_k - t_0). \quad (20)$$

Коэффициенты разложения:

$$C_\alpha = \sum_{\beta=1}^m A_{\alpha\beta}^{-1} \sum_{k=1}^n p_k w_k x_k f_\beta(t_k - t_0).$$

Здесь  $w_k$  - "веса наблюдений", обратно пропорциональные квадратам оценки точности наблюдения:  $w_k = \sigma_o^2 / \sigma_k^2$ . Для наших задач неизвестна точность каждого наблюдения и принимается  $w_k = 1$ .

Множитель  $p_k(t_0)$  - функция фильтра, с помощью которой учитывается то обстоятельство, что при локальной аппроксимации берем наблюдения не во всем интервале, а лишь в окрестностях точки  $t_0$ . Здесь - полуширина фильтра.

Естественно, существует бесконечное множество функций, которые можно было бы использовать в качестве функции фильтра, и выбор между ними связан не со статистическим обоснованием, а с дополнительными соображениями, в частности с требованиями гладкости, минимальной систематической погрешности и максимального веса попадающих в данный интервал наблюдений.

Для практических приложений использовалась функция

$$p_k(\varepsilon) = \begin{cases} (1 - \varepsilon^2)^2, & \text{где } \varepsilon^2 \leq 1, \\ 0, & \text{где } \varepsilon^2 \geq 1, \end{cases}$$

которая является сплайном четвертой степени дефекта 3, поскольку она и ее первая производная непрерывны, а вторая, третья и четвертая

- терпят разрыв при  $\varepsilon = \pm 1$ . Она может быть примером компромисса между требованием гладкости функции и наибольшего суммарного веса наблюдений при фиксированном значении  $\Delta t$ .

Для нахождения оптимального  $\Delta t$  необходимо вычислить зависимости от  $\Delta t$  следующих параметров:  $\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5$ , (они описаны ниже) и отношения сигнал/шум. Здесь  $\sigma_1, \sigma_2$  - оценки ошибок единицы веса наблюдений, вычисленные в соответствии с уравнениями, приведенными в [3]. Величина  $\sigma_1$  определяется с учетом отклонений наблюдений от каждой дуги параболы в интервале  $[t_0 - \Delta t, t_0 + \Delta t]$ :

$$\Phi = \Phi_t + \left( \sum_{k=1}^n p_k - \sum_{\alpha\beta=1}^m A_{\alpha\beta}^{-1} G_{\alpha\beta} \right) \sigma_0^2 = \Phi_t + \phi \sigma_0^2,$$

где

$$G_{\gamma\varepsilon} = \sum p_k^2 w_k f_\gamma(t_k - t_0) f_\varepsilon(t_k - t_0),$$

$$\Phi = \sum_{k=1}^n p_k w_k [x_k - \phi(t_k, t_0, \Delta t)]^2 -$$

полная сумма квадратов невязок,

$$\Phi_t = \sum_{k=1}^n p_k w_k [x_{kt} - \phi_t(t_k, t_0, \Delta t)]^2 -$$

сумма квадратов невязок, обусловленная систематическим отклонением сглаживающей кривой от наблюдений;

$A_{\alpha\beta}^{-1}$  - матрица, обратная матрице нормальных уравнений,  $p_k$  - значение функции фильтра,  $w_k$  - весовой множитель, связанный с точностью наблюдений,

$$\phi(t_0, t_0, \Delta t) = x_c(t_0);$$

$\sigma_2$  - определяется с учетом отклонений наблюдений от сглаживающей кривой только в каждой центральной точке  $t_0$ :

$$\langle w_k (x_k - x_c(t_k))^2 \rangle = \sigma_0^2 \left( 1 + w_k \sum_{\alpha\beta=0}^m f_\alpha(0) f_\beta(0) (R_{\alpha\beta} - 2 p_k A_{\alpha\beta}^{-1}) \right),$$

где

$$R_{\alpha\beta} = \sum_{\gamma\varepsilon=0}^m A_{\alpha\gamma}^{-1} G_{\gamma\varepsilon} A_{\varepsilon\beta}^{-1}.$$

