Mon. Not. R. Astron. Soc. **000**, 1–14 (2009) Printed 23 March 2009 (MN LATEX style file v2.2)

Light curve analysis of Variable stars using Fourier coefficients and Principal Component Analysis

Sukanta Deb and Harinder P. Singh⋆
Department of Physics & Astrophysics, University of Delhi, Delhi 110007, India

Received on ; Accepted on

ABSTRACT

We present an efficient and robust method of light curve analysis of variable stars. While several studies have successfully employed Fourier decomposition technique for the structural analysis of a variety of light curves, a few studies have performed a principal component analysis (PCA) by using Fourier components as inputs to reduce the dimensionality further. In this paper, we employ the PCA technique to a large database of RR Lyrae and Cepheid light curves directly and compare the results with those obtained by the Fourier decomposition technique. We show that the first few principal components (PCs) allow us to extract most of the information when the input matrix is an array of magnitudes at different epochs rather than the Fourier coefficients. The first few PCs are thus enough to reconstruct the original light curves, reducing the data dimensionality by a factor of 2 to 3 compared to the Fourier coefficients. The analysis enables us to separate out the modes of RR Lyrae stars and to look for previously known and suspected resonances of Cepheid variables. We also demonstrate that the PCA technique can be used to classify variables into different classes in an automated, unsupervised way, a feature that has immense potential for large databases of the future.

Key words: Methods: data analysis – RR Lyraes – Cepheids

1 INTRODUCTION

The recent interest on the structure and properties of the light curve of variable stars has increased a lot because of the large flow of observational data from different variable star projects like OGLE, MACHO, ASAS and NSVS. In addition, new techniques for tagging variable objects expected in huge numbers from satellite missions like CoRoT, Kepler, and Gaia in a robust and automated manner are being explored (Debosscher et al. 2007, Sarro et al. 2009). Fourier decomposition technique is a reliable and efficient way of describing the structure of light curves of variable stars. Schaltenbrand & Tammann (1971) derived UBV light curve parameters for 323 galactic Cepheids by Fourier analysis. The first systematic use of Fourier technique was made by Simon (1979) for analyzing the observed light variations and radial velocity variation of AI Velorum. The first-order amplitudes and phases from the Fourier fits were then compared with those obtained from linear adiabatic pulsation models to obtain the mass of AI Vel. Simon & Lee (1981) made the first attempt to reconstruct the light curves of Cepheid variables using the Fourier decomposition and to describe the Hertzsprung progression in Cepheid light curves. The method has been applied extensively by various authors for light curve reconstruction, mode discrimination and classification of pulsating stars (Antonello et al. 1986, Mantegazza & Poretti 1992, Hendry, Tanvir & Kanbur 1999, Poretti 2001, Ngeow et al. 2003, Moskalik & Poretti 2003, Jin et al. 2004, Tanvir et al. 2005). However, Fourier decomposition by itself is not perfectly suitable for classification of variable stars in large databases as the method works for individual stars, but can be used as a preprocessor for other automated schemes (Kanbur et al. 2002, Kanbur & Mariani 2004, Sarro et al. 2009).

The principal component analysis transforms the original data set of variables by way of an orthogonal transformation to a new set of uncorrelated variables or principal components. The technique amounts to a straightforward rotation from the original axes to the new ones and the the principal components are derived in decreasing order of importance (Singh et al. 1998). The first few components thus account for most of the variation in the original data (Chatfiled & Collins 1980, Murtagh & Heck 1987). The technique has been used for stellar spectral classification (Murtagh & Heck 1987, Storrie-Lombardi et al. 1994, Singh, Gulati & Gupta 1998), QSO spectra (Francis et al. 1992) and for galaxy spectra (Sodré & Cuevas 1994, Connolly et al. 1995, Lahav et al. 1996, Folkes, Lahav & Maddox 1996). There

⋆ E-mail: hpsingh@physics.du.ac.in

have been a number of studies on the use of PCA in analyzing Cepheid light curves (Kanbur et al. 2002) and RR Lyrae light curves (Kanbur & Mariani 2004). In both these studies, the input data to the PCA are the Fourier coefficients rather than the light curves themselves. Nevertheless, it was noted that the PCA was able to reproduce the light curves with about half the number of parameters (PCs) needed by the Fourier technique.

In this paper, we show for the first time, the use of PCA directly on the light curve data of more than 13,000 stars (RRab, RRc, fundamental and first overtone Cepheids) taken from the literature and different existing databases. We also use Fourier decomposition technique to these light curve data to determine the Fourier decomposition parameters. We use (i) Fourier coefficients (ii) Fourier coefficients as inputs to PCA and (iii) light curves (phase vs. magnitude) as input to PCA and compare relative performance of their ability to separate out the modes of pulsating stars, finding resonances in Cepheids and in classification of different types of variables. We have performed independent automated Fourier analysis of all the data sets described in the paper using a computer code developed by us. We have shown that PCA is a robust and powerful tool for the detection of changes in the structure of light curves and resonances. The PCA technique has also been used in the reconstruction of the light curves of Cepheids as well as RR Lyraes. Another aim of this paper is to analyze the performance of PCA as a fast, automated and unsupervised classification tool of variable stars. We present Fourier decomposition technique using Levenberg-Marquardt algorithm for non-linear least square fitting (Press et al. 1992) in section 2. We also describe the unit-lag auto-correlation function for finding out the optimal order of the fit and discuss the criteria for selection of stars for analysis. Section 3 describes the Principal component analysis for dimensionality reduction and light curve reconstruction. Section 4 describes the results obtained by the Fourier and PCA technique applied to RR Lyraes and Cepheids. Lastly in section 5, we present important conclusions of the study.

2 FOURIER DECOMPOSITION TECHNIQUE

Since the light curves of the variable stars are periodic, they can be written as a sum of cosine and sine series :

$$m(t) = A_0 + \sum_{i=1}^{N} a_i \cos(i\omega t) + \sum_{i=1}^{N} b_i \sin(i\omega t), \quad (1)$$

where $m(t)$ is the observed magnitude at time t, A_0 is the mean magnitude, a_i, b_i are the amplitude components of $(i-1)^{th}$ harmonic, ω=2π/P is the angular frequency, and N is the order of the fit. Obviously, equation (1) has $2N+1$ unknown parameters which require at least the same number of data points to solve for these parameters. Equivalently, we can write equation (1) as

$$m(t) = A_0 + \sum_{i=1}^{N} A_i \cos(i\omega t + \phi_i), \quad (2)$$

where $A_i = \sqrt{{a_i}^2 + {b_i}^2}$ and $\tan\phi_i = -b_i/a_i$. Since period is known, the observation time can be folded into phase (Φ) as (cf. Ngeow et al. 2003)

$$\Phi = \frac{(\mathrm{t} - \mathrm{t}_0)}{\mathrm{P}} - \mathrm{Int}\left(\frac{\mathrm{t} - \mathrm{t}_0}{\mathrm{P}}\right),$$

where t_0 is the epoch of maximum light. The value of Φ is from 0 to 1, corresponding to a full cycle of pulsation. Hence, equations (1) and (2) can be written as (Schaltenbrand & Tammann 1971)

$$m(t) = A_0 + \sum_{i=1}^{N} a_i \cos(2\pi i\Phi(t)) + \sum_{i=1}^{N} b_i \sin(2\pi i\Phi(t)), \quad (3)$$

$$m(t) = A_0 + \sum_{i=1}^{N} A_i cos[2\pi i\Phi(t) + \phi_i], \quad (4)$$

with relative Fourier parameters as

$$R_{i1} = \frac{A_i}{A_1};\ \phi_{i1} = \phi_i - i\phi_1$$

where $i > 1$. The combination of coefficients R_{i1}, ϕ_{i1} where $i = 2, 3, 4...$ can be used to describe the progression of light curve shape in case of Cepheids, RR Lyraes and other variables and can be used for variable star classification. In Table 1 we list the variable star data that has been subjected to the analysis.

The estimation of optimal number of terms to be used in the Fourier decomposition of the individual light curve is not straight forward. As has been pointed out by Petersen (1986), if N is chosen too small, a larger number of Fourier parameters can be calculated from a given observation and the resulting parameters will have systematic deviations from the best estimate. On the other hand, if N is chosen too large, we are fitting the noise. Following Baart (1982), Petersen (1986) adopted the calculation of unit-lag auto-correlation of the sequence of the residuals in order to decide the right N so that the residuals consist of noise only. It as defined as

$$\rho := \frac{\sum_{j=1}^{n} (v_j - \bar{v})\,(v_{j+1} - \bar{v})}{\sum_{j=1}^{n} (v_j - \bar{v})^2}$$

