Measurement of the distance to the Large Magellanic Cloud using Classical Cepheids

Eliseu Antonio Kloster Filho
(Northwestern University)

June 8, 2023

Abstract

Classical Cepheids exhibit regular pulsations in their radii, which cause corresponding variations in their luminosity. The relationship between pulsating period and luminosity is called the period-luminosity relationship (PLR), and it's a global property of stars in this group. Since luminosity is also related to flux and distance measurements, Cepheids provide a way to indirectly measure distances in astronomy. In this project, we calibrate the PLR for a group of Cepheids in the galactic disk using photometric data from the Optical Gravitational Lensing Experiment (OGLE) survey and parallax data from the Gaia space observatory. Corrections for extinction due to interstellar dust are applied throughout the project using dust maps by Schlegel et al. With this method, we obtain the PLR given by $M = -(2.2 \pm 0.1)\log_{10}(P) - (2.48 \pm 0.07)$, in which M represents the absolute magnitude, and P represents the period measured in days. Subsequently, we use this calibration to measure the distance to a larger group of Cepheids in the Large Magellanic Cloud (LMC), which is also observed by OGLE. These measurements are combined using a median to obtain 43.8 ± 0.2 kpc as our best estimate for the distance to the LMC.

1 Introduction

Measuring distances in astronomy can help us to determine the scale and structure of the universe, as well as the motion and evolution of celestial bodies (Pietrzyński et al., 2019). In this context, the Large Magellanic Cloud (LMC) is vital to the field. It's a satellite galaxy of the Milky Way that, due to its proximity, allows us to compare various distance measurement methods, such as those involving Cepheids, RR Lyraes, and Eclipsing Binaries (Pietrzyński et al., 2019). It is also one of the first steps in the cosmic distance ladder, relating scales in our galaxy to those in extra galactic space. The use of the Cepheid period-luminosity relationship (PLR) to measure distances, in particular, relies on zero-point calibration procedures. Therefore, more precise calibrations can reduce measurement uncertainties. Before 2013, the best Cepheid zero-points were calculated with the help of parallax measurements from the Hubble Space Telescope Fine Guidance Sensor (HST FGS), which has a pointing precision of 2 milliarcseconds (Freedman et al., 2001). This allowed astronomers to estimate the LMC distance modulus in the range $(m-M)_0 = 18.40$–18.50 mag (Walker, 2012). More recently, the Gaia mission, launched in 2013, helped to narrow the interval down to 18.472 ± 0.091 mag (Owens et al., 2022). Gaia aims to create the largest and most precise three-dimensional map of our galaxy (Owens et al., 2022). Its high-precision parallax measurements provided the necessary data to obtain a precise PLR for Cepheid variable stars, which allowed the improvement in LMC's distance modulus.

In this project, we perform a similar measurement, contributing to the ongoing efforts to refine the LMC distance estimates. To obtain a precise calibration, we combine Gaia's parallax data and the high-quality photometric data from the Optical Gravitational Lensing Experiment (OGLE) survey (Owens et al., 2022) (Soszyński et al., 2020). OGLE's apparent magnitude observations have typical uncertainties of thousandths of a magnitude, while the smallest oscillation amplitudes for Cepheids reach tenths of a magnitude (Soszyński et al., 2020). Therefore, although OGLE's goal is to search for exoplanetary gravitational microlensing events, this level of precision photometry can be useful to analyze variable stars, especially when combined with the vast catalogue of observations made by OGLE (A. Udalski1 et al., 2015).

