Measuring the Distance to the Large Magellanic Cloud using Cepheid Variables

Kyle Soni
(Northwestern University)

June 8, 2023

Abstract

The Large Magellanic Cloud (LMC) is a nearby satellite galaxy of the Milky Way. Measuring its distance from Earth is important for understanding the broader structure of the universe and how galaxies interact. Since the LMC contains many Cepheid variables, the fixed period-luminosity (P-L) relationship of Cepheids is useful in calculating this distance. To calibrate this relationship, I used parallax measurements from the Gaia mission to find distances to Cepheids in the galactic disk. I then used light curves from the Optical Gravitational Lensing Experiment (OGLE) to find the pulsation period (via Lomb-Scargle) and apparent magnitudes of these Cepheids, successfully calibrating the P-L relationship as $M = (-2.15 \pm 0.04) \log(P) + (-2.35 \pm 0.01)$, where M is absolute magnitude and P is period. Using OGLE light curves from LMC Cepheids, I found their pulsation periods, allowing me to calculate their absolute magnitudes as well. Combining this with the apparent magnitude data from OGLE yielded distance measurements for these stars. The median of these measurements was 41.24 ± 2.23 kiloparsec (kpc), showing that the distance to the LMC is comparable to the size of the Milky Way.

1 Introduction

Measuring distance to astronomical objects is a fundamental task in astronomy. With accurate distances, we can understand more about the properties of objects, how they interact with other objects, and how they fit into the large-scale structure of the universe. To achieve this, it is useful to construct a "cosmic distance ladder" with the distances to several key objects, such as stars outside of the solar system or galaxies outside of the Milky Way. One ideal candidate for this distance ladder is the Large Magellanic Cloud (LMC), a satellite galaxy of the Milky Way. Due to its close proximity to the Milky Way, calculating the distance to it will serve as a useful benchmark for understanding the scale of galaxies and their neighbors.

The most popular method for measuring distance (dating back to ancient Greece) uses parallax measurements, which are apparent shift in the position of an object due to the Earth's motion around the Sun (Benedict et al., 2007). With two observation times at different times of the year, we can measure the parallax of an object and use it in the following equation:

$$d = \frac{1}{p} \tag{1}$$

where d is distance (measured in parsecs) and p is parallax (measured in arcseconds). Because the distance to an object is much greater than the distance between the two observation points on Earth, we can approximate $\tan(p) \approx p$.

However, an important limitation for parallax measurements is their lack of accuracy for long distances. Since parallax and distance are inversely proportional, parallax measurements for far away objects are too small for modern instruments to achieve a high signal-to-noise ratio. Consequently, this method is impossible for calculating distance to the LMC directly. Previous work has used other options to perform this measurement, including the use of type 1a supernovae or eclipsing binary stars (Pietrzyński et al., 2013). However, because the LMC is home to a large amount of Cepheid variables, calculating distance using Cepheids will provide a more precise measurement.

Cepheid variables are a type of star that pulsates radially, changing brightness at a well-defined stable period. In 1908, Henrietta Swan Leavitt discovered that there is a relationship between the period of pulsation of a Cepheid variable and its brightness (Leavitt, 1908). This Period-Luminosity (P-L) relationship is outlined by the following equation:

$$M = \delta \log(P) + \rho \tag{2}$$

where M is the absolute magnitude, P is the period, and δ and ρ are the slope and intercept respectively (Lanoix et al., 1999). As shown, this relationship needs to be calibrated with known distances, and it will vary depending on the identity of the Cepheid (classical or Type II). However, this relationship is believed to be relatively consistent across Cepheids of the same type, meaning we can calibrate using stars in the Milky-Way that have parallax measurements and known distances. We can then use this relationship along with measurements of the pulsation periods of Cepheids in the LMC to find their absolute magnitudes. Since absolute and apparent magnitudes are related by distance, this allows for a direct calculation of distance to the LMC.

