

THE M33 SYNOPTIC STELLAR SURVEY. II. MIRA VARIABLES

XIAMI^{1,*}, XIAOYU¹

Draft version January 26, 2016

ABSTRACT

We have applied the newly developed semi-parametric periodogram ([Citation Coming Soon](#)) on M33 optical observation to search for Miras. We developed M33 data specific models to calculate the probabilities of being Mira for all possible targets. The training data set was constructed by simulating the M33 Mira light curves with known periods and non-Mira light curves in order to train our models. We found **XXX** Mira candidates, with a **XX**% discovery rate and **XX**% purity, which are estimated from the training data set. We tentatively classified the Oxygen-rich and Carbon-rich Miras and give their Period-Luminosity relations.

1. INTRODUCTION

Mira variable stars (Miras) belong to one type of long period variables which is commonly, and empirically, defined as periodic variables with *V*-band variation greater than 2.5 mag. The pulsating periods of Miras typically range from 100 days to 700 days, but extreme cases can reach beyond 1500 days. They are involved medium or low mass stars at the asymptotic giant branch (Feast 2009).

Thousands of Miras have been identified in the directions of the Milky Way bulge and Magellanic Clouds by the Optical Gravitational Lensing Experiment [Udalski et al. (1992); hereafter OGLE] and MACHO project (Alcock et al. 1993). The Mira light curves obtained by these projects are usually characterized by long time span, full phase coverage, and hundreds of epochs. The Mira light curves are not strictly periodic, but show long-term variations in the mean magnitude and cycle-to-cycle variations. The curve shape for different Miras also varies. Some examples of the light curve variation were described in Huemmerich & Bernhard (2012).

The Mira Period-Luminosity relation (P-L relation) was initially suggested by Gerasimovic (1928). Recent studies (Glass & Evans 2003) have shown this relation exhibits small scatter at *K* ($\sigma \sim 0.13$ mag), which can be used as distance indicator. The near-infrared Mira P-L relations in Large Magellanic Cloud (LMC), Small Magellanic Cloud, and the Galactic bulge have been explored by Soszyński et al. (2009), Soszyński et al. (2011), and Soszyński et al. (2013) respectively. However, extragalactic Mira P-L relations have not yet been calibrated. Bloom (2013) have proposed a dedicated Mira search in megamaser galaxy NGC 4258 to obtain the infrared Mira P-L relations. In the near future, the LSST project will cover dozens of nearby galaxies where Mira P-L relations can be obtained (LSST Science Collaboration et al. 2009).

In this paper, we conducted a Mira search in M33 using the DIRECT and its follow-up observations [Kaluzny et al. (1998); Pellerin & Macri (2011); hereafter M33 observations]. Although the observational baseline is up to nine years, the number of observations and data quality are significantly lower than those of the OGLE and

MACHO projects. At low cadence, the non-periodic variation components of the Mira light curves obstruct period detection. To address this problem, we adopted a semi-parametric periodogram ([Citation Coming Soon](#)) to search for Miras and their periods. This periodogram, which is designed for low-cadence Mira observations, has a Gaussian process component to account for the deviations from exact periodicity, and gives better performances than Lomb-Scargle (Lomb 1976; Scargle 1982) periodogram for M33 observations.

To help identify Miras among other type of stars, we carried out a comprehensive simulation and built classification models using machine learning methods. We extracted **XXX** features from simulated light curves and their frequency spectra to train the models, then applied the models on the M33 observations to classify Miras. We obtained **XXX** Mira candidates in M33 with a **XX**% misclassification rate.

This paper is organized as follows. Section 2 introduces the observations of M33. In Section 3 we report the methods of simulating M33 observations with prior knowledge of light curve classes and periods if they are Miras. Section 4 describes the models which were used to identify Miras and estimate their periods. Section 5 gives the results on real M33 observations.

2. OBSERVATIONS AND PHOTOMETRY OF M33

The detailed descriptions of observations and photometry procedures can be found in the first paper of the series (Pellerin & Macri 2011). Here we briefly summarize them.

Almost the entire disk of M33, which spans approximately half square degree in the sky (Roberts 1895), was observed by the DIRECT project and follow-up observations. The original DIRECT project used the Fred L. Whipple Observatory 1.2m telescope and the Michigan-Darmouth-MIT 1.3m telescope to observe M33 during 1996 September to 1999 November. Additional observations was obtained at the Wisconsin-Indiana-Yale-NOAO Observatory 3.5m telescope between 2002 August to 2006 August. In this paper we refer them collectively as M33 observations. We only used the *I*-band data to search for Miras. The number of valid *I*-band measurements for individual objects is shown in Figure 1. It is common that there is more than one observations in one night, and as a result the median number of nights with valid mea-

¹ DAHAI

* Corresponding author, HUP0@JIANGHE.edu

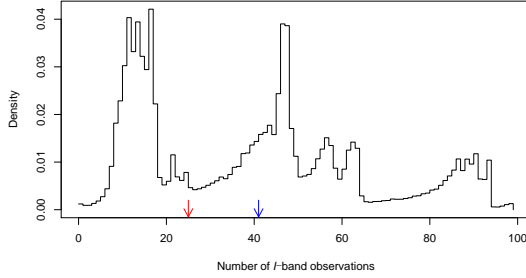


FIG. 1.— The distribution of number of valid I -band measurements for the objects in M33. Objects without detection in certain frames or mismatched cases do not count as valid measurements. The blue arrow indicates the median number of valid measurements per object, while the red arrow indicates the median number of nights with valid measurement.

measurements is even smaller. The observations consist of 29 fields, and different fields have different distributions of number of measurements.