Таким образом,  $\sigma_1$  всегда больше, чем  $\sigma_2$ . Здесь  $\sigma_3$  - среднеквадратичное отклонение наблюдений от аппроксимации; коэффициент пропорциональности  $R$  между ошибкой единицы веса и оценкой погрешности сглаживающей функции  $\sigma[m(t)]$ . Так как используются две величины  $\sigma_1$  и  $\sigma_2$ , то существуют две оценки величины  $\sigma[m(t)]$ :  $\sigma_4 = R \sigma_1$  и  $\sigma_5 = R \sigma_2$ . Параметр  $\sigma_c / \sigma_5$  - это отношение "сигнал/шум".

Существует несколько критериев выбора величины  $\Delta t$ : нахождение значения  $\Delta t$ , при котором отношение "сигнал/шум" максимально; минимизация оценки среднеквадратичного отклонения наблюдений от сглаживающей кривой; минимизация оценки точности сглаженного значения  $\sigma_4$  и  $\sigma_5$ .

Зависимости этих величин от  $\Delta t$ , полученные для аппроксимации всей кривой блеска R Cyg (13075 точек), показаны на рис.3. Величина, соответствующая максимуму отношения "сигнал/шум" ( $S/N = 30.5$ ) -  $\Delta t = 111^d.7$ . Эта величина более, чем в 2 раза меньше, чем значение  $0.54P = 230^d$ , ожидаемое для тригонометрического полинома первой степени, вследствие несинусоидальной формы кривой блеска. Здесь мы использовали значение периода пульсаций  $P = 426^d.45$  [13]. Другой критерий -

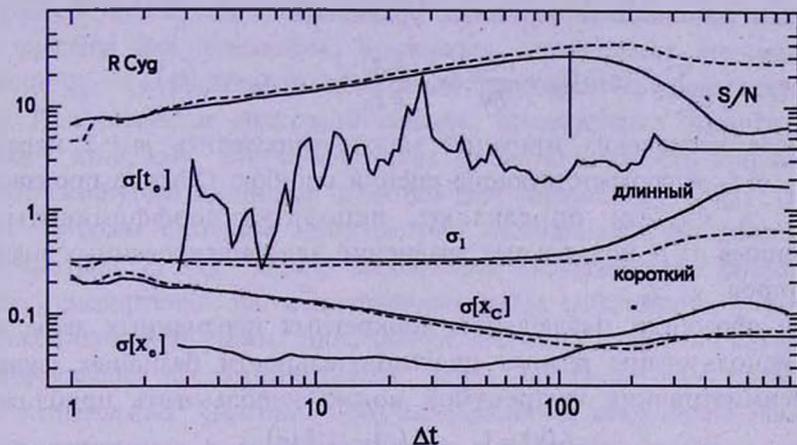


Рис.3. Зависимость от полуширины фильтра  $\Delta t$  параметров аппроксимации методом "скользящих парабол". Сплошными линиями показаны тест-функции  $S/N$ ,  $\sigma_1$  и  $\sigma[x_c]$  для "длинного" ряда, пунктиром - для "короткого" вблизи одного из минимумов (рис.1). Также показаны зависимости от  $\Delta t$  оценок погрешностей определения момента экстремума  $\sigma[t_e]$  и экстремального значения сглаживающей функции  $\sigma[x_e]$ . Вертикальная линия показывает величину  $\Delta t = 112^d$ , соответствующую максимуму отношения "сигнал/шум" для "длинного ряда".

минимизация оценки средней точности сглаженного значения  $\sigma_4$ . Этот параметр минимален при  $\Delta t = 111^d.1$ , что практически совпадает со значением, полученным для предыдущего критерия, и равен  $0^m.059$ . Смещение в область меньших значений несущественно. Такое поведение согласуется с иллюстративными примерами, приведенными в [3] и показывает согласованность оценок  $\Delta t$ , полученных по разным критериям.

