

# Structural properties and classification of variable stars: A study through unsupervised machine learning techniques

Suman Paul,<sup>1</sup>\* Tanuka Chattopadhyay,<sup>1</sup>†

<sup>1</sup> Department of Applied Mathematics, University of Calcutta, 92 A.P.C. Road, Kolkata 700009

Accepted XXX. Received YYY; in original form ZZZ

## ABSTRACT

The advancement in the field of data science especially in machine learning along with vast databases of variable star projects like the Optical Gravitational Lensing Experiment (OGLE) encourages researchers to analyse as well as classify light curves of different variable stars automatically with efficiency. In the present work, we have demonstrated the relative performances of principal component analysis (PCA) and independent component analysis (ICA) applying to huge databases of OGLE variable star light curves after obtaining 1000 magnitudes between phase 0 to 1 with step length 0.001 for each light curves in identifying resonances for fundamental mode (FU) and first overtone (FO) Cepheids and in the classification of variable stars for Large Magellanic Cloud (LMC), Small Magellanic Cloud (SMC) as well as Milky Way (MW). We have seen that the performance of ICA is better for finding resonances for Cepheid variables as well as for accurately classifying large data sets of light curves than PCA. Using K-means clustering algorithm (CA) with respect to independent components (ICs), we have plotted period-luminosity diagrams and colour-magnitude diagrams separately for LMC, SMC and MW and found that ICA along with K-means CA is a very robust tool for classification as well as future prediction on the nature of light curves of variable stars.

**Key words:** stars: variables: general – methods: data analysis – Astronomical Data bases

## 1 INTRODUCTION

Over the past few decades, the importance of variable stars in the history of stellar evolution has been perceived and from the observed light curves of these variable stars, astronomers derived pieces of information not only to interpret their life histories but also that of galaxies. In the context of interpretation, analyzing the structural changes of different variable stars as well as their more accurate classification through different available statistical techniques is crucial. The available databases of variable star projects like MACHO (Massive Compact Halo Object), NSVS (Northern Sky Variability Survey), OGLE (Optical Gravitational Lensing Experiment), ASAS (All Sky Automated Survey), etc. together with satellite missions like CoRoT (Convection Rotation and Planetary Transits), Gaia, Kepler, etc. become primary resources as they contain a large amount of light curve data present in our Galaxy as well as Magellanic clouds.

Fourier decomposition (hereafter, FD) is a reliable, efficient and widely used technique in the study of stellar pulsation. Schaltenbrand & Tamman (1971), in their paper introduced this technique for the first time to derive UVB light curve parameters of 323 Galactic Cepheids from photoelectric observations. However, Simon & Lee (1981) developed this method to reconstruct light curves as well as to describe the Hertzsprung progression in Cepheid light curves. The Fourier method had been applied for different purposes like investigating a large sample of light curves of RR Lyrae field stars (Simon & Teays 1982), for analysis of the velocity curves of classical Cepheids (Simon & Teays 1983), for the exploration of the structural properties

of s-Cepheids (Antonello & Poretti 1986) as well as different kinds of Cepheid variables (Andreasen & Petersen 1987; Andreasen 1988; Simon & Kanbur 1995; Deb & Singh 2009), for the computation of the distance moduli of RR Lyrae stars (Kova'cs & Jurcsik 1997), for the determination of central period of Hertzsprung progression (Welch et al. 1997; Beaulieu 1998), for a detailed study of light curve properties of RRab (fundamental mode pulsators) and RRc (first overtone pulsators) (Poretti 2001; Moskalik & Poretti 2003), reconstructing Cepheid light curves to deal with possible metallicity dependence (Antonello et al. 2000; Ngeow et al. 2003). Some of the recent studies demonstrated the involvement of Fourier coefficients in the determination of physical parameters like luminosity, absolute magnitude, effective temperature, metallicity, etc. as a function of period, wavelength, etc. (Deb & Singh 2010; Nemec et al. 2011; Bhardwaj et al. 2015; Bhardwaj et al. 2017) as well as chemical and structural analysis for fundamental mode RR Lyrae stars of LMC (Deb & Singh 2014). Fourier decomposition technique is not perfectly adequate in the sense of enlargement of light curve databases of variable stars. However, some authors like (Hendry et al. 1999; Kanbur et al. 2002; Kanbur & Mariani 2004; Tanvir et al. 2005; Sarro et al. 2009) applied this method as a preprocessor for principal component analysis (PCA) scheme.

Principal component analysis (PCA) is an unsupervised machine learning method that reduces high dimensional data without losing much information and also helps to identify hidden trends in the data. This technique has been applied in different astronomical problems such as spectral classification of stars (Storrie-Lombardi et al. 1994; Singh et al. 1998), galaxies (Galaz & de Lapparent 1998), quasars (Yip et al. 2004), analysis of galaxy velocity curves (Kalinova et al.

\* E-mail: spappmath\_rs@caluniv.ac.in

† E-mail: tchatappmath@caluniv.ac.in

2017), characterization of Type Ia Supernova light curves (He et al. 2018), etc. As far as the variable stars are concerned, Hendry et al. (1999) found that Fourier decomposition encoded via PCA provides an efficient method to reconstruct noisy and sparse Cepheid light curves. Tanvir, Ferguson & Shanks (1999) applied PCA to statistically characterize V and I band of Hubble space telescope (HST) data for observed Cepheids in M96. In their paper, Kanbur et al. (2002) showed that PCA is more efficient than Fourier decomposition method at bringing out changes in light curve shape as a function of period. They also suggested that PCA will be more capable for searching resonances in observed Cepheid light as well as velocity curves. Leonard et al. (2003) successfully applied the PCA approach to estimate mean magnitudes for HST observed Cepheids in NGC 1637. To analyse the structures of the RRab light curves using PCA, Kanbur & Mariani (2004) found that the correlation between principal components was significantly smaller than the correlation between Fourier amplitudes, leading to a more accurate estimate of absolute magnitudes. Tanvir et al. (2005) applied the PCA approach to confirm that Cepheids light curve shapes are different for Milky Way and the Large and Small Magellanic Clouds at the same period range. However, in the studies like Leonard et al. (2003), Kanbur & Mariani (2004), Tanvir et al. (2005), the Fourier coefficients had been used as input for PCA analysis rather than the light curves themselves. In their paper, Deb & Singh (2009) used the original light curve data as input to the PCA since there is no advantage in the use of PCA where the input data be the Fourier coefficients. Meanwhile, they checked the relative performance of PCA as compared to the Fourier method for finding resonances in Cepheids and classification of different types of variable stars and concluded that the PCA technique could be applied to a larger and more diverse databases because of its efficiency and accuracy in a hierarchical classification. Recently, the PCA method has been applied in many studies including variable stars especially Cepheid variables at optical and near-infrared wavelengths (Bhardwaj et al. 2017) and RR Lyrae light curves (Hajdu et al. 2018).

Independent component analysis (ICA) is also an unsupervised machine learning algorithm similar to PCA, but the main focus is on the independence of components whereas PCA focuses on maximizing the variance of the data points. Recently, ICA method had been applied in many studies in modern astronomy such as source separation (Pires et al. 2006; Sheldon & Richards 2018), unsupervised classification (Das et al. 2015; Chattopadhyay et al. 2019), dimension reduction (Sarro et al. 2018), etc. In their paper, Deb & Singh (2009) demonstrated the ability of PCA over the Fourier decomposition method in the case of classifying stars into different variability classes and this happened may be due to the fact that the principal components (PCs) are uncorrelated. However, in the case of ICA, the components are independent. In this sense, ICA might have some advantage over PCA, that is the motivation behind this paper.

In this paper, we collect huge amount of available light curve data sets of variable stars (Cepheids, RR Lyrae stars, eclipsing binaries, Mira variables) mainly from OGLE-IV photometric databases for Large Magellanic Cloud (LMC), Small Magellanic Cloud (SMC) and Milky Way (MW) galaxies and use PCA and ICA directly to the data and compare the relative performances and also compare with Fourier method used in the above studies.

The main features of this paper are following:

- Structural analysis for Cepheids (FU, FO) classes of variables separately for LMC, SMC and MW using PCA and ICA both and comparison of the results with Fourier method mentioned in the literature.

- Classification of huge amount of data sets through PCA and ICA individually and to compare the results.
- Application of K-means clustering analysis on independent components (ICs) to plot the period-luminosity diagram as well as the colour-magnitude diagram (CMD).

The work has been arranged in the following manner. We present the data sets in Section 2. The methods have been described in Section 3. Results and discussions of structural analysis including classification and use of K-means Clustering have been narrated in Section 4. Finally, we conclude the study in Section 5.

## 2 DATA SET

The ultimate purpose of the Optical Gravitational Lensing Experiment (OGLE) is to search for dark matter using microlensing phenomena, introduced by Paczyński (1986). However, a wealth of information about variable stars has been collected for the Milky Way as well as nearby irregular galaxies, namely Large Magellanic Cloud (LMC) and Small Magellanic Cloud (SMC) as a by-product of this project. The error in the photometric measurements of a variety of variable stars has been less ( $\sim 0.006\text{--}0.2$  mag), making the OGLE database a rich resource for analysing the properties of these stars. In this paper, we use an enormous amount of light curve data of RR Lyrae (RRab and RRc), Cepheids (fundamental mode (FU), first overtone (FO), Type-II), eclipsing binaries (EB) and long-period variable star like Mira variables separately for LMC, SMC and Milky Way galaxies, collected from the publicly available online database of OGLE<sup>1</sup> project.

For Milky Way, we collected both the galactic bulge and the galactic disk light curve data separately and add them to use in this analysis. The description of the data set including references is included under Table 1.