The idea behind using Cepheids to measure distances in this project is simple: for a set of nearby stars, we use parallaxes from Gaia to compute distances, and then combine those with apparent magnitudes from OGLE to compute absolute magnitudes. Based on the flux measurements over time provided by the OGLE survey, we also calculate the periods of oscillation. This is enough information to calibrate Leavitt's Law (Equation 1), which is the PLR (Macri et al., 2015). Afterwards, we calculate periods and apparent magnitudes for Cepheids in the LMC. Plugging these periods in the PLR yields absolute magnitudes for stars in the LMC. Since apparent magnitude, absolute magnitude, and distance are related, we can solve for the estimated distance of each Cepheid in the LMC. The combined information from all the stars analyzed is what allows us to perform the final measurement in Section 4. Of course, additional corrections and processing are necessary to obtain meaningful results. These are described in Sections 2 and 3.

2 Data

Two sets of data are fundamental for this project: (i) light-curves for all Cepheids used in the calibration and measurement procedures, (ii) parallax measurements for all Cepheids used in the calibration procedure. The first, light curves, comes from the OGLE Collection of Variable Stars (OCVS). The second, parallaxes, comes from the Gaia Data Release 3 (Gaia DR3). We also use infrared emission dust maps created by Schlegel et al. (1998). Specifically, we use the recalibration of Schlegel et al. made by Schlafly & Finkbeiner (2011) using data from the Sloan Digital Sky Survey (SDSS). In the following subsections, we describe these sources of data, their data acquisition procedures, and any pre-processing operations applied to them.

2.1 Light Curves

Light curves consist of flux measurements over time. Since 1997, a 1.3-m telescope at the Las Campanas Observatory in Chile is used by the OGLE project to collect them (A. Udalski1 et al., 2015). The observations are made available through the OCVS, the dataset used for this project.

Light curves from stars located in the Milky Way are used for calibration, since Gaia can't measure parallaxes for stars farther than 25 kpc (Lindegren, L. et al., 2018). Moreover, Gaia can rarely resolve source pairs that are less than 1 arcsec away (Fabricius, C. et al., 2021). This is not uncommon for the stars of interest in the galactic bulge (A. Udalski1 et al., 2015). Therefore, to ensure reliable parallax measurements, we choose to use only Cepheids in the galactic disk (GD) for calibration. For the distance measurement, the only restriction is that the Cepheids must be located in the LMC. In addition to these position requirements, we choose to use only I-band (i.e. near-infrared) photometry data to mitigate interstellar dust effects. Dust particles in the interstellar medium scatter and absorb shorter wavelengths more effectively than longer wavelengths. This causes the extinction and reddening of Cepheid's light, which can lead to inaccurate distance measurements. I-band observations, being in the redder part of the spectrum, are less susceptible to these effects, allowing for more precise measurements of the luminosity (Fraknoi et al., 2022).

The OGLE project classifies Cepheids into distinct groups according to their position, including two groups for those located in the GD, and in the LMC. Therefore, there is no need to sort them manually, all data can be easily downloaded from a web interface located at `https://ogledb.astrouw.edu.pl/~ogle/OCVS/`.

In total, the OCVS has 1848 stars in the GD and 4709 stars in the LMC, as of June 2023. Consolidated information for each Cepheid is made available through a downloadable table encoded in a text file. This table contains unique IDs assigned by OGLE, J2000 Right Ascension (RA), J2000 Declination (DEC), Mean Apparent Magnitude (m), and modes of pulsation. Photometry is also provided as text files, which can be downloaded as two compressed files for each group. We load and organize everything using *python* and two *pandas DataFrames* (one for calibration, other for measurement) whose rows are associated with each star and whose columns include the basic data and object references to the photometry *DataFrames*. Some stars do not have photometry in the I-band, those are immediately removed from the analysis.

2.2 Parallaxes

Parallaxes measure the difference in apparent position of an object when the Earth is at different points of its orbit. Gaia's parallax measurements can reach precision of up to 24 microseconds for the faintest sources it can detect, which is much better than previous missions (Owens et al., 2022). It also collects photometric measurements in the range 30–1050 nm using a blue and a red photometer. These are used to estimate effective temperatures and other stellar properties (Lindegren, L. et al., 2021).