, where v_j is the j^{th} residual, $\bar{v}$ is the average of the residuals and $j = 1, ...n$ are the number of data points of a light curve. The value of v is basically the residuals of the fitted light curve

$$v = m(t) - [A_0 + \sum_{i=1}^{N} A_i cos(2\pi i\Phi(t)\ + \phi_i)]$$

. It should be noted that for the calculation of ρ we must choose the ordering of v_j given by increasing phase values rather than ordering given by the original sequence. A definite trend in the residuals will result in a value of ρ equal to 1, while uncorrelated residuals give smaller values of ρ. In the idealized case of residuals of equal magnitude with alternating sign, ρ will be approximately equal to -1. The suitable value of ρ can be chosen using the Baart condition. According to this, a value of $\rho \geqslant [n-1]^{-1/2}$ (where n is the number of observations) is an indication that it is likely that a trend is present, whereas a value of $\rho \leqslant [2(n-1)]^{-1/2}$ indicates that it is unlikely that a trend is present. Baart therefore used the following auto-correlation cut-off tolerance

$$\rho_c = \rho(cut) = [2(n-1)]^{-1/2} \quad (5)$$

Table 1. Data selected for the present analysis

Data	References	No. of stars	No. of good data
I. RR Lyrae			
A. I band (SMC RRab)	Soszyński, I. et al. (2002)	479	478
B. I band (LMC RRab)	Soszyński, I. et al. (2003)	5835	5797
C. I band (LMC RRc)	Soszyński, I. et al. (2003)	1751	1751
II. Fundamental Cepheids			
A. I band (LMC)	Soszyński, I. et al.(2008)	1804	1697
B. I band (SMC)	Soszyński, I. et al.(2003)	1319	1255
C. V band (Galaxy)	Berdnikov (2008)	528	389
D. V band (Galaxy)	Moffett & Barnes (1984)	111	101
E. V band (LMC)	Martin et al. (1979)	56	6
F. V band (LMC+SMC)	Moffett et. al (1998)	22	19
III. Overtone Cepheids			
A. I band (LMC)	Soszyński, I. et al. (2008)	1228	1148
B. I band (SMC)	Udalski, A. et al. (1999)	828	800

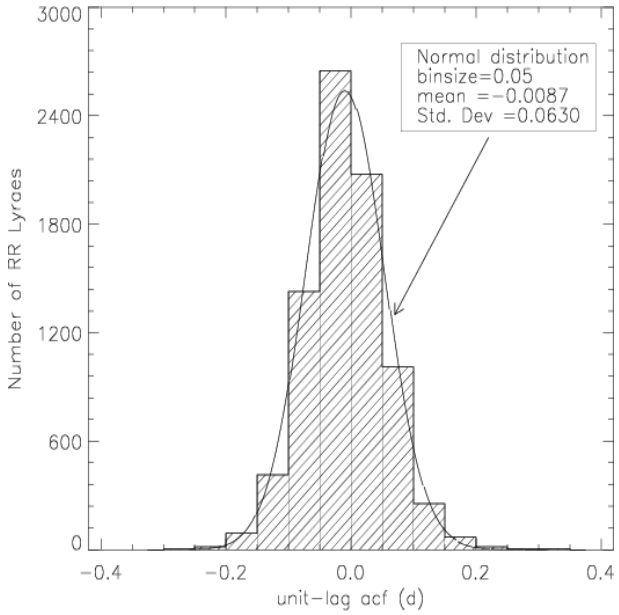

Figure 1. Unit-lag auto-correlation function of LMC and SMC RR Lyrae stars. The solid continuous curve is the best-fit normal distribution.

While computing the Fourier parameters of the light curve data for the present analysis we have taken care of the fact that the Baart's condition is satisfied. A histogram plot of d satisfying the Baart's condition for the RR Lyrae data sets of the present analysis is shown in Fig 1. The order of the fit (N) is 4 for SMC RR Lyraes and 5 for LMC RR Lyraes in data set I.

All the data sets in Table I are finally fitted with the optimal order of the fit and the fitted light curves are used to derive the Fourier phase and amplitude parameters from the Fourier coefficients. The goodness of fit is based on the χ^2 minimization where a value of $\chi^2_\nu \leqslant 5$ was used to select the stars having good phase coverage and smoothly varying data points, where ν is the degree of freedom and is equal to the number of data points minus the number of parameters used to fit the data. Generally a $\chi^2_\nu \sim 1$ indicates good measurement of the fitted data points. However, we have noticed that the presence of an outlier sometimes in the data set increases the χ^2_ν value to some extent. Therefore, such light curves were passed through visual inspection and values of χ^2_ν checked. All light curves with $\chi^2_\nu \geqslant 5$ were rejected. The number of stars with good data are listed in Table I. Fig. 2 shows the examples of good and bad Cepheid light curves. Fig. 3 shows the fitted light curves of fundamental mode Cepheids. Although the phase coverage for examples of these long period Cepheids is poor, the fits are reasonably good.

Thus the Fourier decomposition of the data sets in Table 1 is done by the computer code we have developed and the Fourier decomposition parameters (a_i, b_i) have been computed based on the optimal order of the fit by the calculation of the unit-lag auto-correlation function. The data having sparse phase coverage and noise are rejected based on the high ${\chi^2}_\nu$ between the fitted and the original light curves. Figs. (4) & (5) show the histograms of ${\chi^2}_\nu$ for all the 8026 RR Lyrae stars and 5415 Cepheids respectively.

3 PRINCIPAL COMPONENT ANALYSIS

The principal component analysis transforms the original set of p variables by an orthogonal transformation to a new set of uncorrelated variables or principal components (PCs). It involves a simple rotation from the original axes to the new ones resulting in principal components in decreasing order of importance. The first few q components ($q \ll p$) contain most of the variation in the original data (Chatfield & Collins 1980, Murtagh & Heck 1987). This feature of the PCA has been used in astronomical data analysis primarily for the purpose of reducing the dimensionality of the data

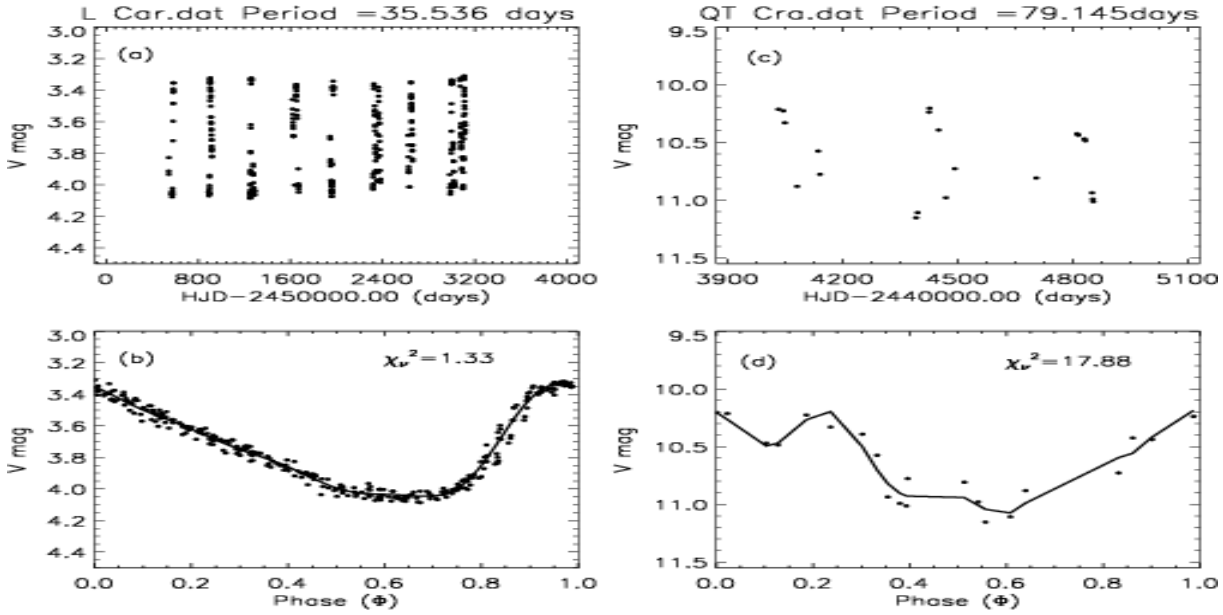

Figure 2. Examples of good (a) and bad (c) Cepheid data sets and their Fourier fitted light curves are shown in (b) and (d) respectively. Data sets such as (c) with $\chi_\nu{}^2 \geqslant 5.0$ have been rejected for analysis for both Cepheids and RR Lyrae variables.

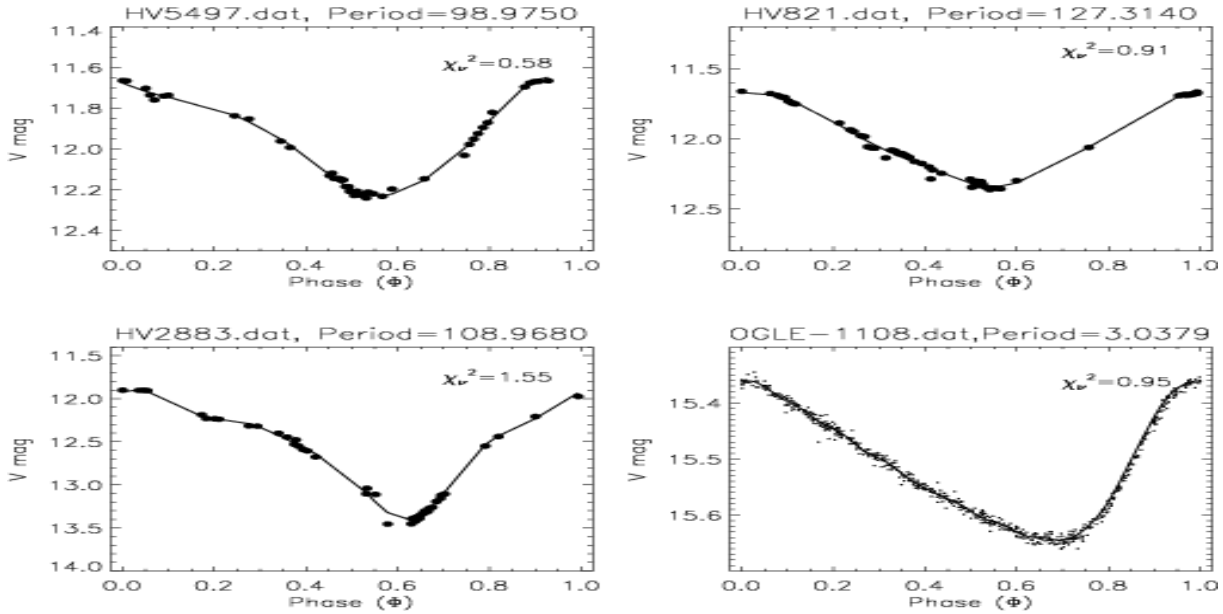

Figure 3. Fitted light curves for fundamental mode long period Cepheids from Moffett et al. (1998). Although the phase coverage is poor, the fits are reasonably good. The lower right panel shows the example of a short period fundamental mode Cepheid from OGLE-III database which has a good phase coverage.

and as a preprocessor for other automated techniques like Artificial Neural Networks (ANN). The application of PCA to the light curve analysis of variable stars has been limited to a few studies (Hendry et al. 1999, Kanbur et al. 2002, 2004, Tanvir et al. 2005). In the following, we briefly describe the transformation.