## 3 METHODOLOGY AND STATISTICAL ANALYSES

### 3.1 Conversion of data sets into matrix form

In this paper, we have used enormous amount of I-band light curve data collected from recent OGLE surveys for different variable stars like RR Lyrae, Cepheids, Mira, Eclipsing binary, etc. in LMC, SMC, and Milky Way galaxies to analyse the structural patterns as well as classification of these stars. Since the input to PCA or ICA has to be a matrix, we first convert each of the data sets into matrix form. Each of the light curve data file mainly consists of Heliocentric Julian Days as observational times, observed I-band magnitudes and uncertainty of magnitudes. The conversion procedure is structured below.

- Convert the observation time ( $t$ ) to phase ( $\phi$ ) with the help of equation (1).

$$\phi = \frac{(t - t')}{P} - \text{int}\left[\frac{(t - t')}{P}\right] \quad (1)$$

(Ngeow et al. 2003; Deb & Singh 2009) where  $\phi$  ranges from 0 to 1, corresponds to one full cycle of pulsation and  $\text{int}$  refers to the integer part of the content.  $P$  and  $t'$  are respectively the period (in days) and epoch of the maximum light, taken from literature of the respective databases referred in Table 1.

<sup>1</sup> <https://ogle.astrouw.edu.pl/>

**Table 1.** Description of data sets collected from OGLE surveys

Data		LMC				SMC				Milky Way			
I-band	stars selected	References		Set	Stars selected	References		Set	Stars selected	References		Set	
RRab	28,858	Soszyński et al. (2016a)		1A	5,193	Soszyński et al. (2016a)		2A	19,568	Soszyński et al. (2011a)		3A	
RRc	9,903			1B	831			2B	7,231	Soszyński et al. (2019)		3B	
FU	2,428	Soszyński et al. (2015)		1C	2,739	Soszyński et al. (2015)		2C	1,184	Soszyński et al. (2020)		3C	
FO	1,766			1D	1,783			2D	542			3D	
Type-II	286	Soszyński et al. (2018)		1E	53	Soszyński et al. (2018)		2E	1,374	Soszyński et al. (2020)		3E	
EB	26,121	Graczyk et al. (2011)		1F	7,938	Pawlak et al. (2016)		2F	17,590	Soszyński et al. (2016b)		3F	
Mira	1,663	Soszyński et al. (2009)		1G	352	Soszyński et al. (2011b)		2G	6,528	Soszyński et al. (2013)		3G	

- Interpolate magnitudes of each light curve using spline interpolation method between phase 0 to 1 with step length 0.001 to obtain 1000 magnitudes. Mathematically, linear splines are fitted at n equally spaced phase intervals to give tabulated function

$$z_i = z(y_i), \quad i = 1, 2, 3, \dots, n_j \quad (2)$$

with j being the jth light curve. Therefore, for  $i=1,2,3,\dots,n_j$  we get  $(n_j - 1)$  linear functions of the form

$$f_i(y) = p_i z_i + q_i z_{i+1}, \quad y \in [y_i, y_{i+1}] \quad (3)$$

where  $f_i(y_i) = z_i$  and  $f_i(y_{i+1}) = z_{i+1}$ . The constants  $p_i$  and  $q_i$  satisfies

$$p_i = \frac{y_{i+1} - y}{y_{i+1} - y_i} \quad \text{and} \quad q_i = 1 - p_i = \frac{y - y_i}{y_{i+1} - y_i}, \quad (4)$$

$i=1,2,3,\dots,(n_j - 1)$ .

We repeat the process with newly formed interpolating points and old ones together until obtaining desired numbers. Thus interpolation error decreases by following the above procedure (Cassisi et al. 2012). Example of one of the interpolated light curves for variable stars has been demonstrated in Fig. 1.

- After interpolating all the light curves corresponding to a data set, we have normalized the magnitudes of each light curve between 0 and 1 using equation (5).

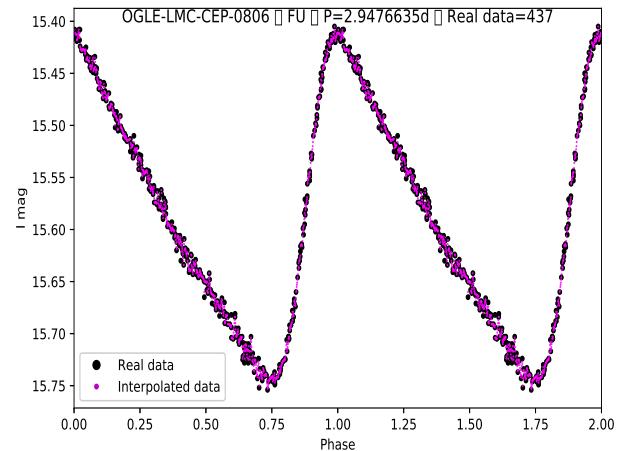
$$m_{norm} = \frac{m - m_{min}}{m_{max} - m_{min}} \quad (5)$$

- Finally, we organize the normalized magnitude data into the matrix form of order  $M \times 1000$ , where M is the number of light curves for further analysis. We have used a high-level programming language in Python for our analysis.

### 3.2 Principal component analysis (PCA)

In the present work, PCA has been used in two different ways. Firstly, to dig out the hidden structural changes from the variable star data and to classify between different variable stars with respect to their periods. In another case, PCA has been performed prior to ICA to obtain optimum number of ICs.

PCA is basically a way of visualizing hidden patterns of complex high dimensional (say k) data sets by transforming the original data sets orthogonally into new set of variables (principal components)



**Figure 1.** Spline interpolation of LMC FU light curve to obtain 1000 I-band magnitudes.

of fewer dimensions ( $r \ll k$ ) which basically retain most of the variation in the original data. The method has been described below in short.

Let us consider a matrix  $M_{k \times q}$  consisting of  $l_{ij}$ 's, be the q normalized magnitudes of k light curves. We perform standardization on the said matrix using the following equation

$$r_{ij} = \frac{l_{ij} - \bar{l}_j}{s_j \sqrt{k}} \quad (6)$$

where  $r_{ij}$ 's are the elements of standardized matrix  $T_{k \times q}$  and

$$\bar{l}_j = \frac{1}{k} \sum_{i=1}^k l_{ij} \quad (7)$$

$$s_j^2 = \frac{1}{k} \sum_{i=1}^k (l_{ij} - \bar{l}_j)^2 \quad (8)$$

are the mean value and standard deviation of any element of the matrix  $M_{k \times q}$ . The corresponding Correlation matrix D is given by

$$D_{jh} = \sum_{i=1}^k r_{ij} r_{ih} = \frac{1}{k} \sum_{i=1}^k (l_{ij} - \bar{l}_j)(l_{ih} - \bar{l}_h) / (s_j s_h) \quad (9)$$

The principal components are the eigen vectors ( $e_i$ ) corresponding to the eigenvalues ( $\lambda_i$ ) of the equation

$$De_i = \lambda_i e_i \quad (10)$$

The eigen vector associated with the largest eigenvalue corresponds to the axis of maximum variance and is called the first principal component. Similarly, the second principal component is the eigenvector associated with the second largest eigenvalue and it corresponds to the axis orthogonal to first and so on. The variance-ratio explained by the  $j^{th}$  eigenvalue is given by  $\lambda_j/S$ , where S be the sum of eigenvalues (Singh et al. 1998).

### 3.3 Independent component analysis (ICA)

Independent component analysis (ICA) is also a technique which reduces the dimension of a higher dimensional matrix as similar to PCA but the primary difference is that the components are mutually independent in the former case.

Mathematically, let,  $Y_1, Y_2, \dots, Y_k$  be k random vectors (here light curves) and q (here 1000) be the number of magnitudes of each  $Y_i$ , ( $i=1,2,3,\dots,k$ ).

Let,  $Z = BQ$ , where  $Q = [Q_1, Q_2, \dots, Q_k]'$  is a random vector consisting of hidden components  $Q_i$ , ( $i=1,2,3,\dots,k$ ) such that  $Q_i$ 's are mutually independent and B is non singular matrix. Then,

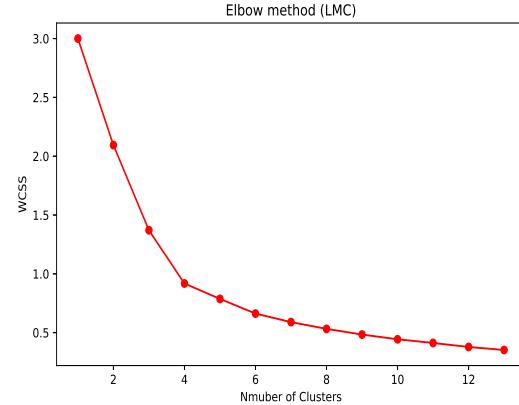
$$Q = B^{-1}Z = WZ \quad (11)$$

Where W is the unmixing matrix (Comon 1994; Chattopadhyay et al. 2013, and references therein). The main objective of ICA is to find Q with the help of W such that  $cov(g_1(Q_i), g_2(Q_j)) = 0$  for  $i \neq j$ , where  $g_1(Q_i), g_2(Q_j)$  are any two non linear functions.

Choice of optimum number of ICs is difficult (Kairov et al. 2017). In this work, we use PCA to choose optimum number of ICs. The optimum number of PCs are choosen in a way that the total variance obtained by PCs are above 85% and the choice of optimum number of ICs are based on choice of optimum number of PCs (Albazzaz and Wang 2004; Chattopadhyay et al. 2013; Chattopadhyay et al. 2019). Other choices of optimum number of ICs leads to unsatisfactory results as being found by multiple run of the algorithm. For PCA, PCs contained variance of the data set in a decreasing order i.e. PC1 contained maximum variance, next PC2 contained less variance than PC1 and so on, however in the case of ICA, there is no definite pattern for ICs for which the variance in the data set is maximum. As the variation among ICs has no order like PC components so by visualizing the plots for individual ICs corresponding to each data set, we have selected that component of ICA which has the maximum variation for comparison with PC1, which has maximum variation. So the independent component with maximum variation may be different for different data sets. For example, in Fig. 5, we plot PC1 against log P (in days) and IC3 against log P (in days) as PC1, IC3 both have maximum variance in their respective analysis. Here we have used scikit-learn machine learning library for Python programming for both PCA and ICA.