Access to the Gaia DR3 database is available through the *python* package *astroquery*. For each of the stars obtained from OGLE, we query Gaia and search for sources in a 2 arcsec radius. The closest source to the input coordinates is selected. Parallax (p), parallax uncertainty (σ_p), and effective temperature (T_{eff}) measurements are downloaded from Gaia.

The guidelines presented in Lindegren, L. et al. (2018) to obtain an astrometrically "clean" subset of Gaia data suggest filtering out the sources with $p/\sigma_p < 10$, which we did. Moreover, because they need a very specific structure to exhibit the pulsation behavior, Cepheids typical effective temperatures lie in a well-defined range: 5000–7000 K (Bono et al., 2000). Those outside this range or with no temperature data were filtered out as well. Gaia data is used only for calibration, of course.

2.3 Dust Maps

To correct for the extinction effects mentioned in Section 2.1, we use dust maps. The map created by Schlafly & Finkbeiner (2011) is available through the python package *dustmaps*, and provides reddening measurements ($E(B-V)$) for all sky coordinates. OGLE's I-band filter closely reproduces the standard Kron-Cousins I-band filter (A. Udalski1 et al., 2015). The appropriate conversion factor between reddening and extinction for this filter is found on Schlafly & Finkbeiner's Table 6, and it's equal to $A/E(B-V) = 1.56$ for $R_V = 3.1$. Because the map measures the total extinction across the galaxy, we remove all data that have $A > 3$ on the basis that it's possible that the stars are in front of the dust, and we would overestimate their luminosity by considering the total extinction Schlafly & Finkbeiner (2011).

3 Methods and Analysis

Methodology is split into two steps: (i) calibration of the PLR for Classical Cepheids, (ii) use of this calibration to measure the distance to the LMC. Calibration makes use of the Cepheids in the GD. Distance measurement uses the Cepheids in the LMC. Table 1 shows the available data in each of these sections after all pre-processing operations made in Section 2. The individual procedures are described in the next two subsections.

Table 1: Data used in each of the methods and analysis sections.

	RA	DEC	m	p	σ_p	T_{eff}	Photometry
Calibration	✓	✓	✓	✓	✓	✓	✓
Distance Measurement	✓	✓	✓				✓

3.1 Calibration of the PLR

3.1.1 Period and Apparent Magnitude

We start by calculating the pulsation period of each star with a Lomb-Scargle periodogram (Lomb, 1976) and (Scargle, 1982). This method is a powerful tool for detecting and characterizing periodic signals in unevenly sampled data, making it ideal for this study (VanderPlas, 2018). It works by fitting a sinusoidal model to the data at a range of different periods, before choosing the frequency that maximizes the goodness of the fit (i.e. minimizes χ^2). To find the best fit periods through this algorithm, we used *astropy*'s built in Lomb-Scargle functions. A frequency grid ranging $\frac{1}{\Delta t}$–5 d^{-1} and an interval of $\frac{1}{5\Delta t}$ d^{-1} were used, in which Δt represents the total time span of data for an individual star in days. This guarantees we are searching for typical Cepheid periods (Fraknoi et al., 2022).

For each frequency f, the model fitted by Lomb-Scargle is shown in Equation 1. $m(t)$ is the apparent magnitude at time t. m, K, and ϕ are fit parameters. From this equation, we can extract the mean apparent magnitude m. Moreover, the χ^2 quantity that we minimize is given by Equation 2, in which the apparent magnitude measurements made by OGLE at time t_n are represented by m_n, while their uncertainties are represented by σ_{m_n}.

$$m(t) = m + K\sin(2\pi f(t-\phi)) \tag{1}$$

$$\chi^2 = \sum_n \left(\frac{m_n - m(t_n)}{\sigma_{m_n}}\right)^2 \tag{2}$$

For illustration purposes, a periodogram of a random star in the dataset is shown in Figure 1, together with its corresponding phase-folded plot in Figure 2.