*Асимптотические параболы.* Для определения характеристик кривых блеска переменных звезд, имеющих асимметричные экстремумы, нами был разработан метод "асимптотических парабол" [16].

В данной модели, коэффициенты  $C_n$  в системе нормальных урав-

нений МНК (1-3) зависят как от наблюдений  $x_k$  в моменты  $t_k$ , которые выбраны для использования вблизи экстремума, так и от выбора базисных функций  $f_\alpha(z)$ , а также начальной эпохи  $t_0$  и "масштаба времени"  $\Delta t$ , которые обеспечивают переход к безразмерным аргументам  $z = (t - t_0)/\Delta t = t/\Delta t + z_0$ .

Среднеквадратичная погрешность  $\sigma$  сглаживающей функции (4) и ее производных порядка  $s$  (5) по параметру  $z$  (в том числе в момент экстремума  $t_e$ ) может быть определена по формуле (6).

Оценка дисперсии момента экстремума  $t_e$  определяется формулой (8).

Эти формулы могут быть обобщены на случай, когда сглаживающая функция зависит от дополнительных параметров  $t_0$ ,  $\Delta t = v - 1$ . Нужно оценить их начальные значения, а потом определить дифференциальные поправки из системы условных уравнений

$$\sum_{\alpha=1}^m f_\alpha(z) \delta C_\alpha + \frac{\partial x_c}{\partial v} \delta v + \frac{\partial x_c}{\partial z_0} \delta z_0 = x_k - x_c(t_k). \quad (21)$$

После нескольких итераций можно определить  $m + 2$  параметра  $C_\alpha$ ,  $v$ ,  $\delta z_0$  и соответствующие оценки ошибок. Обычно производные  $\partial x_c / \partial v$  и  $\partial x_c / \partial z_0$  определяют, используя коэффициенты  $C_\alpha$ , полученные из  $m$  нормальных уравнений для фиксированных значений параметров  $v$ ,  $z_0$ .

Для обработки наблюдений конкретных переменных звезд могут быть использованы разные последовательности базисных функций. Для асимметричных экстремумов можно использовать приближение

$$f_1(z) = 1, \quad f_2(z) = -f_3(z)$$

с асимптотами для больших положительных  $z$

$$f_2(-z) = 0, \quad f_2(+z) = z,$$

Например, может быть использован "линейный сплайн" (LS) и "гиперболический тангенс" (НТ), для которых значения базисной функции и ее производной следующие:

	$z < -1$	$(-1 < z < 1)$	$(z > 1)$	
$f_1(z) =$	0	$(1/4)(1 + z)^2$	$z$	LS
$f_2^{(1)}(z) =$	0	$(1/2)(1 + z)$	1	
$f_1(z) =$	$(1/2)(z + \ln(e^z + e^{-z}))$			НТ
$f_2^{(1)}(z) =$	$(1/3)(1 + \text{th}z)$			

Величина  $2\Delta t$  (рис.1) может быть интерпретирована как "эффективная продолжительность" перехода с одной "асимптоты" на другую. Как показывают модельные расчеты, использование функции НТ приводит к вдвое большей ошибке экстремума  $\sigma[t_e]$ , чем при использовании LS, поэтому LS-аппроксимация была выбрана для практического применения

к обработке наблюдений долгопериодических переменных звезд. В дальнейшем мы будем использовать для нее название "метод асимптотических парабол", сокращенно AP.

Дополнительной сложностью для практических вычислений является определение положения точек перехода  $\xi_1$ ,  $\xi_2$  между параболой и линиями. В многократно использованной ранее программе *Asymp* [16], для каждого участка наблюдений, время линейно преобразуется к интервалу  $[0,1]$ , затем вычисляется тест-функция  $\sigma^2(\xi_1, \xi_2)$ , равная дисперсий невязок, для сетки пар значений  $\xi_1$ ,  $\xi_2$ . Их окончательные значения уточняются методом дифференциальных поправок, что позволяет использовать оценки погрешностей методом наименьших квадратов.