2 Data

There are two important data sets for measuring the distance to the LMC. The source of parallax measurements is the Gaia mission, which measures positions, distances, and motion of approximately one billion stars in the Milky Way Galaxy. The Gaia spacecraft is equipped with two telescopes, along with astrometric, photometric, and spectroscopic instruments, allowing it to provide the most accurate parallax measurements currently available (Gaia Collaboration et al., 2016). I acquired this data through the `astroquery.gaia` python package, which allows for direct selection from the data release. I used the Gaia Data Release 3 (GDR 3) since it was the most recent and complete dataset. As a preprocessing step, I excluded any star with a parallax signal-to-noise less than 10 ($\frac{p}{\sigma_p} < 10$). This was based on the suggestion of Gaia researchers themselves, who stated that signals with less than that threshold are unlikely to be accurate (Eyer et al., 2022).

The other data set was from the Optical Gravitational Lensing Experiment (OGLE), a long-term astrophysical project that studies various phenomena with ground-based telescopes. They provide light curves on thousands of Cepheid Variables in the Milky Way, LMC, and Small Magellanic Cloud (SMC) through their Collection of Varaiable Stars (Paczynski et al., 1996). The light curves provide measurements of the apparent magnitudes of different stars over time, allowing us to examine the period of oscillation in the luminosity. The photometry data in these light curves is directly accessible and available for download on their website.

To narrow down the OGLE data set, I made a few important choices. The first was to look at classical Cepheids (as opposed to Type II). The different Cepheid types have different P-L relationships, so they must be calibrated separately. In addition, classical Cepheids tend to be brighter and more common in both the Milky Way and LMC, making them better candidates for study. Similarly, I chose to focus on the I band for photometry measurements, since OGLE had subtstantially more photometric measurements for that wavelength. For the Milky Way galaxy, stars in the bulge were excluded, leaving only stars in the galactic disk as the main focus. Due to a large presence of dust is the bulge of Milky Way, measurements from those stars tend to have large noise when compared to stars in the disk.

However, stars in the LMC and disk are still subject to extinction/reddening from interstellar dust, so I needed dust maps to properly correct for these. I retrieved the necessary dust maps from the galactic Dust Reddening and Extinction Service provided by Caltech. Their maps are based on data from Schlafly and Finkbeiner's analysis of the Sloan Sky Survey in 2011 (Schlafly & Finkbeiner, 2011).

3 Methods and Analysis

The first step was calibrating the period-luminosity relation using stars in the Milky Way. Before performing any analysis, it was necessary to properly match the data between Gaia and OGLE. To do this, I used `astroquery.gaia` to find the closest match in gaia for a set of right ascension and declination coordinates. Running this on the set of coordinates from disk stars in OGLE yielded the corresponding parallax measurements and errors for those stars. If the angular distance between the OGLE coordinates and best Gaia match was greater than 1e-4, the star was discarded as it is likely that they are different astronomical objects. However, this only affected three stars. With these parallax measurements, I used Equation 1 to find the distance for each star and a corresponding uncertainty.

With the distance, I'm able to calculate the absolute magnitude for each star. I started with averaging the photometry data in the I band for each star to get the average apparent magnitude and a corresponding uncertainty. I then found the dust correction using the dust maps provided by Caltech. Inputting the right ascension and declination for a star gave a value for $E(B-V)$, the interstellar reddening of the star. This is directly related to the extinction with the equation $\frac{A_I}{E(B-V)} \approx 1.516$, where A_I is the total extinction in the I band (Schlafly & Finkbeiner, 2011). Subtracting A_I from the apparent magnitude (m) corrects for the extinction. With all these values, I was able to calculate the absolute magnitude (and corresponding uncertainty) using the equation:

$$M = m - A_I + 5 - 5\log(d) \tag{3}$$

where M is absolute magnitude, m is apparent magnitude, A_I is the total extinction, and d is distance. When looking at these absolute magnitude calculations, I found a few key outliers that are likely due to errors from OGLE and the dust maps. First, some of the stars had a magnitude greater than -1 or less than -20. Very few Cepheid variables have a magnitude greater than -1 so these stars are likely classified incorrectly as Cepheids by OGLE and were excluded from this calibration (Vilardell et al., 2007). For the extremely bright stars, some calculations of A_I from the dust map yielded magnitudes of 5-15, creating these extremely bright absolute magnitudes. Therefore, it is likely that these dust maps are inaccurate, so any star with an $A_I > 3$ was excluded as well.