Let m_{ij} be the p magnitudes corresponding to n light curves. Let us define the $n \times p$ matrix by $\mathbf{X} = x_{ij}$,

$$x_{ij} = \frac{m_{ij} - \overline{m}}{s_j \sqrt{n}},$$

with

$$\overline{m} = \sum_{i=1}^{n} m_{ij},$$

and

$$s_j{}^2 = \frac{1}{n} \sum_{i=1}^{n} (m_{ij} - \overline{m_j})^2,$$

where $\overline{m_j}$ is the mean value and s_j is the standard deviation.

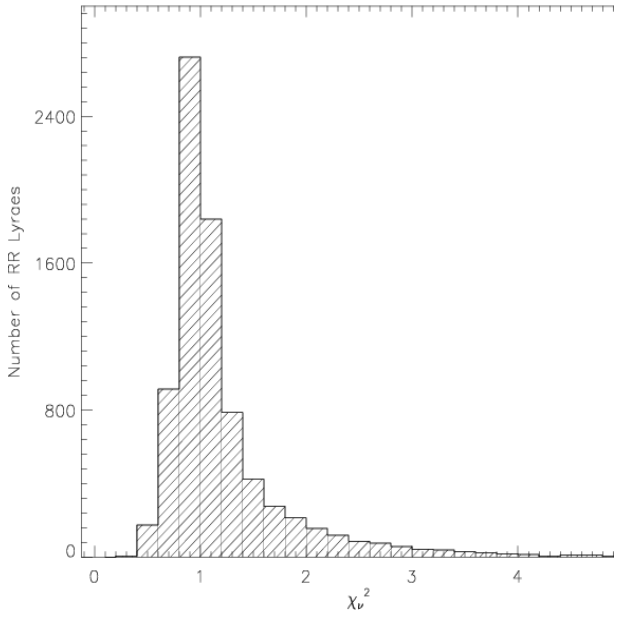

Figure 4. Histogram of $\chi_\nu{}^2$ for the 8026 RR Lyrae stars used in the analysis.

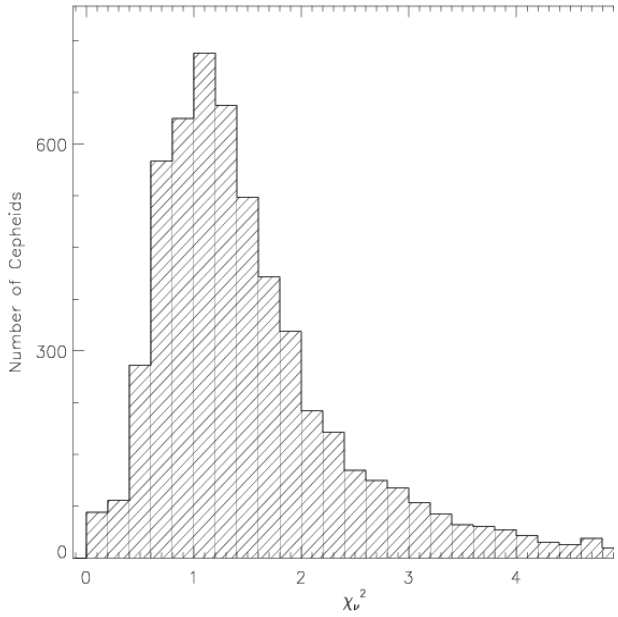

Figure 5. Histogram of the $\chi_\nu{}^2$ for the 5415 Cepheids (Data sets II and III) used in the analysis.

Using such standardization we find the principal components from the correlation matrix (cf. Murtagh & Heck 1987)

$$\mathrm{C_{jk}} = \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{x_{ij}x_{jk}} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i=1}}^{\mathrm{n}} (\mathrm{m_{ij}} - \overline{\mathrm{m_j}})(\mathrm{m_{ik}} - \overline{\mathrm{m_k}})/(\mathrm{s_j s_k}), \quad (6)$$

with the axis of maximum variance being the largest eigenvector $\mathbf{e}_1$ associated with the largest eigenvalue λ_1 of the equation

$$C\mathbf{e_1} = \lambda_1 \mathbf{e_1}. \quad (7)$$

The next (second) axes is to be orthogonal to the first and another solution of equation (7) gives the second largest eigenvalue λ_2 and the corresponding eigenvector or the principal component $\mathbf{e}_2$. Hence the proportion of the total variation accounted by the j^{th} component is λ_j/p, where p is also the sum of the eigenvalues (Singh et al. 1998).

Table 2. The first 4 eigenvectors, their percentage of variance and the cumulative percentages of variance of the 478 SMC RRab stars of the OGLE data base. The input matrix for the PCA is the Fourier coefficients (a_i, b_i, i=1,4) for the data set IA in Table I.

Component	Eigenvalue	Percentage	Cum. Percentage
1	5.76	71.94	71.94
2	0.94	11.70	83.64
3	0.68	8.49	92.13
4	0.24	2.97	95.10

Let us suppose that the first q principal components are sufficient to retain the information on the original p variables. Therefore, we now have a $(n \times q)$ matrix E_q of eigenvectors. The projection vector $\mathbf{Z}$ onto the q principal components can be found by

$$\mathbf{Z} = \mathbf{xE}_{\mathrm{p}}, \quad (8)$$

where $\mathbf{x}$ is vector of magnitudes defined by

$$\mathrm{x_{ij}\, s_j}\sqrt{\mathrm{n}} + \overline{\mathrm{m}} = \mathrm{m_{ij}},$$

and can be represented by

$$\mathbf{x} = \mathbf{Z}{\mathbf{E}_{\mathrm{p}}}^{\mathrm{T}}. \quad (9)$$

We obtain the final light curve x_{rec} by multiplying x with $\mathrm{s_j}\sqrt{\mathrm{n}}$ and adding the mean. Z is a $(n \times \mathrm{q})$ matrix and ${E_{\mathrm{p}}}^T$ is a $(q \times \mathrm{p})$ matrix and hence the reconstructed light curve is the original $(n \times \mathrm{p})$ matrix.

4 ANALYSIS OF LIGHT CURVES

Having determined the Fourier coefficients (a_i, b_i) for all the data given in Table I, one can apply the PCA procedure to various data sets separately by taking these Fourier coefficients as input parameters. The input matrix to the PCA thus contains (a_i, b_i) determined for different data sets. We have computed the PCs from the direct light curves as well. With the phase (Φ) as epoch for each light curve now available, we interpolate and obtain 100 values of the magnitude for phase 0 to 1 in steps of 0.01. Therefore, each light curve now consists of 100 data points as magnitude with the average magnitude subtracted from each data point.

The analysis of light curves of different classes of RR Lyrae and Cepheids is thus carried out by using the following three new processed sets of parameters:

(a) Fourier coefficients (a_i, b_i),
(b) PCs using (a_i, b_i) as input,
(c) PCs with 100 values of magnitude from phase 0 to 1 as input.

Reconstruction of some of the light curves in each of the above processed data sets are also done to show how PCs describe the light curve shape effectively and smoothly with only few components. As an example, light curves of

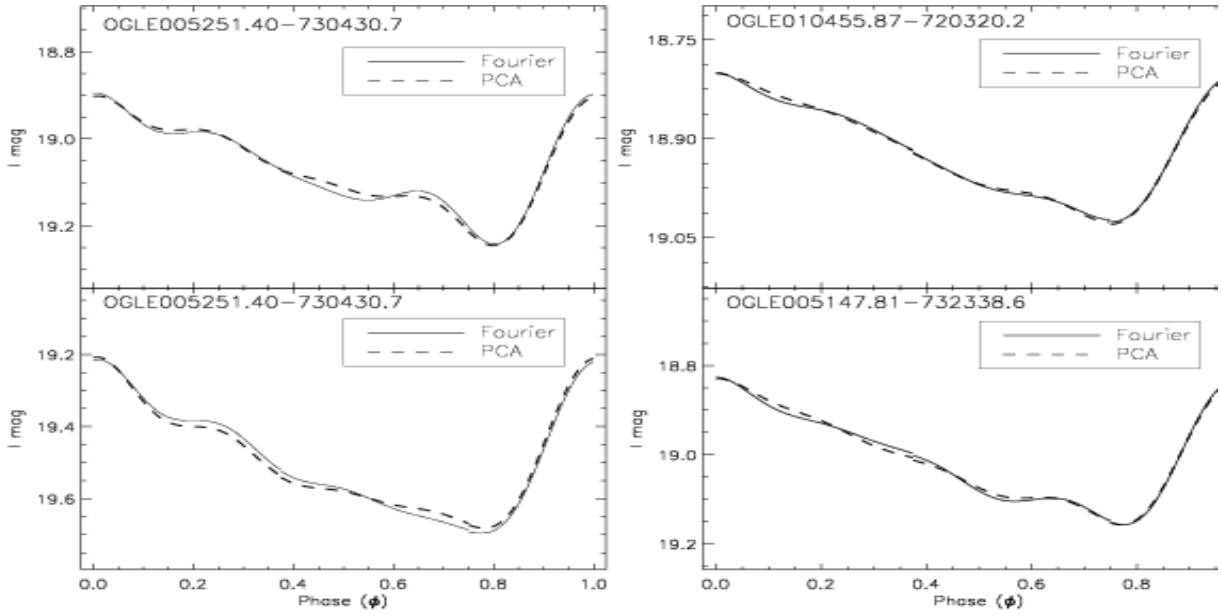

Figure 6. Reconstruction of the light curves of some RR Lyrae stars from data set IA using Fourier decomposition (solid lines) and PCA (dashed lines). The input matrix for the PCA is the Fourier coefficients (a_i, b_i, i=1,4) for the data set IA.