Для дополнительного тестирования метода асимптотических парабол, была разработана новая программа AP, в которой в базисный алгоритм были внесены два новшества. Во-первых, определение оптимальных значений  $\xi_1$ ,  $\xi_2$  проводится локальным перебором с уменьшающимся шагом. При равенстве дисперсий невязок, предпочтение отдается более широкому интервалу. Это обеспечивает большую гладкость кривых при наличии вытянутого минимума функции двух переменных  $\sigma^2(\xi_1, \xi_2)$  типа "оврага". Вторым отличием является учет нелинейности аппроксимации по параметрам  $\xi_1$ ,  $\xi_2$ . После вычисления сглаживающей функции и невязок, генерируются 100 искусственных рядов наблюдений, в которых к сглаженным значениям добавляются невязки в другой случайной неповторяющейся последовательности. Затем, для такого ряда проводится поиск оптимальных значений. Среднеквадратичное отклонение моментов времени экстремума и соответствующего сглаженного значения сигнала принимаются за оценку точности выборочного экстремума.

Эти две модификации далее обозначаются как AP1 и AP2.

Аппроксимация тестового минимума с помощью метода асимптотических парабол также показана на рис.1. Для длинных рядов, согласие сглаживающих функций по пересекающимся интервалам показывает как качество избранной аппроксимирующей функции, так и правильность выбора интервала вблизи экстремума. В случае плохого согласия, рекомендуется переопределить интервалы для сглаживания.

На рис.1 для сравнения показаны аппроксимации отдельного минимума методами: асимптотическими параболой (AP), полиномом 4-й степени (P5), скользящими параболой (RP). Первые два метода имеют по 5 определяемых параметров, таким образом, их сравнение статистически обосновано. В табл.1 приведены значения и погрешности момента и звездной величины экстремума для этих методов. Отметим, что степень полинома на единицу меньше, чем число определяемых параметров, так что, например, обозначение P3 (полином с  $m = 3$ )

соответствует полиному второй степени (параболе). Кроме того,  $P_{opt}$  показывает полиномиальную аппроксимацию с автоматическим определением степени, соответствующим наименьшей статистической погрешности момента экстремума.

Анализ табл.1 показывает, что для столь асимметричного экстремума, наихудшими по точности определения момента экстремума являются

Таблица 1

### ЗНАЧЕНИЯ И ПОГРЕШНОСТИ МОМЕНТА И ЗВЕЗДНОЙ ВЕЛИЧИНЫ ЭКСТРЕМУМА, ПОЛУЧЕННЫЕ РАЗЛИЧНЫМИ МЕТОДАМИ

Метод	$t_i$	$\sigma[t_i]$	$m_i$	$\sigma[m_i]$
P3	46986.2	2.0	13.331	0.061
P4	47007.0	0.9	13.831	0.040
P5	47015.9	1.3	13.849	0.037
P6	47016.9	1.7	13.834	0.041
RP	47015.6	2.0	13.929	0.041
AP1,2	47013.4	1.5	14.045	0.038

P3 (парабола) и RP (скользящая парабола). Однако RP дает значение, более близкое к ожидаемому.

Формально наилучшей по точности является кубическая парабола P4, однако, сам момент сильно смещен относительно значений, соответствующих группе около 47015 (здесь и далее время выражается в HJD-2400000).

Кроме обработки наблюдений долгопериодических переменных, метод также может быть успешно использован для аппроксимации минимумов затменных переменных звезд.

Для сравнения трех методов - P, RP и AP был вычислен средний вес по индивидуальным весам экстремумов  $\bar{w}_i = 1/\sigma_i^2$  где  $\sigma_i$  - статистическая оценка погрешности параметра цикла номер  $i$ . Средняя погрешность определения экстремума:

$$\bar{\sigma} = \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{\sigma_i^2} \right)^{-1/2},$$

где  $n$  - число исследуемых экстремумов.

Эти величины приведены на примере звезды R Cyg в табл.2, из которой видно, что для AP1-метода средняя погрешность определения момента времени  $\sigma_t$  в несколько раз меньше, чем для RP, а средняя погрешность звездной величины  $\sigma_m$  - в ~1.2 раза больше. Аналогичные соотношения были получены для других звезд.