We used the photometry products from Pellerin & Macri (2011). They performed PSF photometry using the DAOPHOT, ALLSTAR, and ALLFRAME (Stetson 1987, 1994), and carried out photometric calibration based on Massey et al. (2006). We used the I -band light curves with Stetson’s variability index $L > 0.6$ (Stetson 1996) to carry out Mira search.

3. SIMULATION OF M33 OBSERVATIONS

To build models for searching Miras from the M33 observations, we simulated Mira and Non-Mira light curves as training data. The simulated light curve includes Miras, constant stars, and semiregular variable stars (SRVs). Upon the simulated light curves, we calculated the variability index and kept those with $L > 0.6$. The number of light curves was chosen to match real observations, other properties such as sampling patterns and measurement uncertainties were also derived from the M33 observations. Below we describe the simulations of different kind of light curves.

3.1. Mira light curves

We simulated 5,000 Mira light curves for as part of the training data. The simulation method is similar to that in (cite GP), which applies the sampling patterns and noise levels of the M33 observations to LMC Mira light curves from the OGLE observations. The sampling patterns we used in this paper is slightly different from those in (cite GP) because we removed some noisy frames of the M33 observations from analysis. Here we briefly describe the method we used to simulate Mira light curves.

We generated templated LMC Mira light curves based on the OGLE observations, and derived their occurrence in the M33 observations. The LMC Miras can be considered as a complete sample and are well observed, with a baseline longer than 7.5 years and hundreds of observations at relatively small uncertainty. However, the light curves are by no means continuous, which are needed for simulating sparsely observed light curves at given sampling patterns. To obtain continuous Mira templates, we used the (model name) to fit the LMC Mira light

curves and predict their magnitude at any given time inside their time domain. We then shifted the template light curves by +6.2 magnitude to account for the distance difference between LMC and M33. For the M33 observations, most faint Miras can not be detected. We used the completeness functions of the M33 observations derived in (cite GP) as occurrence rate of the template light curves.

The sampling patterns were randomly selected from the light curves of M33 observations. We shifted the observation dates of the templates by some random amount to cover the selected M33 sampling dates. The random shifts made the simulated light curves different from each other, even when the same template and same sampling pattern were used.

The magnitude uncertainties of the M33 observations are quite different from the LMC light curves due to differences in overall apparent magnitude and in configurations of telescopes. To simulate the noise level of the M33 observations, we fit the relations between magnitude uncertainty σ and apparent magnitude m with the following empirical function for each frame of the M33 observations

$$\sigma(t_i, f_k) = a(t_i, f_k)^{[m - b(t_i, f_k)]} + c(t_i, f_k), \quad (1)$$

where a , b , and c are free parameters and varies with frames identified by time t_i and field f_k . The magnitude uncertainties of simulated light curves were determined by Equation 1 and normally distributed noises with standard deviations $\sigma(t_i, f_k)$ were added to the simulated light curves.

3.2. Constant star light curves

We simulated $\sim 5 \times 10^5$ light curves for constant stars, which matches the number of observed stars in M33. For the M33 observations, the magnitude uncertainties reported by the photometry software were sometimes either under-estimated or over-estimated. Since both the L index and the Mira periodogram utilize the value of magnitude uncertainties, we took the inaccuracy of reported magnitude uncertainties into account for the simulated light curves. For constant stars the inaccuracy of uncertainties can be described by

$$\chi_\nu^2 = \frac{1}{\sqrt{n-1}} \sum_{i=1}^n \frac{(m_i - \bar{m})^2}{\sigma_i^2}, \quad (2)$$

where n is the number of observations. Figure 2 shows the distribution of χ_ν^2 as a function of magnitude for all stars in one field of the M33 observations.