### 3.4 K-means cluster analysis

K-means cluster analysis (CA) is a multivariate unsupervised machine learning technique, basically applied for finding coherent



**Figure 2.** Representation of within cluster sum of squares (WCSS) against number of clusters to determine optimum value of K. For LMC, the optimum value of clusters (K) is 4.

groups in a data set (Chattopadhyay et al. 2007, 2010, 2012, 2019; Das et al. 2015). The main objective of this method is to find K clusters such that all the objects are distributed among those K clusters with the following two properties :

- Each cluster must have at least one object.
- Each object must be contained in exactly one cluster.

The algorithm of this method is described as below :

Step 1: All objects are randomly distributed among K (given) groups.

Step 2: A particular object is selected randomly from any of those groups at first. Then parametric distance (parameters are the number of ICs choosen for the analysis) is calculated between the selected object and the remaining objects. If the distance measured between objects in the group is greater than that for objects in other groups, they are alternated.

Step 3: The previously described procedure is applied for all objects and for all groups.

Step 4: Repeat steps 1 and 3 until no further change of objects between groups.

Since the number of clusters (K) is pivotal input to the above algorithm, it is crucial to find K accurately. Thus we opted for Elbow method described below.

#### 3.4.1 Determination of optimum value of K

The optimum number of clusters is selected by K-means clustering (MacQueen 1967) in association with the method proposed by Sugar & James (2003). In this method, distortion measure ( $d_k$ ) is calculated for  $K = 1, 2, 3, \dots$  etc with the help of equation (12),

$$d_k = (1/P) \min_x E[(x_k - c_k)'(x_k - c_k)], \quad (12)$$

where P is the total number of variables (here the number of ICs). In other words, it is the distance of  $x_k$  vector (values of parameters) from centriod  $c_k$  of the corresponding cluster. The optimum value of K is selected for which  $d_k$  (within cluster sum of squares, WCSS) first starts to diminish at the 'elbow' of the pattern. For SMC, LMC (see Fig. 2) and MW, we found  $K = 4$ .

## 4 RESULTS AND DISCUSSIONS

### 4.1 Identifying resonances for Cepheid light curves: a comparison between PCA and ICA

The usage of Cepheid variables are profound in various astrophysical investigations such as determination of Hubble's constant to a certain accuracy (Freedman & Madore 2010; Riess et al. 2018) as well as extragalactic distance scaling through period-luminosity (P-L) relation (Leavitt and Pickering 1912). Also, more specifically classical Cepheids are being utilized as the convenient touchstone for verification of stellar pulsation theories (Antonello 1994; Bono et al. 1999). Radial modes of pulsation i.e. periodic expansion and contraction in outer layers of Cepheids lead to the variations of stars luminosity. These inward-outward movements might be due to opacity driven mechanism. In the dim phase, due to a sufficiently hot environment helium become fully ionized which causes an increase in opacity of the entire layer and absorbs radiation outpouring from the core of the star, leading to the expansion of the star's radius. This expansion triggers cooling of the outer layer which helps to recombine the electrons of helium to produce single ionized helium, which reduces the opacity as well as the ability of absorption, causing a brighter appearance of the star as well as contraction of the layers and this contraction provokes the next expansion. Fundamental mode is the simplest radial mode in which the expansion/contraction happens simultaneously to the entire environment of the star whereas in the case of the first overtone, while the outer zone of the star is expanding, the inner zone shrinks and vice versa.

Systematic changes in the shape of classical Cepheid light curves arranged according to increasing pulsation period is known as Hertzsprung progression (HP) (Hertzsprung 1926). There is a secondary bump appears in the descending branch for light curves of short period ( $P < 10$  days) Cepheids and as the period increases the bump moves backward in phase and appears at the maximum light, called the centre for HP and finally the bump moves to the rising branch of long period ( $P > 10$  days) Cepheids. This phenomenon is associated with the sharp breaks found around 10 days in the progression of Fourier parameters against period (Simon & Lee 1981) and these could be the consequence of a resonance between the fundamental mode and second overtone Cepheids when the period ratio between the modes is  $\sim 0.5$  (Simon & Schmidt 1976; Simon 1977; Takeuti & Aikawa 1981; Aikawa 1984; Klapp et al. 1985). Since then the well known FD method had been used by many authors (Simon & Lee 1981; Moskalik et al. 1992; Welch et al. 1997; Beaulieu 1998) to identify the possible resonances in FO and FU Cepheids as well as determination of centre of HP by the means of Fourier parameters for different galaxies. By finding out the center of HP in LMC ( $Z = 0.008$ ), SMC ( $Z = 0.004$ ) and MW ( $Z = 0.02$ ), Beaulieu (1998) concluded that a decrease in metallicities shifts the HP center along longer periods. Bono et al. (2000) constructed nonlinear convective models by adopting a fixed chemical composition ( $Y = 0.25$ ,  $Z = 0.008$ ) to investigate bump Cepheids theoretically and showed that the periods of the HP center are in good agreement with observational values. They also suggested that the dependence of the HP behaviour on chemical compositions should be investigated with the help of detailed comparison between theoretical nonlinear Cepheid models and observational evidences. Thus reliable detection of resonances for Cepheids through various modern tools is crucial as they can be used to place constraints on stellar parameters such as luminosity, mass, metallicity etc. (Simon & Kanbur 1995). However, FD method is time consuming in case of large databases as each light curve has to be analysed separately. In recent years, PCA has been widely used as an alternative to FD method in identifying resonances along the

**Table 2.** First 3 eigenvalues, their explained variance and cumulative explained variance for 2428 FU Cepheids

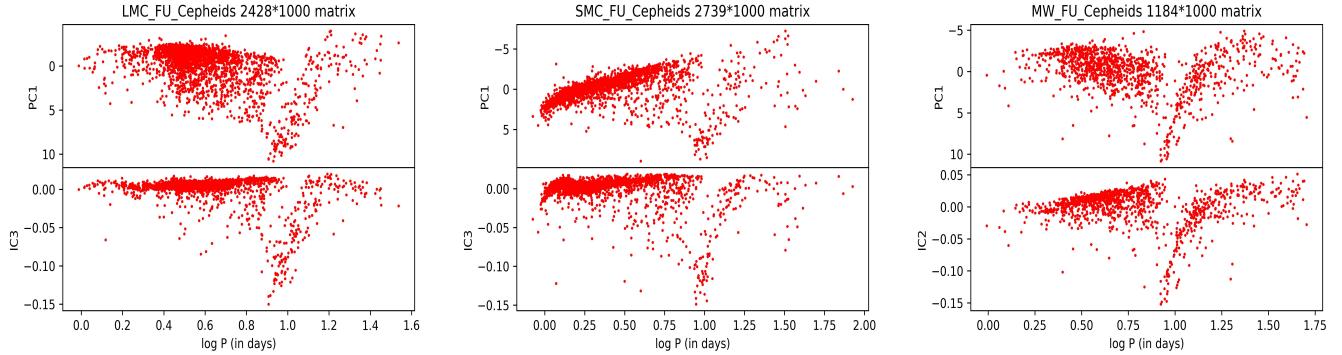
PC	Eigenvalue ( $\lambda$ )	Percentage	Cum. Percentage
1	5.615142	55.9325	55.9325
2	2.348315	23.3916	79.3241
3	0.944975	9.4129	88.7370

light curve shape with the help of vastly available observed light curve data (Kanbur et al. 2002; Kanbur & Mariani 2004; Tanvir et al. 2005; Deb & Singh 2009). ICA is also a dimension reduction technique similar as PCA and hopefully it can also contribute to identify structural changes around FU and FO Cepheids. This is the motivation behind this part. In this section, the capability of both PCA and ICA have been compared in the case of extracting resonances of FU and FO Cepheids individually for LMC, SMC and MW along with FD method mentioned in the paper Deb & Singh (2009).

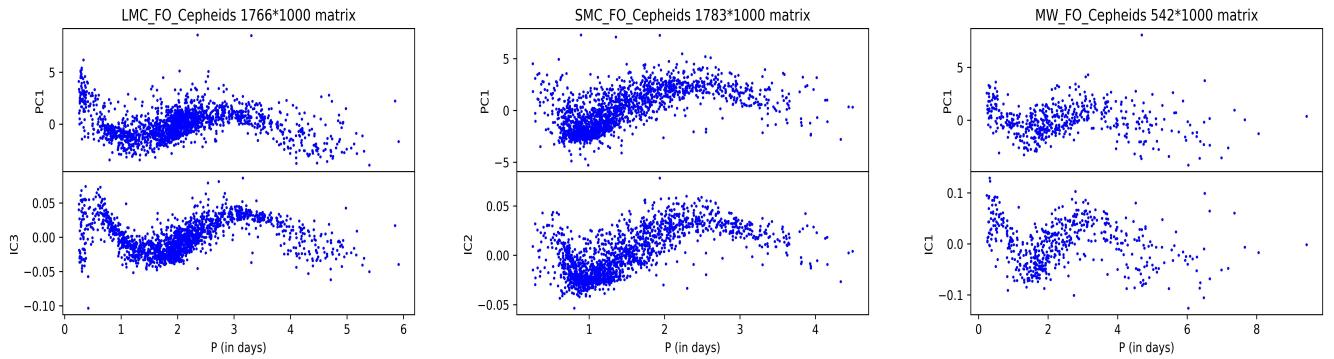
#### 4.1.1 Fundamental mode (FU) Cepheids

All the light curve data have been analysed for FU Cepheids are from latest OGLE-IV database. Deb & Singh (2009) used data from various sources (mostly from OGLE) and add LMC and SMC (I-band and V-band) data altogether for analysis, but we have used only I-band data separately for LMC, SMC and MW (data sets 1C, 2C, 3C respectively from Table 1). For LMC, the input matrix consisting of  $2428 \times 1000$  array corresponding to 2428 light curves for FU Cepheids and each consists of 1000 data points (here magnitudes) between phase 0 to 1 with step length 0.001. Table 2 explained about output after performing PCA on  $2428 \times 1000$  data matrix.