a majority of the 478 RRab stars from the OGLE data base (Soszyński et al. 2002) are completely described by the fourth order Fourier fit which has been justified by calculating the unit-lag auto-correlation function satisfying the Baart's condition. When we perform the PCA analysis on the (478×8) matrix of Fourier parameters, we find that 95 percent of the variance in the data is contained in the first four principal components (Table 2). Fig. 6 shows the reconstruction of some I band SMC RRab light curves using Fourier coefficients and the first four PCs. The input matrix for the PCA is the Fourier coefficients $(a_i, b_i, i = 1, 4)$. Therefore, nearly half of the parameters (PCs) are required as compared to the Fourier coefficients.

In Fig. 7 we have shown the reconstruction of LMC RR Lyrae light curves using the first 1, 5, 7 and 10 PCs. Table 3 shows the PCA results on an (7548×100) array of 7548 stars with 100 I band magnitude values between phase 0 and 1. More than 99 percent of the variance is contained in the first 7 PCs. Data compression ratio is of the order of 13 : 1.

Let us examine the relative merit of applying PCA on the Fourier coefficients and on the light curves directly. In Fig. 8, we have plotted the first principal component computed using 8 Fourier coefficients against the Period and the amplitude of 478 SMC RRab stars of data set IA (Table 1). We can easily see that the first principal component is correlated with the amplitude as well as the period. It is evident that the first PC is a good indicator of the amplitude of the RR Lyrae stars although there is a spread in PC1 at lower amplitudes (Fig. 8a). In Fig. 9a,b we see a similar trend when the PC1 is plotted against period and Amplitude of 7548 LMC RRab and RRc stars. The log (Period) vs. PC1 (first principal component) plot in Fig. 9a shows how well the short period RRc variables are separated from the RRab variables. However, PCA applied to Fourier coefficients is unable to separate the amlitudes of these two classes of variable stars (Fig. 9b). In Fig. 9c, we have plotted the amplitude vs. PC1 where PCs have been computed from the direct light curves instead of through Fourier coefficients. In this case, the amplitudes appear to be better separated for RRab and RRc stars, although there is an overlap at lower amplitudes. Fig. 10 shows the Period against the amplitude and the first three PCs. Again both the groups of stars are well separated by the first three PCs. In the subsequent analysis, we consolidate the fact that PCA applied on the direct light curve of variable stars can act as a robust and powerful classification tool for variable stars as well as extremely useful for structural analysis of their light curves.

Table 3. The first 11 eigenvectors, their percentage of variance and the cumulative percentages of variance of the 7548 RR Lyrae stars in LMC of the OGLE data base. The input matrix is an (7548×100) array .

Component	Eigenvalue	Percentage	Cum. Percentage
1	60.27	60.27	60.27
2	26.63	26.63	86.90
3	4.64	4.64	91.54
4	2.20	2.20	93.74
5	1.41	1.41	95.15
6	1.25	1.25	96.40
7	1.09	1.09	97.49
8	0.84	0.84	98.33
9	0.76	0.76	99.09
10	0.56	0.56	99.65
11	0.29	0.29	99.94

4.1 Structural Analysis & Classification

We use the light curve data for 3467 fundamental mode classical Cepheids from various sources as mentioned in Table 1 (data set II). Majority of the data used in the analysis are from OGLE database. The Fourier decomposition of all

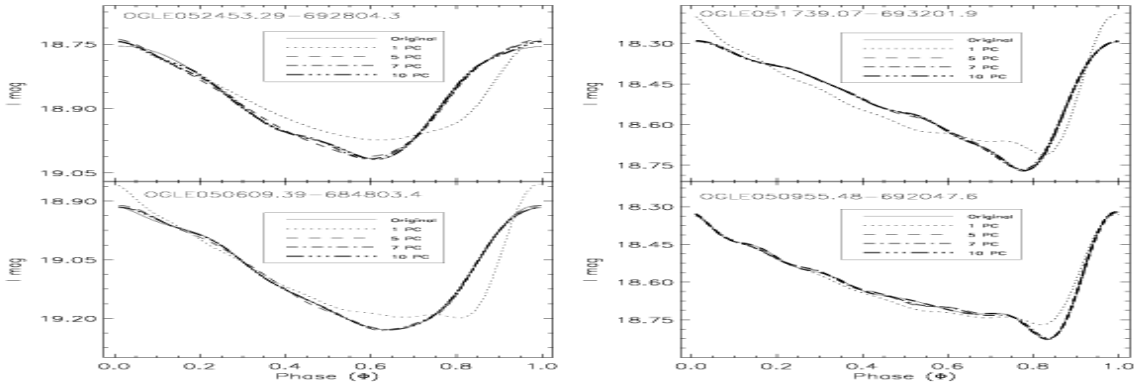

Figure 7. Reconstruction of RR Lyrae light curves using the first 1, 5, 7 and 10 principal components. The input matrix is an array of 7548 rows (stars) and 100 columns (magnitudes from phase 0 to 1). The sum of Eigenvalues is 100, the number of attributes (I band magnitudes).

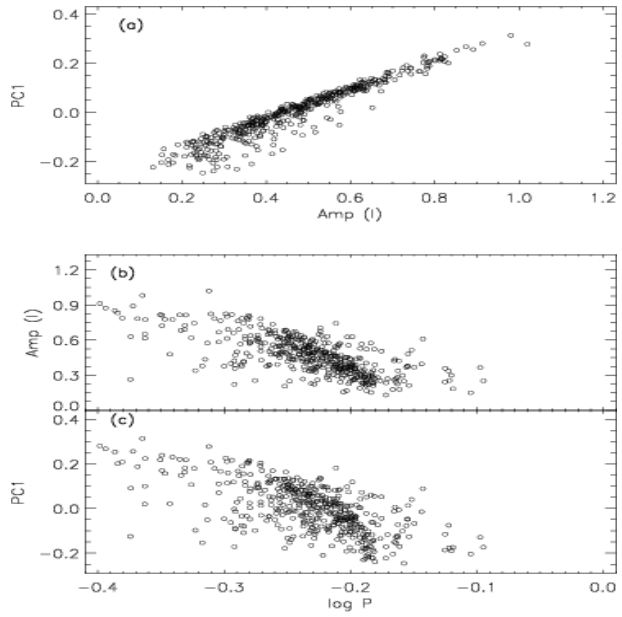

Figure 8. (a)First principal component (PC1) against Amplitude, (b) and (c) show the Amplitude and PC1 plotted as a function of log (period) in days for SMC RRab stars respectively. The input matrix for the PCA is the Fourier coefficients (a_i, b_i, i=1,4) for the data set IA.

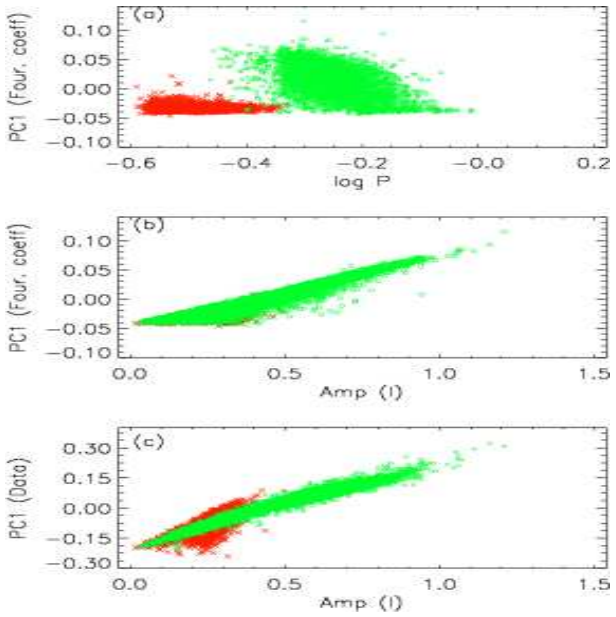

Figure 9. (a) PC1 versus log (Period) for LMC RR Lyrae stars (Data set IB and IC). The input matrix for the PCA is the Fourier coefficients (a_i, b_i, i=1,5). Open circles are LMC RRab stars and crosses are LMC RRc stars which are well separated. (b) PC1 versus Amplitude for the data set as in (a). Low amplitude RRc stars are not separated. (c) PC1 versus Amplitude. The input matrix is an array of 7548 rows (stars) and 100 columns (magnitudes from phase 0 to 1).

the 3467 Cepheid light curves has been independently done by us for the calculation of the Fourier decomposition parameters as described in Sec.2. We have seen that all the Cepheid light curves selected in the present study give satisfactory light curve shape with no numerical bumps or wiggles when reconstructed using the Fourier parameters. PCA is performed on an input matrix consisting of a 3467 × 100 array corresponding to 100 magnitudes from phase 0 to 1 for 3467 fundamental mode Cepheids. The result of the PCA output is shown in Table 4. We see that first six PCs are able to explain nearly 99 percent of the variance in the data. We do not use the Fourier parameters for the PCA analysis.