На рис.1 показан пример аппроксимации экстремума звезды R Cyg

разными методами. Различие между двумя сглаживающими кривыми не превосходит оценку погрешности сглаженного значения, за исключением параболы, показывающей сильные систематические отклонения от хода кривой блеска. На кривой RP наблюдаются малоамплитудные волны, накладывающиеся на основной ход экстремума, что характерно для меньшей ширины фильтра сглаживания. Это и приводит к уменьшению оценки среднеквадратичного отклонения от сглаживающей кривой по

Таблица 2

СРЕДНИЕ ПОГРЕШНОСТИ ОПРЕДЕЛЕНИЯ ЭКСТРЕМУМОВ  
 ДЛЯ МЕТОДОВ P, AP И RP ПРИ АППРОКСИМАЦИИ 83  
 ЭКСТРЕМУМОВ R Cуг:  $\bar{\sigma}_t$  - ПОГРЕШНОСТЬ МОМЕНТА  
 ЭКСТРЕМУМА,  $\bar{\sigma}_m$  - ПОГРЕШНОСТЬ ЗВЕЗДНОЙ ВЕЛИЧИНЫ  
 ЭКСТРЕМУМА

	P3	P4	P5	P6	P <sub>opt</sub>	AP1	AP2	RP
$\bar{\sigma}_t$ , суг	1.28	1.73	1.95	2.41	1.19	0.54	1.89	2.80
$\bar{\sigma}_m$ , mag	53	45	50	54	51	63	54	54

сравнению с кривой AP и увеличению погрешности экстремума, более подверженного влиянию локальных флуктуаций.

Поэтому для основной массы звезд применялся AP-метод для определения характеристик экстремумов. RP-метод применялся в тех случаях, когда на кривой наблюдались сильные неправильности (типа вторичного максимума), которые искажали форму экстремумов.

В качестве еще одного теста для сравнения разных методов мы обработали 100 искусственных рядов по 101 точке:  $t_k = 2(k-1)/100 - 1$ ,  $x_k = t_k^2 + \alpha \epsilon_k$ , где  $\epsilon_k$  - случайная величина, распределенная по нормальному закону с нулевым средним и единичной дисперсией. Результаты анализа приведены в табл.3.

Для "слабого шума"  $\alpha = 0.1$ , асимптотические параболы AP1 по точности определения экстремума почти в два раза лучше полиномиальной аппроксимации P5 с тем же числом степеней свободы, в то время как погрешность экстремального значения сигнала отличается несущественно. Несомненный выигрыш по точности у параболы P3, что объясняется соответствием тестового сигнала аппроксимирующей функции. Как мы видели для асимметричных экстремумов R Cуг, P3 дает формально небольшую оценку погрешности, но реально большие значения систематического сдвига.

Для "очень сильного шума"  $\alpha = 1$ , экстремумы для всех аппроксимаций распределены почти по всему интервалу данных, что приводит

Таблица 3

СРЕДНИЕ ПОГРЕШНОСТИ ОПРЕДЕЛЕНИЯ ЭКСТРЕМУМОВ  
 ДЛЯ МЕТОДОВ AP И RP ПРИ АППРОКСИМАЦИИ  
 ЭКСТРЕМУМОВ МОДЕЛЬНОЙ ЗАШУМЛЕННОЙ ПАРАБОЛЫ:  
 $\bar{\sigma}_l$  - ПОГРЕШНОСТЬ МОМЕНТА ЭКСТРЕМУМА,  $\bar{\sigma}_m$  -  
 ПОГРЕШНОСТЬ ЗВЕЗДНОЙ ВЕЛИЧИНЫ ЭКСТРЕМУМА,  
 $\sigma_l$  и  $\sigma_m$  - СООТВЕТСТВУЮЩИЕ СРЕДНЕМУ ВЕСУ;  
 $\sigma_l$  и  $\sigma_m$  - РЕАЛЬНЫЙ СРЕДНЕКВАДРАТИЧНЫЙ РАЗБРОС