The first 3 PCs (i.e.  $\lambda \geq 1$ ) contain more than 85% of the variance in the data. Thus we consider the optimum number of ICs as 3 for LMC. Similarly by performing PCA on a data matrix of order  $2739 \times 1000$  corresponding to 2739 light curves for the SMC FU Cepheids, we found that first 4 PCs (i.e.  $\lambda \geq 1$ ) explained above 85% of the variance in the data. Thus for SMC, optimum number of ICs is equal to 4. Finally, for MW FU Cepheids ( $1184 \times 1000$ ), almost 85% of the variance explained by 4 PCs (i.e.  $\lambda \geq 1$ ). Therefore in this case the optimum number of ICs is equal to 4. Components correspond to PCA and ICA are plotted against  $\log P$  (in days) in Fig. 3. Deb & Singh (2009), in their paper found a discontinuity in the shape around  $\log P = 1$  by applying FD and PCA techniques separately for LMC FU Cepheids however the discontinuity is more pronounced for the former one. It is evident from Fig. 3 that the discontinuity around  $\log P = 1$  is more prominent for ICA component compared to PC1 although we have used more light curve data for LMC as well as SMC and MW FU Cepheids in our analysis. Due to lack of data points, Kanbur et al. (2002) could not find any resonance suggested by Antonello & Morelli (1996) in the period range  $1.38 < \log P < 1.43$  using PCA on Fourier coefficients. However, structural changes around  $\log P \sim 1.5$  are visible both for PCA and FD analysed by Deb & Singh (2009). Using PCA on a larger data set of LMC classical Cepheids, structural change around  $\log P \sim 1.35$  is visible (left column in Fig. 3) but not so clear because of scattering of data points in the period range  $1.2 < \log P < 1.6$ . On the contrary in the case of ICA, the structural pattern is much more precise and the location of change is relatively clear which is  $\log P \sim 1.3$ , close to the range suggested by Antonello & Morelli (1996). For SMC FU Cepheids, due to scattering of data points, the structural pattern beyond  $\log P = 1$  is not clear for PC1



**Figure 3.** Components of PCA and ICA have been plotted as a function of  $\log P$  (in days) separately for the 2428 I-band FU Cepheids in LMC (left column), 2739 I-band FU Cepheids in SMC (middle column) and 1184 I-band FU Cepheids in MW (right column). Upper panel is for PCA component and lower panel is for ICA component in each of the columns



**Figure 4.** Components of PCA and ICA have been plotted as a function of  $P$  (in days) separately for the 1766 I-band FO Cepheids in LMC (left column), 1783 I-band FO Cepheids in SMC (middle column) and 542 I-band FO Cepheids in MW (right column). Upper panel is for PCA component and lower panel is for ICA component in each of the columns.

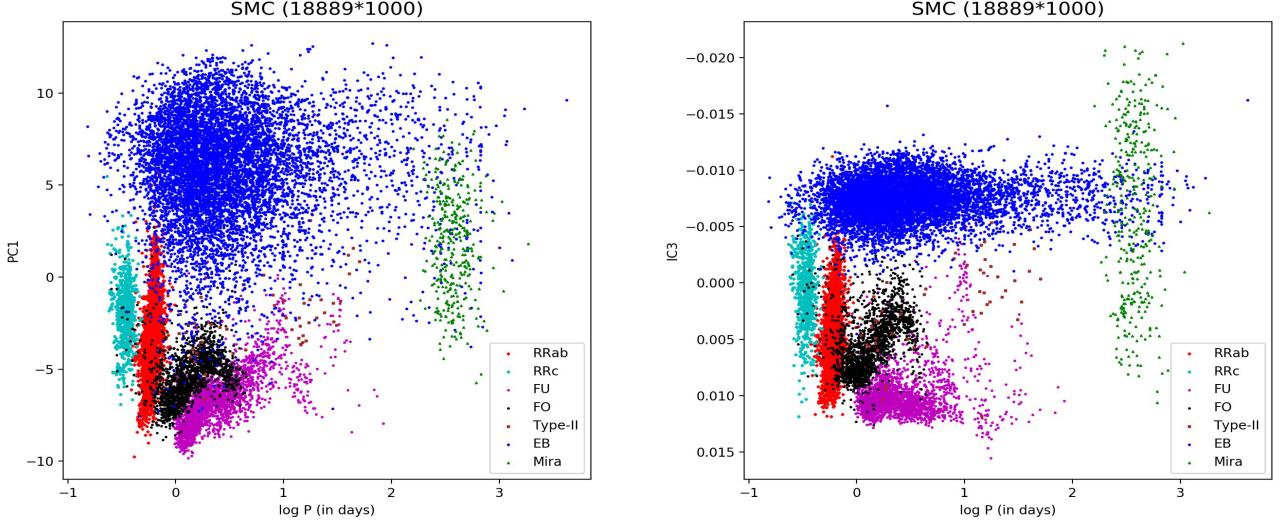
but the structural changes around the period range  $1.3 < \log P < 1.4$  is prominent in the case of IC3 which is almost similar to the range suggested by Antonello & Morelli (1996). Also, there is slight change in pattern somewhere around  $\log P \sim 1.8$  is visible for both the cases in SMC but confirmation of the existence of such resonance is difficult due to lack of data points in these period range. In such scenarios, radiative hydrodynamical models are useful (Kienzle et al. 1999). In the case of MW Cepheids (right column in Fig. 3), performance of ICA and PCA are similar.

#### 4.1.2 First overtone (FO) Cepheids

We have collected 1766 (set 1D in Table 1), 1783 (set 2D in Table 1) and 542 (set 3D in Table 1) I-band light curve data of FO Cepheids towards LMC, SMC and MW respectively from OGLE-IV database. For LMC, first 3 PCs (i.e.  $\lambda \geq 1$ ) corresponds to 85% of the variance in the data. So, we set optimum number of ICs as 3 for our analysis. Performing PCA on a  $1783 \times 1000$  matrix for SMC, first 4 PCs (i.e.  $\lambda \geq 1$ ) contain approximately 85% variance. So, for SMC optimum number of ICs is taken as 4. However for analysis on MW FO Cepheids, it takes 5 PCs (i.e.  $\lambda \geq 1$ ) to counter nearly 85% variance in the data. Thus we have taken 5 as the optimum number of components for ICA. Fig. 4 carries outputs for PCA and ICA respectively for LMC (left column), SMC (middle column) and

MW (right column) data sets. A short period discontinuity emerges around  $P \sim 0.4$  days (similar result found by Soszyński et al. (2008), Deb & Singh (2009) for LMC) for all the sub figures of Fig. 4. For LMC FO Cepheids there is a change in shape around  $P \sim 1.5$  days, evident for IC3 but difficult to visualize in the case of PC1. Also, we found a definite pattern change around a period of  $\sim 3.2$  days for both PCA and ICA analyses could be explained by the occurrence of resonance between first and fourth overtone (Antonello & Poretti 1986; Peterson 1989). However, authors like Kienzle et al. (1999) and Feuchtinger et al. (2000) using hydrodynamical models of FO Cepheids have shown that 3.2 days is not a resonance but Deb & Singh (2009) found this as resonance using FD and PCA both. There is also some change visible around  $\sim 4.5$  days only in the case of PC1 suggested by both Buchler et al. (1996) and Kienzle et al. (1999). However ICA fails to detect such resonance similar as for FD analysed by Deb & Singh (2009). For SMC FO Cepheids, there is clearly a change in shape in period-luminosity relation around  $\sim 2.5$  days (Kanbur et al. 2002; Bharadwaj et al. 2016) visible for both PC1 and IC2. For MW FO Cepheids, the structural changes are visible around 1.5 days, 3.2 days. However IC1 does a better job by bringing out the changes than PC1. ICA performs better in bringing out the resonance around  $\sim 4.5$  days than PCA.

Thus in general, the performance of ICA technique is better than both



**Figure 5.** Classification of different types of variable light curves based on PCA and ICA method separately. The input matrix is of order 18889×1000 for SMC. PC1 plotted as a function of  $\log P$  (in days) (left) and IC3 plotted against  $\log P$  (in days) (right). Types of variable stars are outlined in plot legend.

PCA and FD methods to extract the structural patterns of FU and FO Cepheids for LMC, SMC and MW respectively.