Kanbur et al. (2002) have tried to explain the the res-

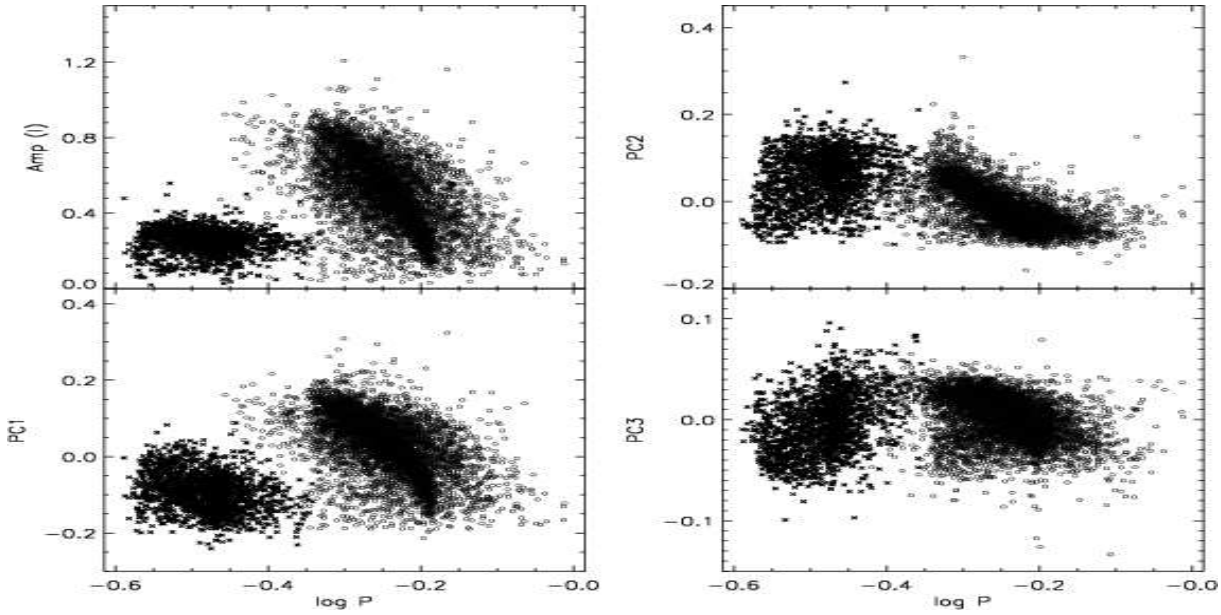

Figure 10. Amplitude and First three principal components versus log (Period) for LMC RR Lyrae stars (Data set IB and IC). Open circles are LMC RRab stars and crosses are LMC RRc stars which are well separated. The input matrix is an array of 7548 rows (stars) and 100 columns (magnitudes from phase 0 to 1).

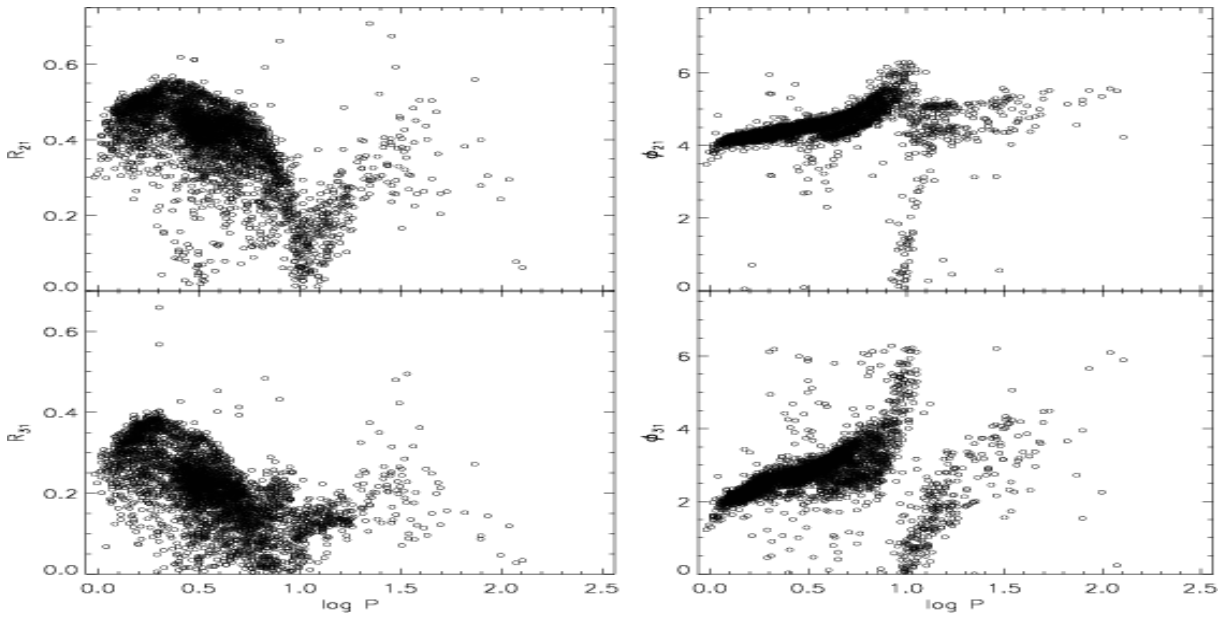

Figure 11. Fourier parameters R_{21}, R_{31}, ϕ_{21}, ϕ_{31} as a function of log (Period) for the 3467 fundamental mode Cepheids (Data set II, Table 1). The changes in the values of the parameters around the period log P = 1.0, 1.5, 2.10 are quite evident.

onances using the PCA. But due to the relatively less number of data points they did not give any definite conclusions about some of the resonances suggested by Antonello & Morelli (1996) in the period range 1.38< log P <1.43. By doing the PCA analysis of same data as used by Antonello & Morelli (1996), Kanbur et al. (2002) could not find any feature in that period range. They also put an open possibility of the presence of this feature in the higher order principal component plot. Based on the available light curves covering a wide range of periods, we have plotted R_{21}, R_{31}, ϕ_{21}, ϕ_{31} versus log P in Fig. 11. It is very evident from the plots that there is a definite structural change in the Fourier coefficients at periods log P $\sim$ 1.0 and 1.5 the second being close to the period range 1.38 < logP< 1.43 suggested by Antonello & Morelli (1996). We see that the Fourier decomposition parameters R_{21} and R_{31} decrease till log P $\sim$ 1.0, increase thereafter till log P $\sim$ 1.5 and after that R_{21} falls gradually again till $\log P \sim 2.10$. Similarly in the ϕ_{21} and ϕ_{31} plane, we see a sharp discontinuity around log P $\sim$ 1.

In Fig. 12 we plot the first eight PCs against log P. For PC1 to PC4 plots, a discontinuity around log P = 1.0 is quite visible. PC3 and PC4 clearly show a change around the pe-

Table 4. The first 8 eigenvectors, their percentage of variance and the cumulative percentage of variance of 3467 fundamental mode Cepheids. The input matrix is an 3467 × 100 array.

Component	Eigenvalue	Percentage	Cum. Percentage
1	38.71	38.71	38.71
2	32.01	32.01	70.72
3	15.64	15.64	86.36
4	7.40	7.40	93.76
5	3.45	3.45	97.21
6	1.60	1.60	98.81
7	0.69	0.69	99.50
8	0.17	0.17	99.67

riods log P $\sim$ 1.5 and also around log P $\sim$ 2.0. Thus, after independent PCA analysis and using a larger domain of the data we have find that in fact there are structural changes around log P $\sim$ 1 and 1.5 and hence there exist resonances around these periods. Kanbur et al. (2002), using the PCA analysis on the Fourier coefficients, did not find any resonances in the period range $1.38 < \log P < 1.43$. They used only the first two PCs and suggested that the resonances may be visible in the higher-order PCs. We have shown that the higher order PCs indeed show a resonance in the period around log P = 1.5. It is difficult to pinpoint the exact location of the resonance because of the spread in the parameter space. Model calculations are necessary to confirm the existence of these resonances. Antonello & Poretti (1996) also used a number of data points of the longer period side and found some evidence of a decrease of R_{21} at longer periods around (log P $\sim$ 2). It is difficult to confirm the existence of such a resonance although we see some change in trend in higher order PCs around this period.

It is a known fact that the first overtone Cepheids in the Galaxy, LMC and SMC show resonances at 3.2d, 2.7d and 2.2d respectively. However by analyzing the Fourier coefficients of the 1228 first overtone LMC Cepheids in the OGLE III database Soszyński et al. 2008 have found that this behaviour appears twice for the first overtone pulsators at period $\sim$ 0.35d and $\sim$ 3.0d. The short-period discontinuity at 0.35d can be explained by presence of the 2:1 resonance between the first and fifth overtones in stars with masses of about 2.5 $M_{\odot}$ (Dziembowski & Smolec 2009) while the second feature around 3.0d is interpreted as the signature of 2:1 resonance between the first and fourth overtones (Antonello & Poretti 1986).