Метод	$\alpha = 1$				$\alpha = 0.1$			
	$\bar{\sigma}_l$	$\sigma_l$	$\bar{\sigma}_m$	$\sigma_m$	$\bar{\sigma}_l$	$\sigma_l$	$\bar{\sigma}_m$	$\sigma_m$
P3	0.084	0.094	0.148	0.140	0.0085	0.0081	0.0149	0.0143
P4	0.136	0.181	0.157	0.141	0.0212	0.0226	0.0149	0.0143
P5	0.118	0.293	0.182	0.201	0.0214	0.0228	0.0185	0.0189
P6	0.097	0.357	0.202	0.292	0.0361	0.0358	0.0186	0.0191
P <sup>ср</sup>	0.071	0.294	0.163	0.293	0.0085	0.0081	0.0150	0.0143
AP1	0.318	0.420	0.235	0.188	0.0092	0.0115	0.0171	0.0150
AP2	0.252	0.459	0.192	0.201	0.0129	0.0150	0.0192	0.0201

к колоссально большим значениям среднеквадратичного разброса. В этом случае наилучшей является P3, но это преимущество теряется, если экстремум сдвинут относительно середины интервала наблюдений или является сильно асимметричным.

Аппроксимация скользящими параболками позволяет отслеживать мелкие детали кривой, но они более чувствительны к разрывам в данных и недостаточному количеству наблюдательных точек.

Метод AP удобен для аппроксимации кривых блеска мирид, в том числе имеющих асимметричные экстремумы, которые могут сильно искажаться при сглаживании другими методами, в том числе и RP.

**5. Характеристики индивидуальных циклов.** Для каждого цикла переменности были определены такие основные параметры:

- моменты минимума и максимума, их фазы и звездные величины (определенные методами AP или RP);
- средний момент, фаза и звездная величина горба на восходящей ветви (если он присутствует). Момент определялся как среднее из крайних точек горба по оригинальной кривой, так как в большинстве случаев с помощью скользящих парабол горб прописывался плохо из-за малого количества наблюдательных точек и большой полуширины фильтра, выбранной оптимальной для всей кривой. Звездная величина - выборочное среднее из всех точек горба. В случае получения характеристик горба (вторичного максимума) как экстремума, с помощью

скользящих парабол, вычислялись как параметры максимума, так и следующего за горбом понижения блеска;

- для определения величин обратных наклонов  $dt/dm$  для индивидуальных циклов на восходящей и нисходящей ветвях были выбраны линейные участки индивидуально для каждого цикла и аппроксимированы прямой линией;

- также использовались моменты и фазы пересечений линейными участками уровней фиксированных звездных величин (для большинства звезд  $10^m$ );

Характеристики индивидуальных экстремумов частично опубликованы в [17], полный вариант каталога можно запросить у авторов.

Кроме основных параметров, вычислялись следующие, связанные с ними, характеристики:

- амплитуда восходящей и нисходящей ветви (разница звездных величин текущего максимума и двух соответствующих минимумов);

- период, как интервал времени, между последовательными максимумами;

- период, как интервал времени, между последовательными минимумами;

- продолжительности восходящей и нисходящей ветвей;

- асимметрия (два значения: как отношение продолжительности восходящей ветви к: 1 - периоду между максимумами, 2 - периоду между минимумами. Для корреляционного анализа, как правило, играл большую роль параметр 1, так как для него вычислялось больше значений, поскольку максимумы блеска наблюдаются в среднем чаще);

- разность звездных величин между последовательными максимумами;

- разность звездных величин между последовательными минимумами;

- средний блеск (звездная величина) на восходящей и нисходящей ветви.

Исследовалась зависимость всех этих величин от времени. При этом обоим периодам и асимметриям приписывался момент текущего максимума, а таким характеристикам, как амплитуда и средний блеск, которые дважды встречаются в одном цикле, приписывался средний момент времени между максимумом и соответствующим минимумом.