#### 4.2 Identification of PCA and ICA components : a comparative study

In this section, we demonstrate the capabilities of unsupervised machine learning techniques like PCA and ICA in classifying different kinds of variable stars light curve data sets collected from OGLE database for LMC, SMC as well as MW, described in Table 1. For SMC, PCA has been performed on a total of 18889 (set  $2A + 2B + 2C + 2D + 2E + 2F + 2G$  in Table 1) light curves containing 1000 magnitudes each. We found that the first 6 PCs (i.e.  $\lambda \geq 1$ ) explain more than 85% of variance in the data set whereas PC1 contains 63.86%. Thus the optimum number of ICs to be chosen as 6. PC1 and IC3 are plotted as a function of  $\log P$  (in days) separately in Fig. 5 for all types of variable stars. There is some overlapping between eclipsing binaries (7938 in total) and other variable classes through out the period range for PC1 but the separation is much more efficient in the case of IC3, however some overlapping between eclipsing binaries and Mira variables are visible in the period range  $2.5 < \log P < 3$ , that might be due to some outliers present in the data set of eclipsing binaries. Slight overlapping between RRab and FO Cepheids are seen in both cases (PCs and ICs) over a small period range. IC3 not only is able to separate FU and FO Cepheids to a large extend and perform better than PC1 but also prominently figure out the discontinuity of FU variable around  $\log P = 1$ . PC1 and IC3 both fail to separate FO and Type-II Cepheids in  $0 < \log P < 1$ . However no overlapping is visible between FU and Type-II Cepheids.

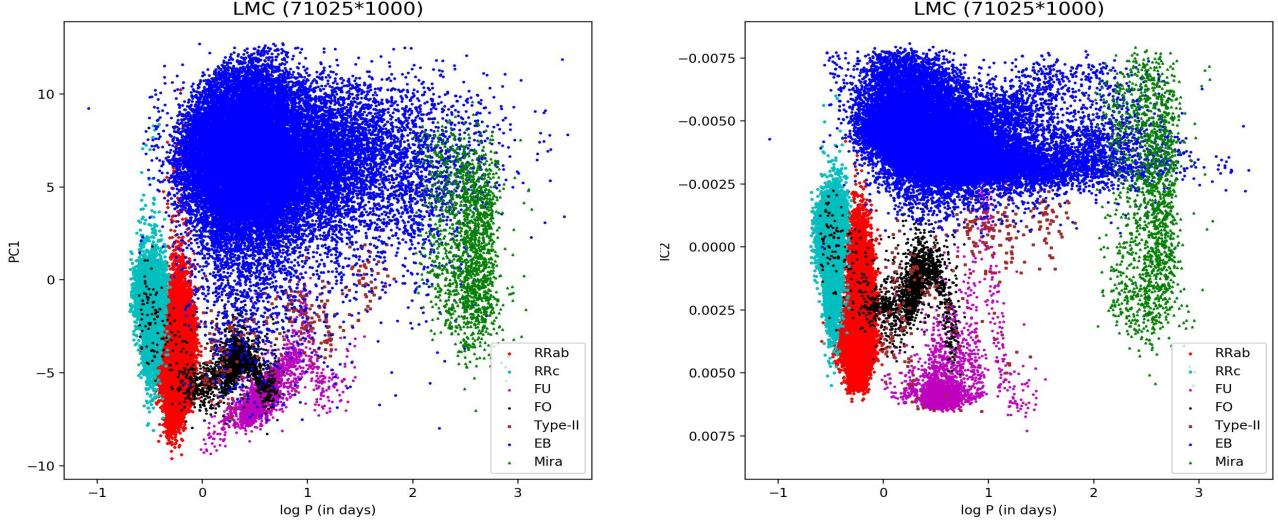
For LMC the input matrix is  $71025 \times 1000$  which is basically a collection of 1000 interpolated magnitude data points having phase 0 to 1 with step length 0.001 each for 71,025 light curves (set  $1A + 1B + 1C + 1D + 1E + 1F + 1G$  in Table 1). We found out that the first 5 PCs (i.e.  $\lambda \geq 1$ ) explain above 85% of variance in the data set whereas PC1 contains 62.36%. Thus the optimum number of ICs to be taken is 5. Deb & Singh (2009) used Fourier decomposi-

tion parameter  $R_{21}$  and PC1 to classify nearly about 17,000 stars and found that the sample of RR Lyraes (RRab and RRc) and Cepheids (FU and FO) could not separate well in the whole data set. From Fig. 6, it can be seen that the amount of overlap between FO and RR Lyrae (RRab and RRc) samples is trivial in terms of sample data used for our analysis, however PC1 is not able to separate FU and FO Cepheids well enough which is not the case for IC2. Also, the discontinuity of FU Cepheids around  $\log P = 1$  is clearly visible for IC2. There is certain mix up between eclipsing binaries with other variable stars for PCA analysis whereas in the period range of 1 to 100 days, eclipsing binaries have distinct IC2 values from other variable stars. Considerable amount of overlapping between eclipsing binaries and Mira variables have been seen in both cases in the range of  $2 < \log P < 3$ , that might be due to presence of some outliers in the data set. Type-II variables are differentiated well for IC2 in the period range  $1 < \log P < 2$  and  $0 < \log P < 0.2$ . There is some overlapping between Type-II and FO for both the cases but in case of IC2 it is much reduced.

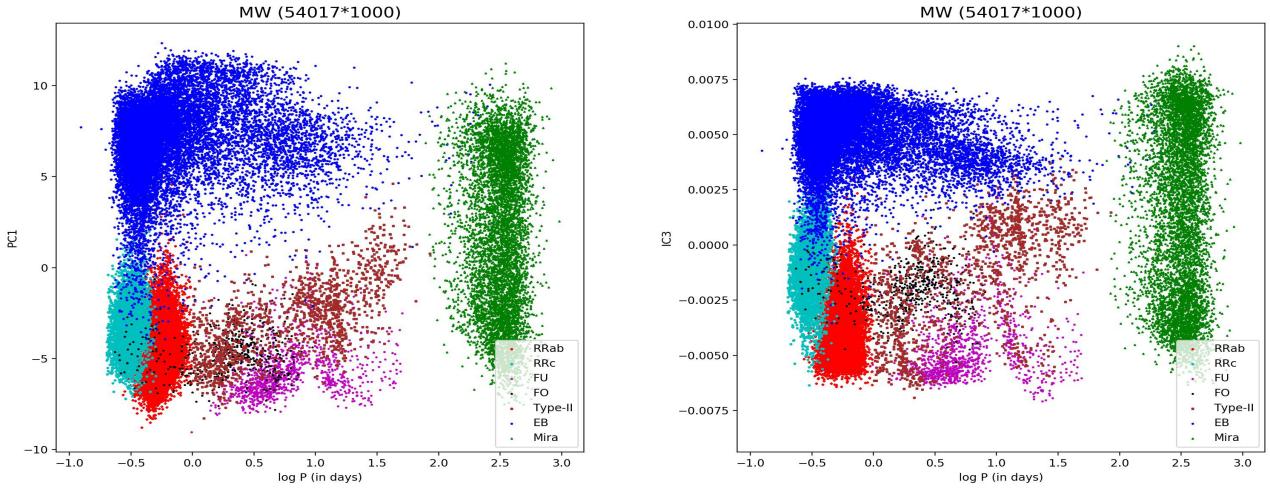
For MW galaxy, the input matrix is of order  $54017 \times 1000$  (set  $3A + 3B + 3C + 3D + 3E + 3F + 3G$  in Table 1). Applying PCA, we figured out that the first 6 PCs (i.e.  $\lambda \geq 1$ ) explain above 85% of variance in the data whereas PC1 contains 51.57%. Thus for MW, the optimum number of ICs to be taken as 6. From Fig. 7, it may be noted that samples of EB, RRab, RRc and Mira variables are well separated more for IC3 than PC1 though there are some overlapping among Cepheids FU, FO and Type-II. Overall it is clear that the performance of ICA technique is much better for classification of big data than PCA and FD methods.

#### 4.3 K Means clustering with respect to ICs

In the present work, we have used huge data sets of the light curves (LC) of variable stars in SMC, LMC and MW and have classified the LC by applying CA with respect to the optimum number of IC components. Interestingly we have found four coherent similar groups irrespective of the host galaxies. They are EB (G1), Cepheids



**Figure 6.** Classification of different types of variable light curves based on PCA and ICA method separately. The input matrix is of order  $71025 \times 1000$  For LMC. PC1 plotted as a function of  $\log P$  (in days) (left) and IC2 plotted against  $\log P$  (in days) (right). Types of variable stars are outlined in plot legend.



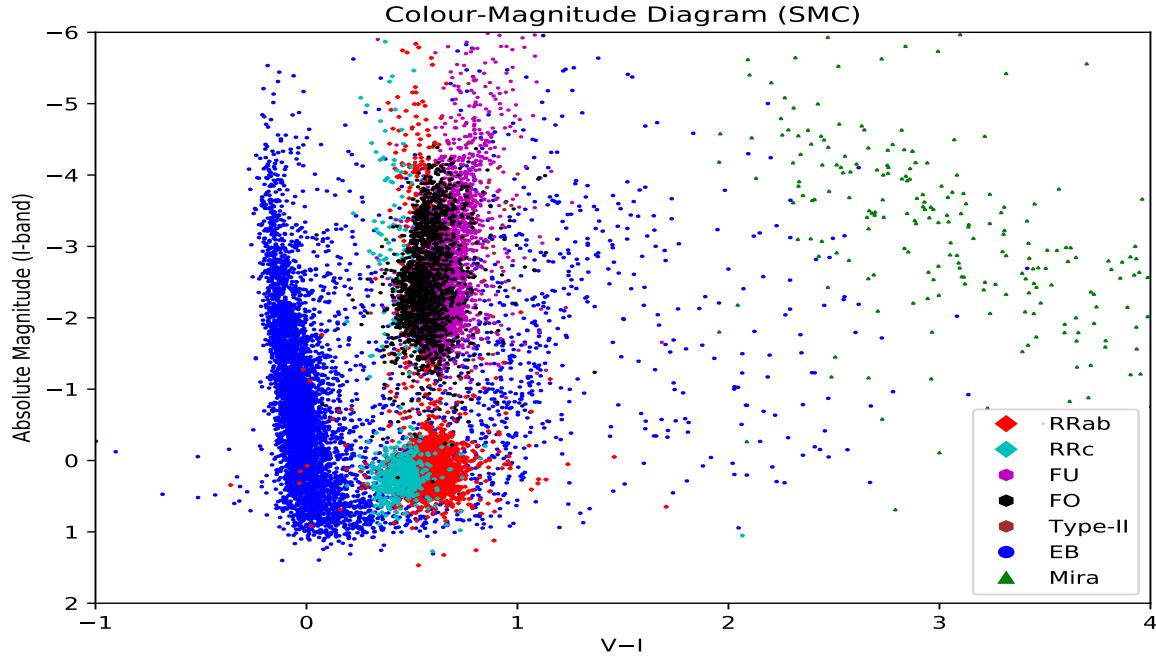
**Figure 7.** Classification of different types of variable light curves based on PCA and ICA method separately. The input matrix is of order  $54017 \times 1000$  For SMC. PC1 plotted as a function of  $\log P$  (in days) (left) and IC3 plotted against  $\log P$  (in days) (right). Types of variable stars are outlined in plot legend.