In Fig. 13, we plot the Fourier parameters R_{21}, ϕ_{21}, R_{31}, ϕ_{31} versus P for 1148 LMC overtone Cepheids (data IIIA in Table I). The optimal order of the fit to the Fourier method has been found to be 12. There is a definitive marked structure of discontinuity in the Fourier plots at periods around 0.35 and 2.7 days. We now try to find out whether our PCA procedure can extract the information about these discontinuities, i.e, whether it is able to separate out the resonances. We carry out the PCA on a 1148 × 100 matrix of 1148 LMC Cepheids with 100 I band magnitudes corresponding to phase 0 to 1 in steps of 0,.01. Fig. 14 shows the the first six PCs versus the period. A sharp discontinuity around the shorter period end near 0.35 day is evident in all the PC plots while PC2 and PC3 show the change in the light curve shape of the LMC overtone Cepheids around a period of 2.7 days.

In Fig. 15 we have plotted the Fourier parameters R_{21}, R_{31} and ϕ_{21}, ϕ_{31} for 800 first overtone SMC Cepheids (Data IIIB, Table 1). It is quite evident from the R_{21} vs. P plot that there is marked discontinuity around the period $\sim$ 2.2d. The same plot leads one to suspect that there could be a short period resonance similar to LMC Cepheids. However, the subsequent PCA does not confirm such a short period resonance for the SMC Cepheids, probably because of the lack of data for the shorter periods. The PCA of the SMC overtone Cepheid data matrix (800 × 100) is shown in Fig. 16. The higher order PCs especially PC2 and PC4 show the resonance around a period 2.2 days.

Lastly we explore the possibility of classification of different classes of variable stars on the basis of the PCA analysis. Recently, Debosscher et al. (2007) reviewed the problem of scientific analysis of variable objects and proposed several methods to classify new objects base on the photometric time series data. Sarro et al. (2009) have done automated supervised classification of variable stars to the OGLE database. Having calculated the Fourier coefficients of all the 10871 stars that we have selected (Table 1) which includes LMC and SMC RRab stars, LMC RRc stars and the LMC fundamental and first overtone Cepheids, we plot the coefficient R_{21} vs. log P in Fig. 17. The Fourier parameter is able to cluster the stars into different regions of the R_{21} - P space although there is some overlap between the short period overtone Cepheids and the RRab variables. The LMC and SMC RRab stars are also indistinguishable. The principal component analysis has been performed on the 10871 × 100 array and the plot of PC1 - log P space shows that the first PC is able to separate RRab, RRc, fundamental and first overtone Cepheids. We hope to add more samples of some other classes of variable classes to carry out the analysis in a subsequent study.

5 CONCLUSIONS

Fourier decomposition is a trusted and much applied technique for analyzing the behaviour of light curves of periodic variable stars. It is well suited for studying individual light curves as the Fourier parameters can be easily determined. However, when the purpose is to tag a large number of stars for their variable class using photometric data from large surveys, the technique becomes slow and cumbersome and each light curve has to be fitted individually and then analyzed. Same is true if the aim is to look for resonances in the light curves in an automated way for a large class of pulsators. It is, therefore, desirable to look for methods that are reliable, automated and unsupervised and can be applied to the available light curve data directly.

Some attempts have been made in the recent past to use the well known Principal Component Analysis for the light curve analysis, but the major drawback of these studies was that they required the calculation of the Fourier parameters which then went as input to the PCA. This meant that the PCA, which was supposed to replace the Fourier decomposition, in fact relied on it.

In this paper, we have presented a new way of calculating the Principal Components which do not require the

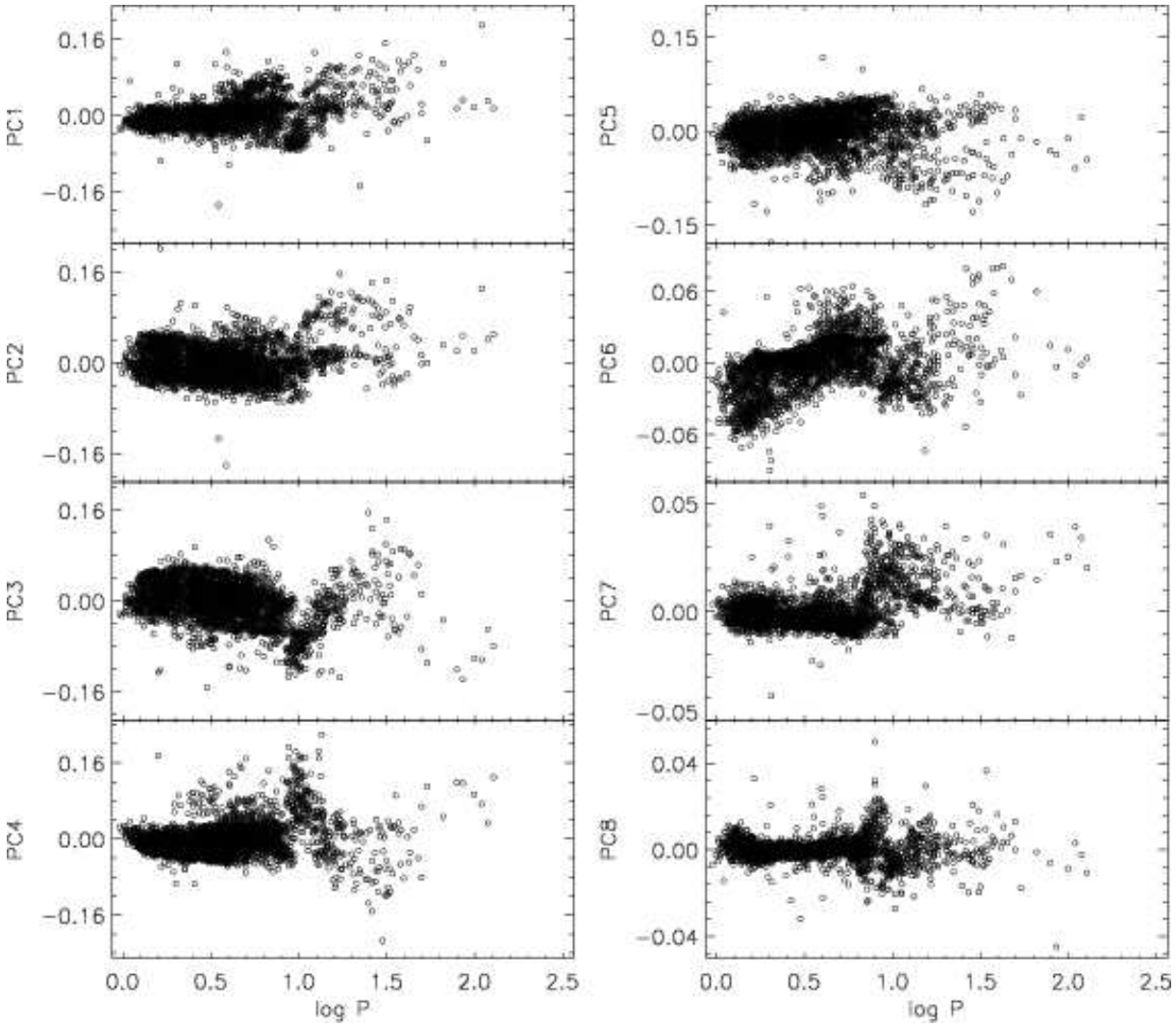

Figure 12. First eight PCs versus log(period) for 3467 fundamental mode Cepheids (Data set II). Higher order coefficients PC3, PC4 and PC7, PC8 clearly show the discontinuities around the period log P = 1.0 and 1.5. The input matrix is an array of 3467 rows (stars) and 100 columns (magnitudes from phase 0 to 1).

pre-calculation of Fourier coefficients. The input matrix of Fourier components was thus replaced by the original light curve data in the form of magnitude and phase. The technique has been first tested on a large database of RR Lyrae stars. It has been found that around 5 PCs are sufficient in most cases to reproduce the light curves. The results are compared with those obtained by the Fourier coefficients alone as well with those obtained by a PCA with Fourier coefficients as inputs.

Having tested our PCA technique, we applied it to study the structure of light curves of fundamental and first overtone Cepheids. By choosing a large data set of a large range of periods we have shown that the structure of the fundamental mode Cepheid light curves shows significant changes around the periods $\log P \sim 1$ and 1.5. The resonance around the period $\log P \sim 1$ is well known. Higher order PCs also show that the behavior of the light curves changes around the period $\log P \sim 1.5$ which is close to the resonance suggested by Antonello & Poretti (1996) in the period range $1.38 < \log P < 1.43$. There is some evidence of the structural change in the light curve shape around the period $\log P \sim 2.0$ also but this can be confirmed only when longer period data become available.

For the first overtone LMC Cepheids, we find a discontinuity at a shorter period of ~ 0.35d. Most of the PCs also show a clear trend of structural changes of the first overtone Cepheids at this short period. We have been able to find this feature because of the availability of significant number of light curves towards the shorter period end of the LMC Cepheids in the OGLE database. Our technique can easily find similar resonances in the Galactic and SMC first overtone Cepheids as and when there is substantial data available for the short period objects of this class.