Когда это необходимо, проводился периодограммный анализ с целью определения характерных времен изменения периода. Для этой же цели изменения характеристик индивидуальных циклов аппроксимировались с помощью RP-метода. Для полуправильных звезд применялся всплеск-анализ [18] для изучения стабильности периода.

Перечисленные выше характеристики использовались нами для корреляционного анализа.

Если имеется  $m$  параметров, то это соответствует  $m(m-1)/2$  независимым значениям коэффициента корреляции. Кроме самого коэффициента корреляции  $\rho$ , вычислялось также значение параметра "сигнал/шум", равное

$$S = \frac{|\rho|}{\sigma_\rho} = |\rho| \cdot \sqrt{\frac{n-2}{1-\rho^2}},$$

где  $n$  - количество пар характеристик. При  $S > 3$ , т.е.  $|\rho| > 3\sigma_\rho$ , корреляция считается значимой [19].

Необходимо отметить, что часть корреляций является тривиальной. Таким образом, выборка значимых независимых корреляций сужается.

Обсуждение результатов для исследуемых звезд планируется привести в последующих работах.

Авторы выражают свою благодарность за полезное обсуждение В.П.Гринину, В.Ю.Теребижу и рецензенту.

Астрономическая обсерватория и кафедра астрономии Одесского национального университета им. И.И.Мечникова,  
Украина, e-mail: il-a@mail.od.ua vlada@mail.od.ua

## INVESTIGATION OF VARIABILITY OF LONG-PERIODIC PULSATING STARS. I. METHODS OF ANALYSIS OF OBSERVATIONS

I.L.ANDRONOV, V.I.MARSAKOVA

The methods of time series analysis of long periodic stars by the observation from databases of variable stars observers are discussed. The methods were compared to obtain the better accuracy of characteristics of individual variability cycles.

Key words: *Stars: long-period: pulsating - methods - observational*

## ЛИТЕРАТУРА

1. *E.Schweitzer*, <http://cdsarc.u-strasbg.fr/pub/afoev/>
2. *D.Nogami*, <http://www.kusastro.kyoto-u.ac.jp/vsnet/>
3. *I.L.Andronov*, *Astron. Astrophys. Suppl. Ser.*, **125**, 207, 1997.
4. *L.S.Shapovalova*, *Odessa Astron. Publ.*, **10**, 131, 1997.
5. *В.Ю.Теребиж*, Введение в статистическую теорию обратных задач. Наука, М., 2005, с.376.
6. *I.L.Andronov*, *Odessa Astron. Publ.*, **7**, 49, 1994.
7. *T.J.Deeming*, *Astron. Astrophys. Suppl. Ser.*, **36**, 137, 1975.
8. *N.R.Lomb*, *Astrophys. Space Sci.*, **39**, 447, 1976.
9. *J.D.Scargle*, *Astrophys. J.*, **263**, 835, 1982.
10. *В.Ю.Теребиж*, Анализ временных рядов в астрофизике, Наука, М., 1992, с.392.
11. *П.Н.Холопов*, Фотографическая фотометрия переменных звезд. Методы исследования переменных звезд, Наука, М., 1971, с.91.
12. *К.Мардиа, П.Земроч*, Таблицы F-распределений, Наука, М., 1985, с.640.
13. *П.Н.Холопов и др.*, Общий каталог переменных звезд, Т.1-3, Изд. 4-е, Наука, М., 1985.
14. *E.Antonello*, *Astron. Astrophys.*, **285**, 140, 1994.
15. *G.Kovacs, I.Shlosman, J.R.Buchler*, *Astrophys. J.*, **307**, 593, 1986.
16. *V.I.Marsakova, I.L.Andronov*, *Odessa Astron. Publ.*, **9**, 127, 1996.
17. *V.I.Marsakova, I.L.Andronov*, *Odessa Astron. Publ.*, **11**, 79, 1999.
18. *И.Л.Андронов*, Кинематика и физика небесных тел, **14**, 490, 1988.
19. *Г.А.Корн, Т.М.Корн*, Справочник по математике для научных работников и инженеров: Пер. с англ., Наука, М., 1970, с.720.