(FU, FO, Type-II) (G2), RRab and RRc (G3) and Mira variables (G4).

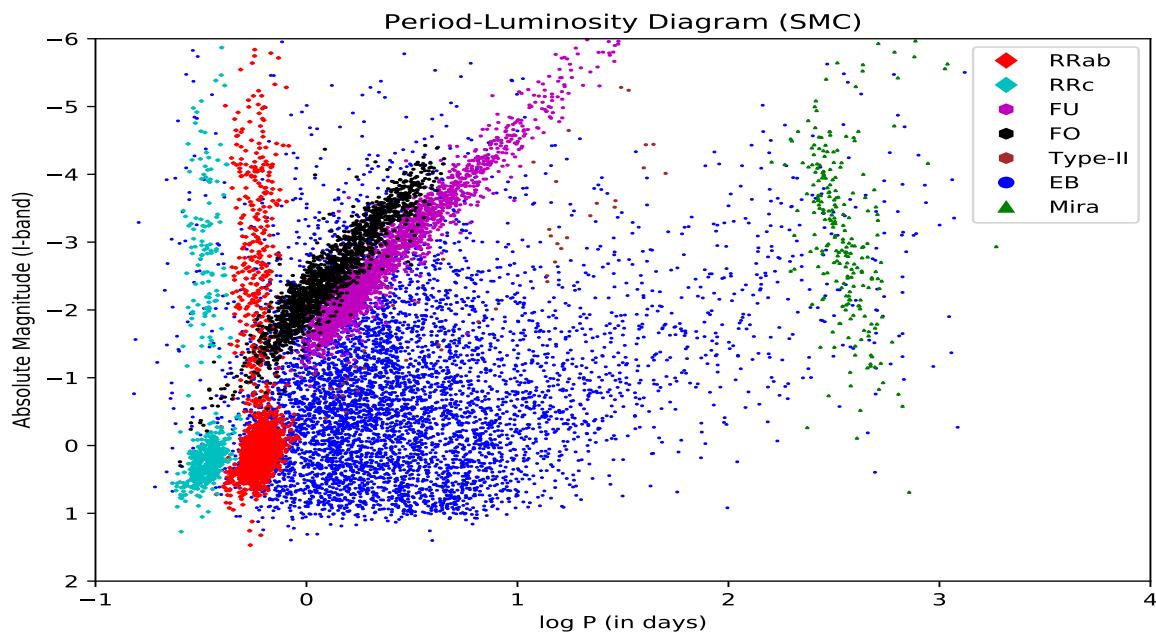
The following features have been observed :

- (i) EB (G1) are mostly main sequence stars and they contain mostly O and B type stars ( $V-I \sim 0$ ) and their luminosity covers a wide range of spectrum ( $M_I \sim -1$  to  $-5$ ) i.e this group has a wide range in the mass. Their periods fall in the moderate range ( $\log P \sim 0$  to  $1$ ). This is consistent with the fact that in young star clusters the contribution of binary stars is significant (Hu et al. 2010, Wang et al. 2020). Very few stars have evolved from the main sequence branch.

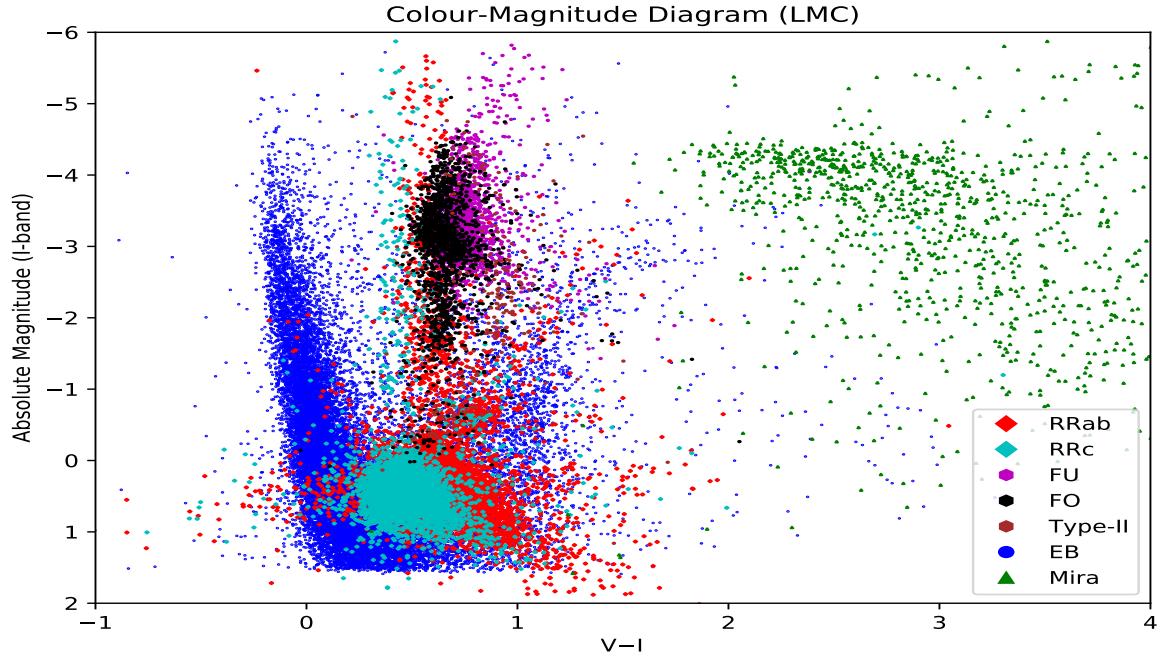
- (ii) Interestingly all Cepheids (FU, FO and Type-II) fall in one group (G2). All the stars are very bright ( $M_I \sim -2$  to  $-4$ ) and are comparatively metal rich. The stars have evolved from the main sequence and lie in the Cepheid strip i.e along giant and super giant branches. The Cepheids of FO type have smaller period and lesser magnitudes compared to FU type Cepheids. FU Cepheids are also redder than FO ones. There are slight differences in the magnitudes and colours of Cepheids and EB in SMC, LMC and MW, e.g. SMC contains brighter Cepheids ( $M_I \sim -4$ ) than LMC and MW. Also, MW, EB are much redder than SMC and LMC, EB and they have



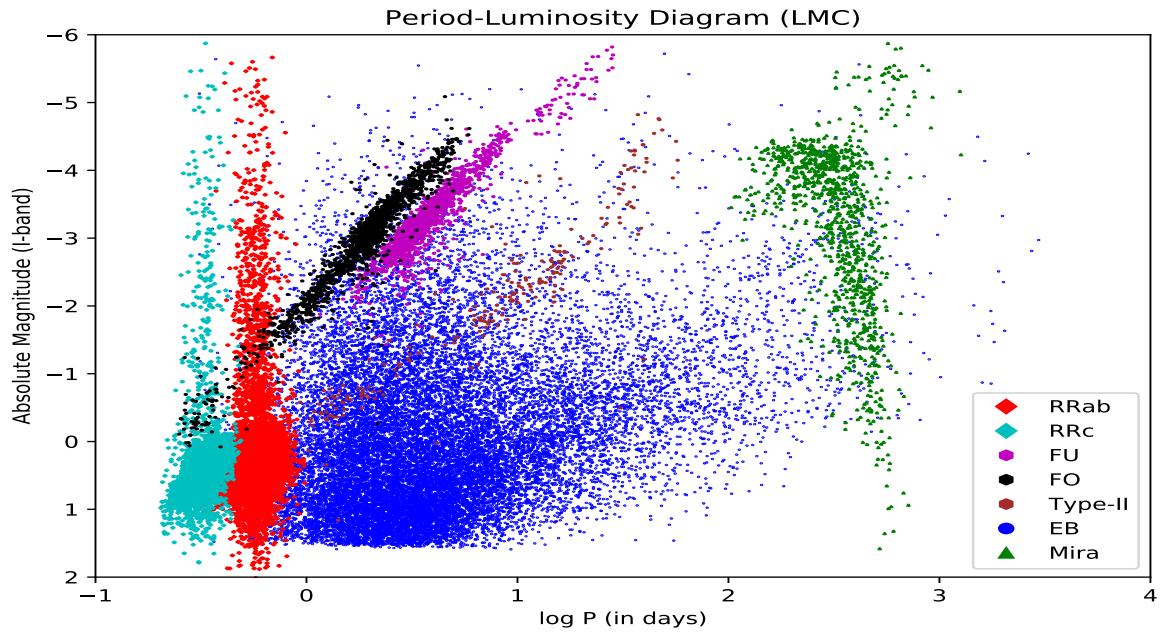
**Figure 8.** Colour-Magnitude diagram for SMC is plotted by applying K-means clustering on six ICs. Different colours represent different variable classes and variables which have similar marker indicator contained in same clusters.