Data compression ratio using PCA on the direct light curve data is enormous, a fact that has great relevance when dealing with large databases of the future. Also, we have shown some preliminary results of variable star classification for an ensemble of 10871 stars. In a future paper, we will describe the application of the PCA technique with a

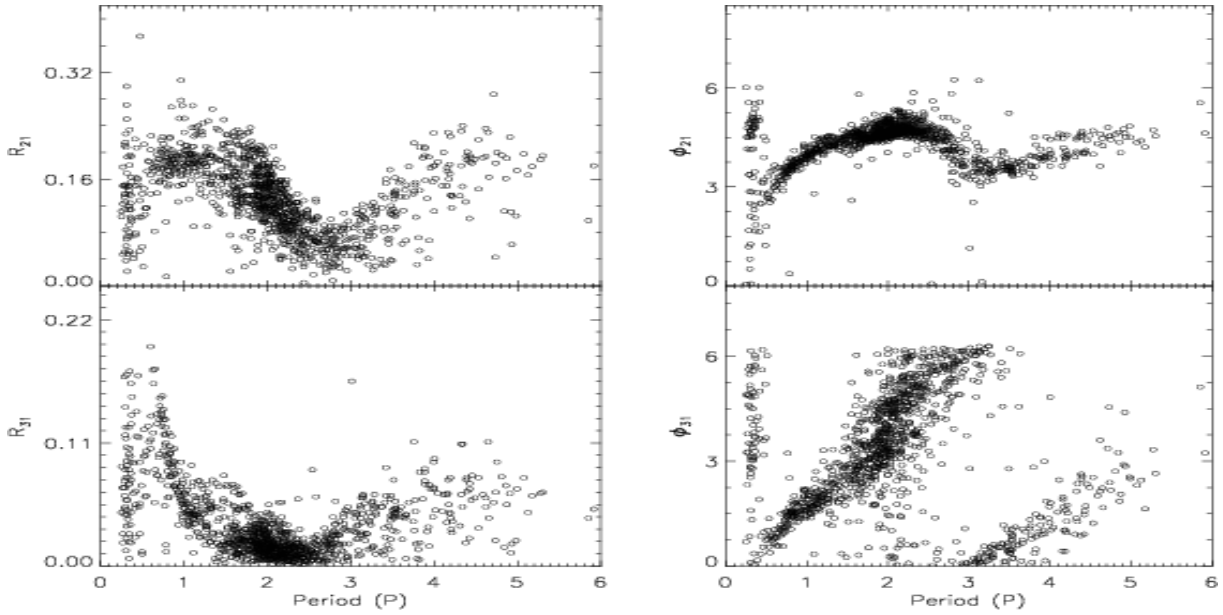

Figure 13. Fourier parameters R_{21}, R_{31}, ϕ_{21}, ϕ_{31} as a function of log (Period) for 1148 LMC overtone Cepheids (Data set IIIA). Sharp discontinuities around period P =0.35 and 2.7 are visible from the plots.

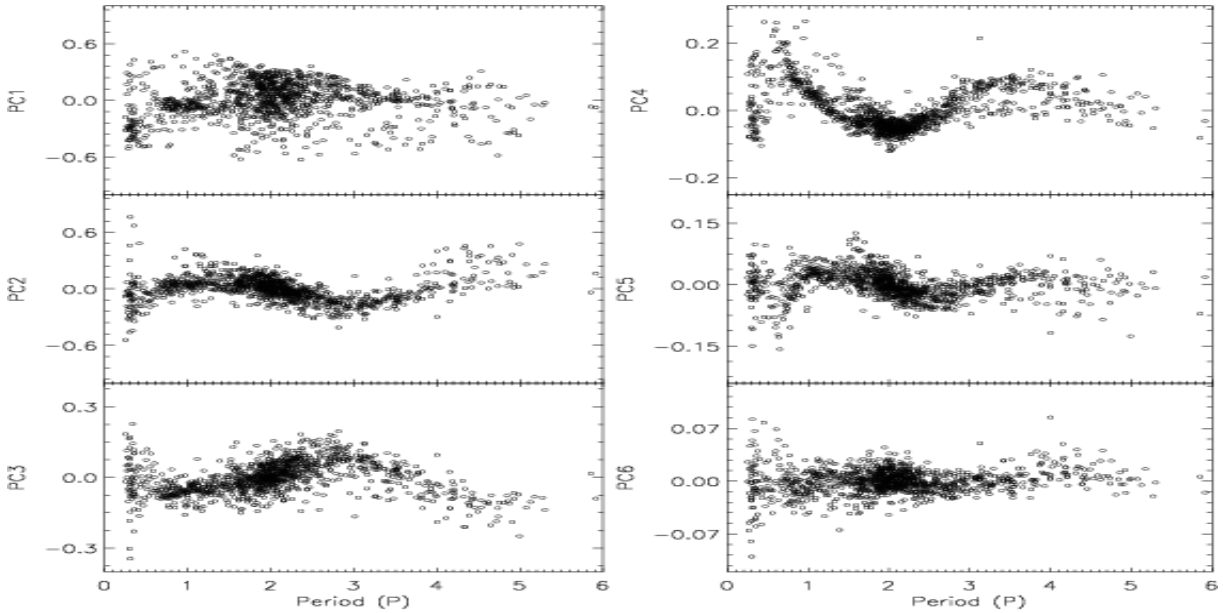

Figure 14. First Six PCs versus Period for LMC overtone Cepheids (Data set IIIA). The change in the light curve shape as shown in Fig. 13 are also seen from the PC plots. The input matrix is an array of 1148 rows (stars) and 100 columns (magnitudes from phase 0 to 1).

larger, more diverse database by looking at the classification accuracy and errors.

ACKNOWLEDGMENTS

SD thanks Council of Scientific & Industrial Research, India for a Senior Research Fellowship.

REFERENCES

Antonello, E., Poretti, E., 1986, A&A, 169, 149

Antonello, E., 1994, A&A, 282, 835

Antonello, E., Morelli, P. L., 1996, A&A, 314, 541

Baart, M. L., 1982, IMA J. Num. Analysis, 2, 241

Berdnikov, L. N., VizieR On-line Data Catalog: II/285. Originally published in: Sternberg Astronomical Institute, Moscow (2008)

Chatfield, C., Collins, A. J., 1980, Introduction to Multivariate Analysis. Chapman & Hall, London

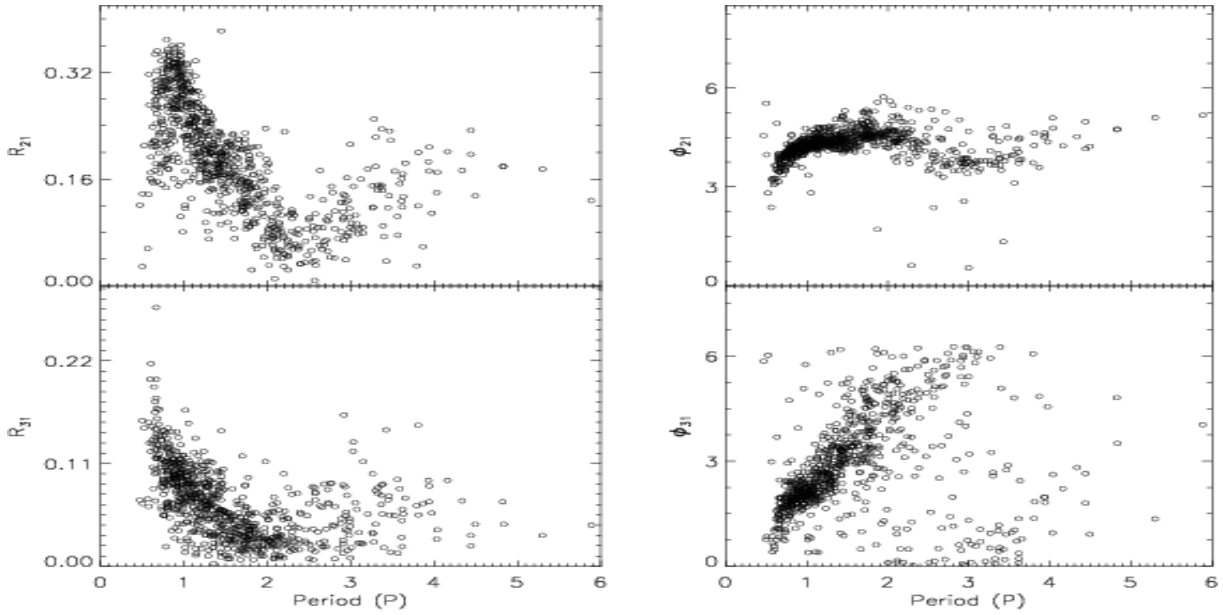

Figure 15. Fourier parameters R_{21}, R_{31}, ϕ_{21}, ϕ_{31} for 800 SMC overtone Cepheids (Data set IIIB). Sharp discontinuities around the period P = 2.2 d are visible from the R_{21} plot.

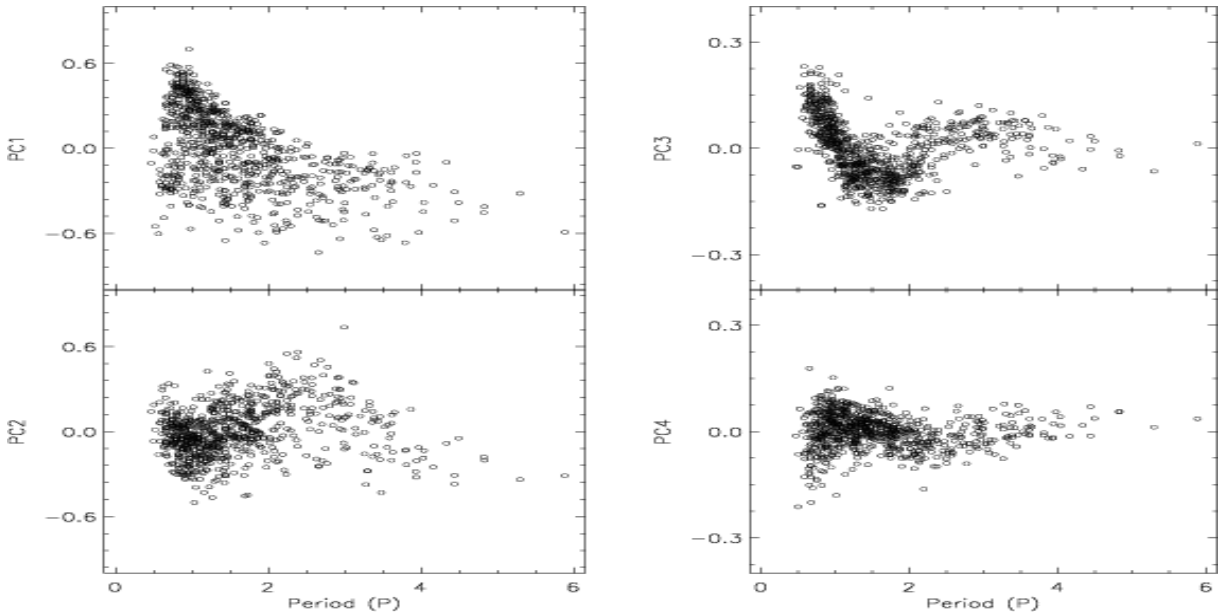

Figure 16. First four PCs for first overtone Cepheids of the SMC (data IIIB). Principal components PC2 and PC4 clearly show the discontinuity around period P = 2.2 d. The input matrix is an array of 800 rows (stars) and 100 columns (magnitudes from phase 0 to 1).