For each star in the M33 observations, we calculated the mean magnitude and χ_ν^2 by assuming that it is a constant star. To simulate a light curve, the magnitude of each epoch m_i was sampled from a normal distribution

$$m_i = \mathcal{N}(\bar{m}, \sqrt{\chi_\nu^2} \cdot \sigma_i), \quad (3)$$

where σ_i is the reported magnitude uncertainty for the i th epoch. We also shuffled the sampling pattern by randomly selecting n time points from all the valid epochs of the corresponding field. The magnitude uncertainties $\{\sigma_i\}$ kept unchanged for the simulated light curve. In this way the simulated light curves gave the same inaccuracy of magnitude uncertainties as the real light curves. The

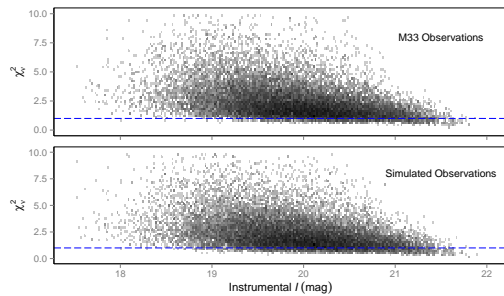


FIG. 2.— χ^2_ν as a function of instrumental I -band magnitude for one field. The depth of color is in log scale of number of stars in each pixel. The blue dashed line indicates the position of $\chi^2_\nu = 1$.

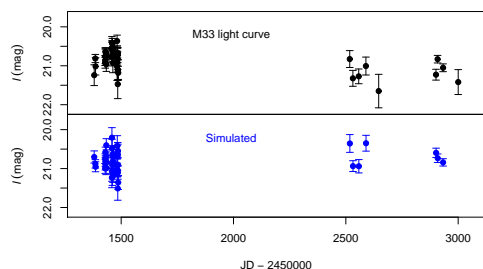


FIG. 3.— An example of simulated constant star light curve (lower) and its corresponding real light curve (upper). The errorbars for the simulated light curve indicate the “reported uncertainties”, which were rescaled before adding noise the light curve.

similarity between the distribution of χ^2_ν for the simulated light curves and that of the real M33 light curves illustrated this consistency (Figure 2). **An example of the simulated constant star light curve is given in Figure 3.**

We note that the assumption of all stars in the M33 observations being constant stars is not correct generally, because there should be some variable stars among them.

But the amount of variable stars are small relative to the entire sample of stars. An advantage of this method is that stars of bad photometry results, which may be due to bad pixels or frame edges, were also simulated.

3.3. SRV light curves

SRVs share some similarities with Miras, and thus may increase the misclassification rate for M33 Mira search. Although SRV light curves are more chaotic and non-periodic than Mira light curves, the magnitudes and the amount of magnitude variations for some SRVs are comparable to Miras. Due to the sparse nature of the M33 observations, some SRVs in M33 may be misclassified as Miras. To improve the ability of discriminating SRVs and Miras, we included simulated SRV light curves in the training data. This would also help us to assess the misclassification rate.

We simulated 5000 SRV light curves with the approach similar to the simulation of Mira light curves. The only difference is that we smooth splined the LMC SRV light curves to obtain template SRV light curves.

4. MODELS

$1 - 5lightcurves - - > Stetson'sJ > 0.75 - - > Simulateddataset$
Apply GP on simulated data set

4.1. model 1

Regression (ridge, lasso ?)

4.2. model 2

Beyes, cart, svm?

4.3. model 3

Deep learning?

5. RESULTS

REFERENCES

- Alcock, C., Allsman, R. A., Axelrod, T. S., et al. 1993, in *Astronomical Society of the Pacific Conference Series*, Vol. 43, *Sky Surveys. Protostars to Protogalaxies*, ed. B. T. Soifer, 291
- Bloom, J. 2013, *Absolute Calibration of the Extragalactic Mira Period-Luminosity Relation*, HST Proposal
- Feast, M. W. 2009, in *AGB Stars and Related Phenomena*, ed. T. Ueta, N. Matsunaga, & Y. Ita, 48
- Gerasimovic, B. P. 1928, *Proceedings of the National Academy of Science*, 14, 963
- Glass, I. S., & Evans, T. L. 2003, *MNRAS*, 343, 67
- Huemmerich, S., & Bernhard, K. 2012, *Open European Journal on Variable Stars*, 149, 1
- Kaluzny, J., Stanek, K. Z., Krockenberger, M., et al. 1998, *AJ*, 115, 1016
- Lomb, N. R. 1976, *Ap&SS*, 39, 447
- LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, *ArXiv e-prints*, arXiv:0912.0201
- Maercker, M., Vlemmings, W. H. T., Brunner, M., et al. 2015, *ArXiv e-prints*, arXiv:1512.01350
- Massey, P., Olsen, K. A. G., Hodge, P. W., et al. 2006, *AJ*, 131, 2478
- Pellerin, A., & Macri, L. M. 2011, *ApJS*, 193, 26
- Roberts, I. 1895, *MNRAS*, 56, 70
- Scargle, J. D. 1982, *ApJ*, 263, 835
- Soszyński, I., Udalski, A., Szymański, M. K., et al. 2009, *Acta Astron.*, 59, 239
- . 2011, *Acta Astron.*, 61, 217
- . 2013, *Acta Astron.*, 63, 21
- Stetson, P. B. 1987, *PASP*, 99, 191
- . 1994, *PASP*, 106, 250
- . 1996, *PASP*, 108, 851
- Udalski, A., Szymanski, M., Kaluzny, J., Kubiak, M., & Mateo, M. 1992, *Acta Astron.*, 42, 253