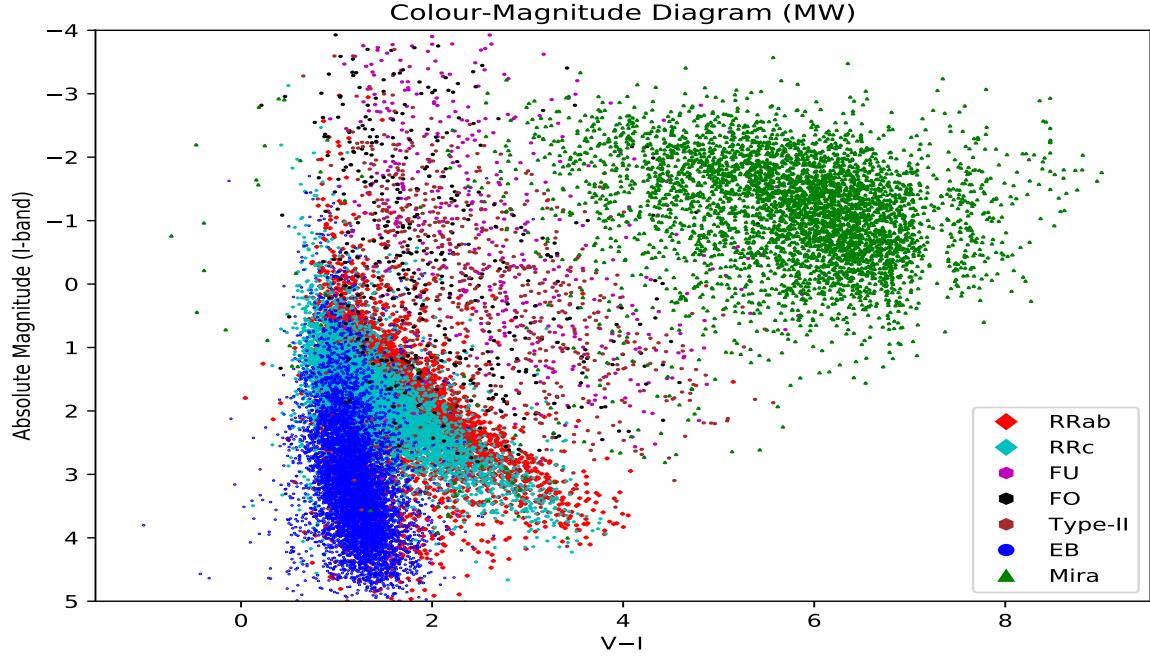
**Figure 9.** Period-Luminosity diagram for SMC is plotted by applying K-means clustering on six ICs. Different colours represent different variable classes and variables which have similar marker indicator contained in same clusters.



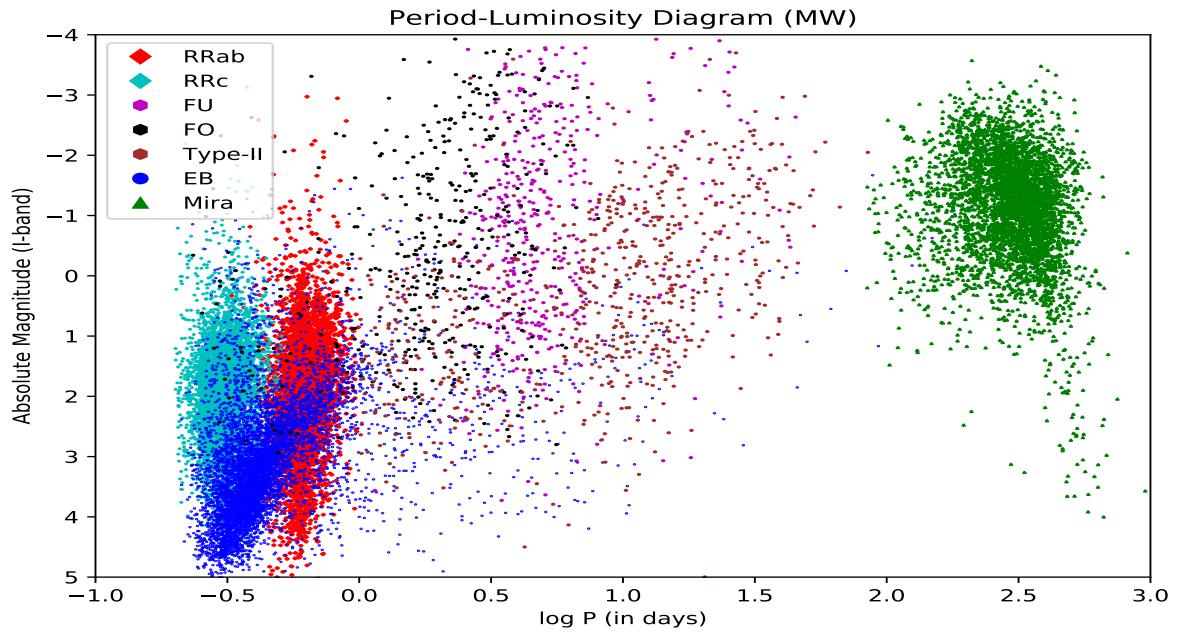
**Figure 10.** Colour-Magnitude diagram for LMC is plotted by applying K-means clustering on five ICs. Different colours represent different variable classes and variables which have similar marker indicator contained in same clusters.



**Figure 11.** Period-Luminosity diagram for LMC is plotted by applying K-means clustering on five ICs. Different colours represent different variable classes and variables which have similar marker indicator contained in same clusters.



**Figure 12.** Colour-Magnitude diagram for Milky Way is plotted by applying K-means clustering on six ICs. Different colours represent different variable classes and variables which have similar marker indicator contained in same clusters.



**Figure 13.** Period-Luminosity diagram for Milky Way is plotted by applying K-means clustering on six ICs. Different colours represent different variable classes and variables which have similar marker indicator contained in same clusters.

much shorter period ( $\log P \sim -0.5$  to 0) compared to SMC and LMC ( $\log P \sim 0$  to 1). This might be due to that they are less massive as seen from their very low luminosity ( $M_I \sim -2$  to 0). The period and luminosity correlation is stronger in SMC and LMC compare to MW.

(iii) All RR-Lyrae variables fall in the same group (G3) though RRab have longer period than RRc variables. RR-Lyrae variable stars just turn off from main sequence branch and on the way to giant branch.

(iv) It is clear from the color magnitude diagram (CMD) that Mira variables form a separate group (G4) occupying a red giant region of the CMD having brightest magnitudes, reddest colours and longest periods. They are old stars (Baud et al. 1981; Feast 2009) and exceeding main sequence luminosity.

So, it is interesting from the above observation that ICA along with CA can clearly classify the four categories of variable stars with great efficiency and very little amount of overlapping and these four groups are robust i.e independent of their host galaxies. This is the success of the present study. Also the period-luminosity diagram clearly shows the correlation for Type-I Cepheids though the correlation is not very clear for Type-II Cepheids due to small size of the sample. Though the difference between FU and FO Cepheids are not clear in CMD but the difference is clear in period-luminosity diagram as two parallel strips which is only due to the mode of their vibration and not due to chemical composition. This is also true for RRab and RRc variables which form a coherent group.

So, we can conclude that ICA along with CA is a very strong and robust tool which can be used for the classification of variable stars and for future prediction on the nature of light curves of variable stars.

## 5 CONCLUSIONS

The present work deals with a huge database of OGLE variable stars light curve data (RR Lyrae, Cepheids, EB, Mira etc.). It consists of Heliocentric Julian Days (HJD), observed I-band magnitudes. Using spline interpolation, we obtain 1000 magnitudes for each light curve between phase 0 to 1 with step length 0.001. The input to PCA and ICA is the matrix form of normalized magnitude data set obtained using equation (5).

The following features have been observed :

(i) PCA and ICA have been applied separately to matrices of order  $2428 \times 1000$ ,  $2739 \times 1000$  and  $1184 \times 1000$  corresponding to light curves for FU Cepheids for LMC, SMC and MW respectively and we observed that the overall performance of ICA is better than PCA and FD (Deb & Singh 2009) for finding out resonances in FU cepheids in LMC and SMC but for MW the performance of ICA and PCA is similar. For FO Cepheids, the input matrices are  $1766 \times 1000$ ,  $1783 \times 1000$  and  $542 \times 1000$  towards LMC, SMC and MW respectively and we find out that the performance of PCA and ICA is similar for LMC and SMC but for MW the overall performance of ICA is better for bringing out the structural changes in the shape of FO Cepheids with respect to their periods.

(ii) We demonstrate the capabilities of PCA and ICA in classifying different variable stars for LMC, SMC and MW separately.

For SMC the input matrix for PCA and ICA is of order  $18889 \times 1000$ . From Fig. 5, it is clear that IC3 outperforms PC1 not only in the separation of FU and FO Cepheids but also separation between eclipsing binary to other variables except for Mira variables due to some outlier present in the data. Similarly, IC2 performs much better than PC1 in the case of LMC where the input matrix is of order  $71025 \times 1000$ . There is no such difference in the performance of ICA and PCA in the case of MW where the input matrix is of order  $54017 \times 1000$ .

(iii) Finally, applying K-Means clustering on ICs we have plotted the period-luminosity diagram and the colour-magnitude diagram separately for LMC, SMC and MW galaxies. Interestingly ICA along with CA introduce four categories of variable stars namely EB (G1), Cepheids (G2), RR Lyrae (G3) and Mira (G4), independent of host galaxies as seen from Fig. (8) - (13).