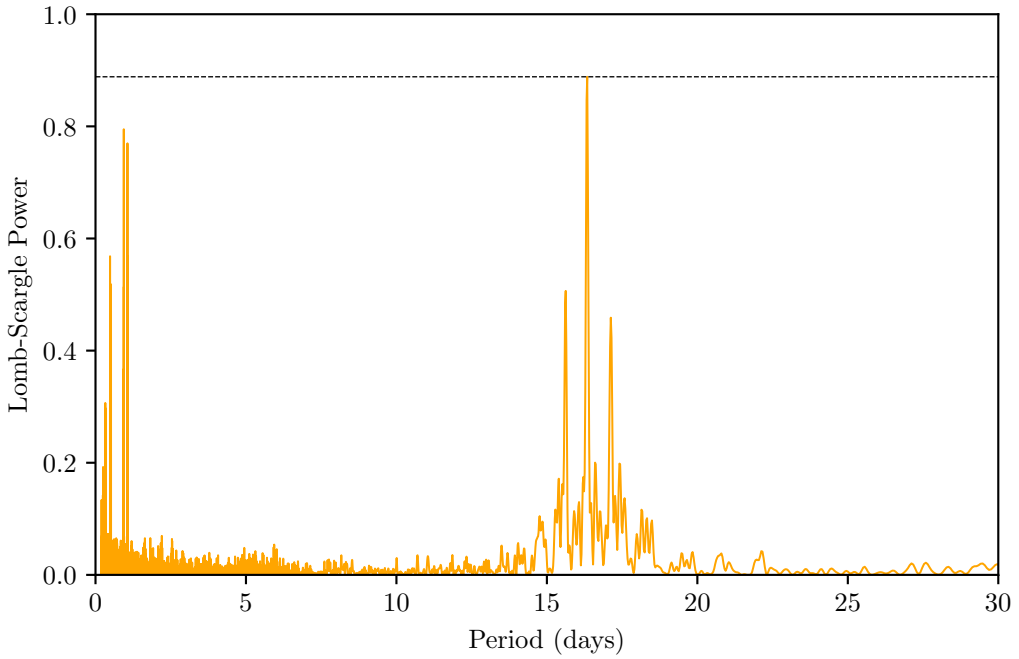

Figure 1: Periodogram for OGLE-GD-CEP-0572. The plot was zoomed around the peak of the Lomb-Scargle Power distribution. The period at which the power peaks is $P = 16.340$ d, so this is the calculated period for the star.