Connolly, A. J., Szalay, A. S., Bershady, M. A., Kinney, A. L., Calzetti, D., 1995, AJ, 110, 1071
Debosscher, J., Sarro, L. M., Aerts, C. et al. 2007, A&A, 475, 1159
Dziembowski, W. A., Smolec, R. 2009, Acta Astron. (in Press)
Folkes, S. R., Lahav, O., Maddox, S. J., 1996, MNRAS, 283, 651
Francis, P. J., Hewett, P. C., Foltz, C. B., Chaffee, F. H., 1992, ApJ, 398, 476
Hendry, M. A., Tanvir, N. R., Kanbur, S. M., 1999, ASP Conf. Ser., 167, Astron. Soc. Pac., San Francisco, p.192
Jin, H., Kim, S. -L., Lee, C. -U., et al. 2004, AJ, 126, 1847
Kanbur, S. M., Iono, D., Tanvir, N. R., Hendry, M. A., 2002, MNRAS, 329, 126
Kanbur, S. M., Mariani, H., 2004, MNRAS, 355,1361
Lahav, O., Naim, A., Sodré, L., Jr., Storrie-Lombardie, M. C., 1996, MNRAS, 283, 207
Mantegazza, L., Poretti, E., 1992, A&A, 261, 137
Martin, W. L., Warren, P. R., Feast, M. W., 1979, MNRAS, 188, 139
Moffett, T. J., Barnes, T. G. III., 1984, ApJS, 55, 389

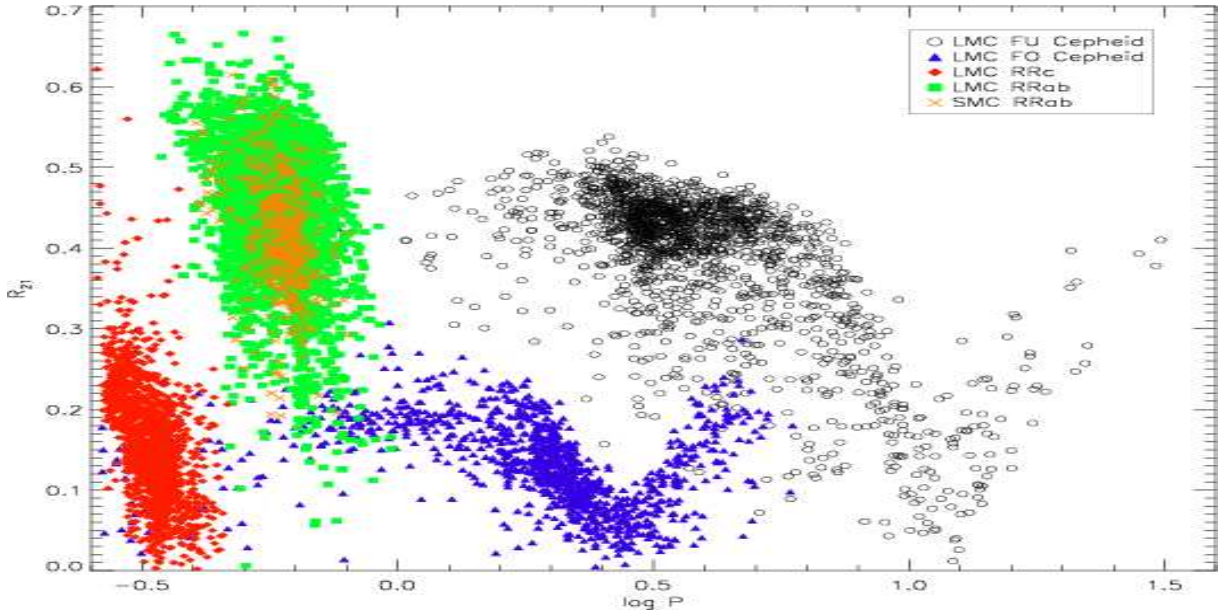

Figure 17. Classification of 10871 stars using Fourier coefficient R_{21} from the OGLE LMC & SMC data sets.

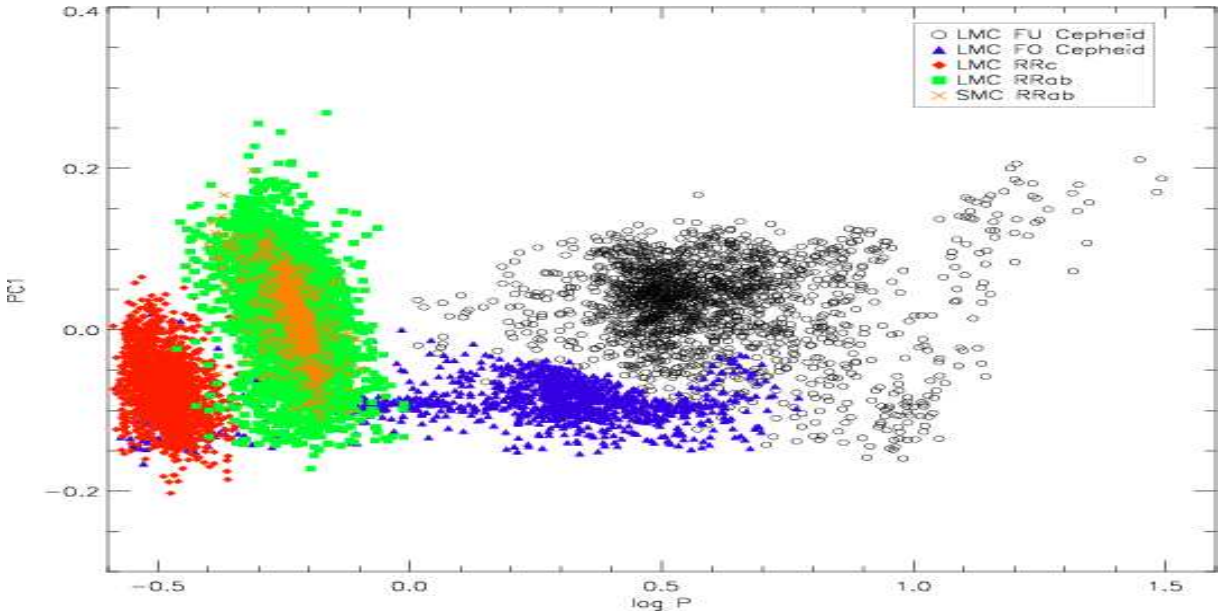

Figure 18. Classification based on PC1 of 10871 LMC & SMC RR Lyraes and Cepheids in the OGLE database.

Moffett, T.J., Gieren, W. P., Barnes, T.G. III, Gomez, M., 1998, ApJS, 117, 135

Moskalik, P., Poretti, E., 2003, A&A, 398, 213

Murtagh, F., Heck, A., 1987, Multivariate Data Analysis. Reidel, Dordrecht

Ngeow, C. -C, Kanbur, S. M., Nikolaev, S. et al., 2003, ApJ, 586, 959

Petersen, J. O., 1986, A&A, 170, 59

Poretti, E., 2001, A&A, 371, 986

Press, W., Teukolsky, S., Vetterling, W., & Flannery, B., 1992, Numerical Recipes in FORTRAN (2nd ed.) Cambridge Univ. Press

Sarro, L. M., Debosscher, J., López, M., and Aerts, C., 2009, A&A, 494, 739.

Schaltenbrand, R., Tammann, G. A., 1971, A&AS, 4, 265

Simon, N. R., 1979, A&A, 74, 30

Simon, N. R., Lee, A. S., 1981, ApJ, 248, 291

Singh, H. P., Gulati, R. K., Gupta, R., 1998, MNRAS, 295, 312

Sodré, L. Jr., Cuevas, H., 1994, Vistas Astron., 38, 287

Soszyński, I., Udalski, A., Szymanski, M. et al., 2002, 52, 369

Soszyński, I., Udalski, A., Szymanski, M. et al., 2003, Acta Astron., 53, 93

Soszyński, I., Poleski, R., Udalski, A., 2008, Acta Astron., 58, 163

Storrie-Lombardi, M. C., Irwin, M, J., von Hippel, T., Storrie-Lombardi, L. J., 1994, Vistas Astron., 38, 331
Tanvir, N. R., Hendry, M. A., Watkins, A. et al., 2005, 363, 749
Udalski, A., Soszy ński, I., Szyma ński, M. et al., 1999, Acta Astron., 49, 437
Udalski, A., Soszyński, I., Szymanski, M. K. et al., 2008, Acta Astron., 58, 89