Therefore we conclude by saying that ICA is a dimension reduction technique which has the capability of identifying possible resonances for large light curve data sets of FU and FO Cepheid and performs better than PCA. Also, ICA outperforms PCA and FD method (Deb & Singh 2009) in classifying big data sets, which will be more suitable when dealing with larger databases with a more diverse set of variable stars in future. Also, ICA along with CA is a robust classifier for variable stars helping for future predictions on the nature of variable stars.

## ACKNOWLEDGEMENTS

Author S.P acknowledges INSPIRE SRF Vide sanction Order No. DST/INSPIRE Fellowship/2017/IF170368 under the Department of Science and Technology (DST) INSPIRE program, Government of India. The authors are grateful to Mr. Prasenjit Banerjee for providing valuable support.

## DATA AVAILABILITY

The data underlying this article are available in OGLE On-line Data Directory, at <https://ogle.astrouw.edu.pl/>.

## REFERENCES

- Aikawa T., 1984, MNRAS, 206, 833
- Albazzaz H., Wang X. Z., 2004, Industrial & engineering chemistry research, 43, 6731
- Andreasen G. K., 1988, A&A, 191, 71
- Andreasen G. K., Petersen J. O., 1987, A&A, 180, 129
- Antonello E., 1994, A&A, 282, 835
- Antonello E., Morelli P. L., 1996, A&A, 314, 541
- Antonello E., Poretti E., 1986, A&A, 169, 149
- Antonello E., Fugazza D., Mantegazza L., 2000, A&A, 356, L37
- Baud B., Habing H. J., Matthews H. E., Winnberg A., 1981, A&A, 95, 156
- Beaulieu J. P., 1998, MmSAI, 69, 21
- Bhardwaj A., Kanbur S. M., Singh H. P., et al. 2015, MNRAS, 447, 3342
- Bhardwaj A., Ngeow C. C., Kanbur S. M., et al. 2016, MNRAS, 458, 3705
- Bhardwaj A., Kanbur S. M., Marconi M., et al. 2017, MNRAS, 466, 2805
- Bono G., Marconi M., Stellingwerf R. F., 1999, ApJS, 122, 167
- Bono G., Marconi M., Stellingwerf R. F., 2000, A&A, 360, 245
- Buchler J. R., Kollath Z., Beaulieu J. P., et al. 1996, ApJ, 462, L83
- Cassisi C., Montalto P., Aliotta M., et al. 2012, Similarity Measures and Dimensionality Reduction Techniques for Time Series Data Mining. Intech Open Limited

- Chattopadhyay T., Misra R., Chattopadhyay A. K., et al. 2007, ApJ, 667, 1017  
 Chattopadhyay T., Sharina M., Karmakar P., et al. 2010, ApJ, 724, 678  
 Chattopadhyay T., Sharina M., Davoust E., et al. 2012, ApJ, 750, 91  
 Chattopadhyay A. K., Chattopadhyay T., De T., Mondal S., 2013, in Springer Series in Astrostatistics Vol.-1. Astrostatistical Challenges for the New Astronomy (Springer-Verlag New York), p. 185  
 Chattopadhyay T., Fraix-Burnet D., Mondal S., 2019, PASP, 131, 108010  
 Comon P., 1994, Signal Process, 36, 287  
 Das S., Chattopadhyay T., Davoust E., 2015, PASA, 32, e041  
 Deb S., Singh H. P., 2009, A&A, 507, 1729  
 Deb S., Singh H. P., 2010, MNRAS, 402, 691  
 Deb S., Singh H. P., 2014, MNRAS, 438, 2440  
 Feast M. W., 2009, in Ueta T., Matsunaga N., Ita Y., eds, AGB Stars and Related Phenomena. p. 48  
 Feuchtinger M., Buchler J. R., Kolláth Z., 2000, ApJ, 544, 1056  
 Freedman W. L., Madore B. F., 2010, ARA&A, 48, 673  
 Galaz G., de Lapparent V., 1998, A&A, 332, 459  
 Graczyk D., Soszyński I., Poleski R., et al. 2011, AcA, 61, 103  
 Hajdu G., Dékány I., Catelan M., et al. 2018, ApJ, 857, 55  
 He S., Wang L., Huang J. Z., 2018, ApJ, 857, 110  
 Hendry M. A., Tanvir N. R., Kanbur S. M., 1999, in Harmonizing Cosmic Distance Scales in a Post-Hipparcos Era Vol. 167, Egret D., Heck A., eds. ASP Conference Series, p. 192  
 Hertzsprung E., 1926, Bull. Astron. Inst. Netherlands, 3, 115  
 Hu Y., Deng L., de Grijs R., Liu Q., Goodwin S. P., 2010, ApJ, 724, 649  
 Kairov U., Cantini L., Greco A., et al. 2017, BMC Genomics, 18, 712  
 Kalinova V., Colombo D., Rosolowsky E., et al. 2017, MNRAS, 469, 2539  
 Kanbur S. M., Mariani H., 2004, MNRAS, 355, 1361  
 Kanbur S. M., Iono D., Tanvir N. R., et al. 2002, MNRAS, 329, 126  
 Kienzle F., Moskalik P., Bersier D., et al. 1999, A&A, 341, 818  
 Klapp J., Goupil M. J., Buchler J. R., 1985, ApJ, 296, 514  
 Kova'cs G., Jurcsik J., 1997, A&A, 322, 218  
 Leavitt H. S., Pickering E. C., 1912, Harvard College Observatory Circular, 173, 1  
 Leonard D. C., Kanbur S. M., Ngeow C. C., et al. 2003, ApJ, 594, 247  
 MacQueen J. B., 1967, in Some Methods for Classification and Analysis of Multivariate Observations. in proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, 281  
 Moskalik P., Poretti E., 2003, A&A, 398, 213  
 Moskalik P., Buchler J. R., Marom A., 1992, ApJ, 385, 685  
 Nemec J. M., Smolec R., Benkő J. M., et al. 2011, MNRAS, 417, 1022  
 Ngeow C. C., Kanbur S. M., Nikolaev S., et al. 2003, ApJ, 586, 959  
 Paczyński B., 1986, ApJ, 301, 503  
 Pawlak M., Soszyński I., Udalski A., et al. 2016, AcA, 66, 421  
 Petersen J. O., 1989, A&A, 226, 151  
 Pires S., Juin J. B., Yvon D., et al. 2006, A&A, 455, 741  
 Poretti E., 2001, A&A, 371, 986  
 Riess A. G., Casertone S., Yuan W., et al. 2018, ApJ, 861, 126  
 Sarro L. M., Debosscher J., Aerts C., et al. 2009, A&A, 506, 535  
 Sarro L. M., Ordieres-Meré J., Bello-García A., et al. 2018, MNRAS, 476, 1120  
 Schaltenbrand R., Tammann G. A., 1971, A&AS, 4, 265  
 Sheldon K., Richards G., 2018, AAS, 231, 250.23  
 Simon N. R., 1977, ApJ, 217, 160  
 Simon N. R., Kanbur S. M., 1995, ApJ, 451, 703  
 Simon N. R., Lee A. S., 1981, ApJ, 248, 291  
 Simon N. R., Schmidt E. G., 1976, ApJ, 205, 162  
 Simon N. R., Teays T. J., 1982, ApJ, 261, 586  
 Simon N. R., Teays T. J., 1983, ApJ, 265, 996  
 Singh H. P., Gulati R. K., Gupta R., 1998, MNRAS, 295, 312  
 Soszyński I., Poleski R., Udalski A., 2008, AcA, 58, 163  
 Soszyński I., Udalski A., Szymański M. K., et al. 2009, AcA, 59, 239  
 Soszyński I., Dziembowski W. A., Udalski A., et al. 2011a, AcA, 61, 1  
 Soszyński I., Udalski A., Szymański M. K., et al. 2011b, AcA, 61, 217  
 Soszyński I., Udalski A., Szymański M. K., et al. 2013, AcA, 63, 21  
 Soszyński I., Udalski A., Szymański M. K., et al. 2015, AcA, 65, 297  
 Soszyński I., Udalski A., Szymański M. K., et al. 2016a, AcA, 66, 131  
 Soszyński I., Pawlak M., Pietrukowicz P., et al. 2016b, AcA, 66, 405  
 Soszyński I., Udalski A., Szymański M. K., et al. 2018, AcA, 68, 89  
 Soszyński I., Udalski A., Wrona M., et al. 2019, AcA, 69, 321  
 Soszyński I., Udalski A., Szymański M. K., et al. 2020, AcA, 70, 101  
 Storrie-Lombardi M. C., Irwin M. J., von Hippel T., et al. 1994, VA, 38, 331  
 Sugar C. A., James G. M., 2003, J. Am. Stat. Assoc., 98, 750  
 Takeuti M., Aikawa T., 1981, Sci. Rep. Tohoku Univ. Eighth Ser., 2, 106  
 Tanvir N. R., Ferguson H. C., Shanks T., 1999, MNRAS, 310, 175  
 Tanvir N. R., Hendry M. A., Watkins A., et al. 2005, MNRAS, 363, 749  
 Wang C., Langer N., Schootemeijer A., et al. 2020, ApJ, 888, L12  
 Welch D. L., Alcock C., Allsman R. A., et al. 1997, in Variable Stars and the Astrophysical Returns from the Microlensing Surveys, Ferlet R., Maillard J. -P. & Raban, B., eds. Gif-Sur-Yvette, France, p. 205  
 Yip C. W., Connolly A. J., Vanden Berk D. E., et al. 2004, AJ, 128, 2603

This paper has been typeset from a  $\text{\TeX}$ / $\text{\LaTeX}$  file prepared by the author.