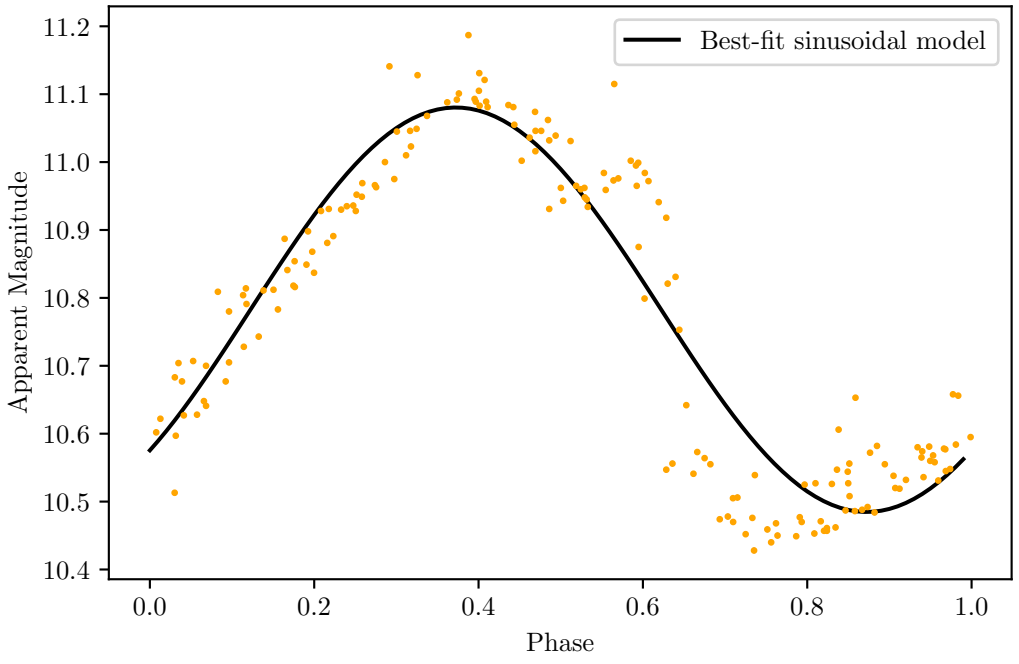

Figure 2: Phase-folded plot for OGLE-GD-CEP-0572. The best-fit sinusoidal model is given by $m(t) = 10.78 - 0.298 \sin\left(\frac{2\pi}{16.340}(t - 2.085)\right)$. This is in the same form as Equation 1.

3.1.2 Distance

Distances d measured in parsec are computed from the parallaxes p measured in milliarcseconds. The formula for this computation and the corresponding propagated uncertainty (σ_d) are shown in Equations 3 and 4, respectively. These approximations work for $p < 10^8$ mas, which is an interval that certainly includes all of our data (Fraknoi et al., 2022).

$$d \approx \frac{1000}{p} \tag{3}$$

$$\sigma_d \approx \frac{1000}{p^2}\sigma_p \tag{4}$$

3.1.3 Absolute Magnitude

We can use m, d, and A to find the absolute magnitude M through the relationship between magnitudes and extinction shown in Equation 5 (Fraknoi et al., 2022). The propagated uncertainty is shown in Equation 6.

$$M = m - A - 5\log_{10}\left(\frac{d}{10}\right) \tag{5}$$

$$\sigma_M = \frac{5}{d\ln(10)}\sigma_d \tag{6}$$

By the same reason Cepheids have a narrow temperature range, they also have a typical range of absolute magnitudes, which are no larger than -1 for most cases (Bono et al., 2000). For this reason, we remove all Cepheids with $M > -1$.

3.1.4 Model fitting

The final step is to fit the PRL model expected for Classical Cepheids. Leavitt's Law says that the absolute magnitude (M) and the logarithm of the period ($\log_{10}(P)$) satisfy a linear relationship (Macri et al., 2015). This model is expressed in Equation 7, for which m and b are fit parameters. P is the period measured in days, and M is the absolute magnitude.

$$M = m\log_{10}(P) + b \tag{7}$$

Using *scipy*'s curve fitting tools, we employ a χ^2 minimization method to fit this model. Figure 3 shows the line fitted to the data points that passed all pre-processing cuts. Equation 8 exhibits the derived PRL. This calibration equation will be used to estimate the absolute magnitudes of Cepheids in the LMC.

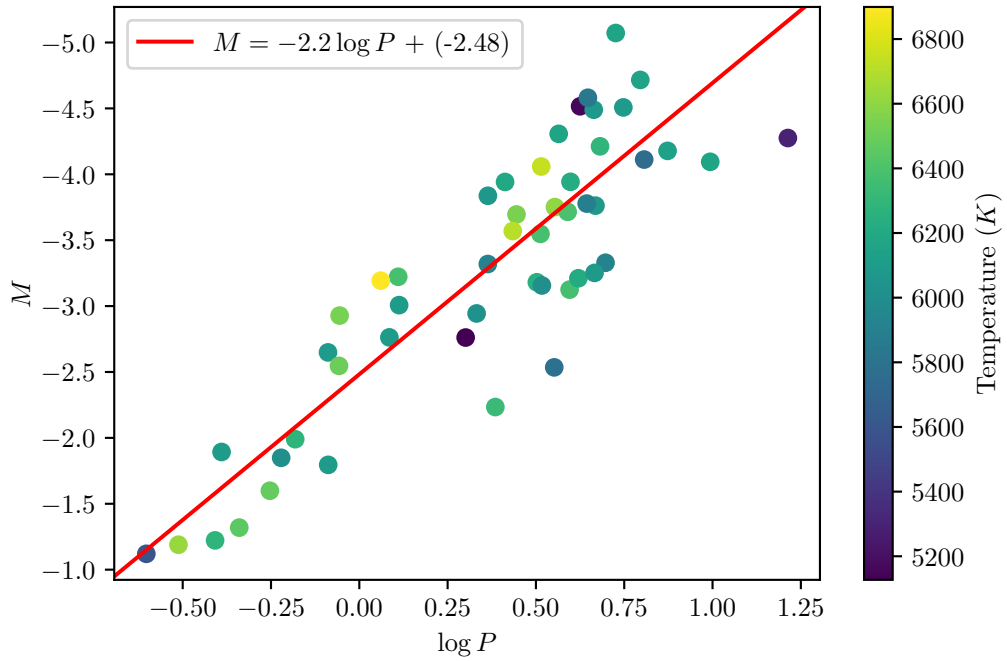

Figure 3: Best fit PLR and the 50 calibration Cepheids. The line equation is given by Equation 8.

$$M = -(2.2 \pm 0.1) \log_{10}(P) - (2.48 \pm 0.07) \tag{8}$$

3.2 LMC Distance Measurement

3.2.1 Period and Apparent Magnitude

Period and apparent magnitude are measured in the exact same way as in the calibration phase.

3.2.2 Absolute Magnitude

Absolute magnitude is computed from the PRL derived in Equation 8. The uncertainty on M is σ_M and computed according to Equation 9. The uncertainties on the coefficients m and b are represented by σ_m and σ_b, respectively.

$$\sigma_M = \sqrt{(\log_{10}(P)\sigma_m)^2 + \sigma_b^2} \tag{9}$$

For the same reasons outlined in subsection 3.1.3, we remove all Cepheids with $M > -1$.

3.2.3 Distance

From the apparent magnitude, absolute magnitude, and extinction values we obtained, the distance to LMC Cepheids and its uncertainty can be estimated using Equations 10 and 11, respectively. Equation 10 is just a rearrangement of Equation 5.

$$d = 10 \cdot 10^{\frac{m-M-A}{5}} \tag{10}$$

$$\sigma_d = 2\ln(10) \cdot 10^{\frac{m-M-A}{5}} \sigma_M \tag{11}$$

3.2.4 Final Measurement

Finally, we combine all the distance measurements using a median. The median is a robust estimator that is less sensitive to the tails of the distribution, where other physical regimes might be in action (Bohm & Zech, 2014). The uncertainty is estimated using a bootstrap method (Bohm & Zech, 2014). This technique involves sampling the data, with replacement, multiple times and calculating the sample median each time. From the resulting distribution of medians, we can calculate the standard deviation, which is the estimated uncertainty. In this project, the sample size used was the same as the data (4510). The number of samples performed was 1000. We noticed that values larger than 1000 did not affect the final results after significant figures were taken into consideration.

4 Results

Using the methodology developed in Section 3, we obtain the distribution of distances shown in Figure 4. The median was calculated using the bootstrap method, resulting in our best estimate for the distance to the LMC:

$$d_{\text{LMC}} = 43.8 \pm 0.2 \text{ kpc}$$

The corresponding distance modulus is:

$$\mu_{\text{LMC}} = 18.20 \pm 0.01 \text{ mag}$$

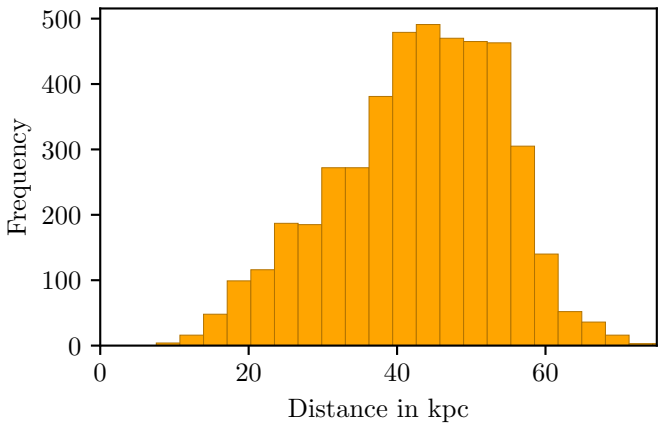

Figure 4: Histogram of the measured distances to the LMC.

This measurement has important consequences in astronomy. It is a first step in understanding the scales of our universe. Distances to other galaxies can be computed

relative to the LMC, so it is important that this measurement is made with accuracy and precision. Moreover, the distance to the LMC serves as a crucial benchmark for various astronomical studies and observations. For example, we can calibrate other methods used to measure astronomical distances using the value we found for the distance to the LMC (Pietrzyński et al., 2019). Finally, precise distance measurements to the LMC contribute to understanding its orbital motion around the Milky Way and the gravitational interactions between the two galaxies. This information helps refine models of galaxy formation and evolution, shedding light on the processes involved in the growth and structure of galaxies (Alves, 2004) and (Mastropietro et al., 2005).

5 Conclusions

New instrumental advancements provided by Gaia open room for improvements in techniques that depend on parallax measurements. In this project, we combined the Gaia DR3 dataset and the OCVS to calibrate the PLR for Classical Cepheids and subsequently measure the distance to the LMC. We made use of parallaxes, effective temperatures, Lomb-Scargle periodograms, dust maps, uncertainty propagation formulas, and the relationship between apparent magnitude, absolute magnitude, and distance. In the end, we consolidated all the individual distances to each Cepheid into a single best estimate of the distance to the LMC. Further work could be done to improve our measurement. For example, Owens et al. explain that metallicity effects on Cepheid luminosity are still poorly understood, and can lead to systematic errors in the measurement. Moreover, instead of using dust maps, we could have used individual extinctions available in the literature for each Cepheids, which could provide additional precision (Owens et al., 2022).

References

A. Udalski1, M.K. Szymański, & G. Szymański. 2015, Acta Astronomica, 65, 1

Alves, D. R. 2004, New Astronomy Reviews, 48, 659

Bohm, G., & Zech, G. 2014, Introduction to statistics and data analysis for physicists (Hamburg: DESY)

Bono, G., Castellani, V., & Marconi, M. 2000, ApJ, 529, 293

Fabricius, C., Luri, X., Arenou, F., et al. 2021, A&A, 649, A5

Fraknoi, A., Morrison, D., & Wolff, S. 2022, Astronomy 2e (OpenStax)

Freedman, W. L., Madore, B. F., Gibson, B. K., et al. 2001, ApJ, 553, 47

Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, A&A, 616, A2

Lindegren, L., Klioner, S. A., Hernández, J., et al. 2021, A&A, 649, A2

Lomb, N. R. 1976, Ap&SS, 39, 447

Macri, L. M., Ngeow, C.-C., Kanbur, S. M., et al. 2015, AJ, 149, 117

Mastropietro, C., Moore, B., Mayer, L., et al. 2005, Mon. Not. Roy. Astron. Soc., 363, 509

Owens, K. A., Freedman, W. L., Madore, B. F., & Lee, A. J. 2022, ApJ, 927, 8

Pietrzyński, G., Graczyk, D., Gallenne, A., et al. 2019, Nature, 567, 200

Scargle, J. D. 1982, ApJ, 263, 835

Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Soszyński, I., Udalski, A., Szymański, M., et al. 2020, Acta Astronomica, 70, 101–119

VanderPlas, J. T. 2018, The Astrophysical Journal Supplement Series, 236, 16

Walker, A. R. 2012, Astrophysics and Space Science, 341, 43