To complete the period-luminosity relation calibration, the pulsation period for the Cepheids is necessary. I started with the photometry data from OGLE in the I band for each star. With `astropy`, I used the Lomb-Scargle periodogram to find a best-fit period. The Lomb-Scargle algorithm is a method for characterizing periodic signals in unevenly spaced data, using a variation of the classical periodogram (VanderPlas, 2018). This period was then checked with the literature (OGLE calculated period) as well as visually with a phase-folded light curve. An example for this process is shown in Figure 1.

When performing these calculations, 19 stars had periods that did not match OGLE. Upon closer inspection, most stars were able to get a well-fitting period with a recalculation using different parameters, so they were included in the data (period differed from OGLE in two cases). In three cases, neither the Lomb-Scargle or OGLE period produced a periodic phase-folded curve, so these three stars were excluded. After all the filtering, I used a set of 133 stars to perform the calibration.

Using a least-squares linear fit (via `scipy`) between the absolute magnitudes (M) and log-scale period ($\log(P)$), I calibrated the values for the constants in Equation 2, finding a relationship of:

$$M = (-2.15 \pm 0.04)\log(P) + (-2.35 \pm 0.01) \tag{4}$$

A graphical representation of the P-L relationship along with the best-fit line is found in Figure 2. As shown, most stars are within two standard deviations of the best-fit

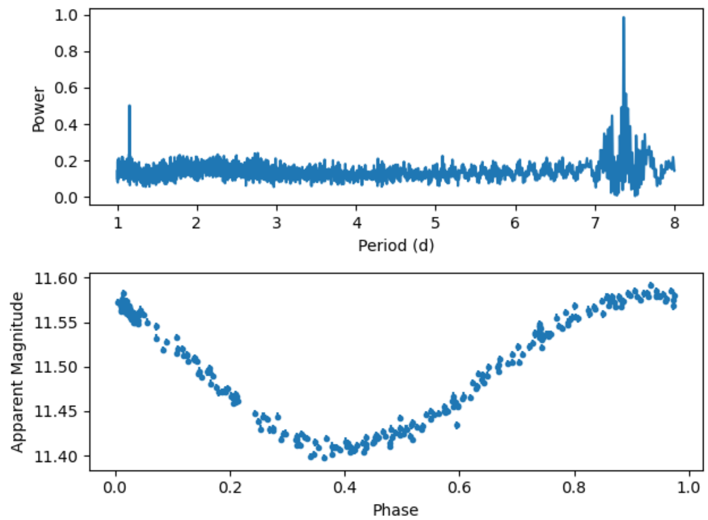

Figure 1: An example calculation of period using Lomb-Scargle on 'OGLE-GD-CEP-0812' from the Galactic disk. (Top) Lomb-Scargle Periodogram. (Bottom) Phase-folded light curve in I band (period = 7.36 d).

line. There are a few outliers with dim magnitudes, but the overall uncertainty in this calibration is low.

To find the distance to the LMC, I used the calibrated P-L relationship and the periods of Cepheids in the LMC. The process of calculating those periods was identical to those in the galactic disk, using Lomb-Scargle periodograms on 4652 stars. Once I had the periods, I calculated the absolute magnitudes using Equation 4. For the distance calculation I used the apparent magnitude measurments (corrected for dust extinction) from the OGLE light curves. I used the same criteria for exclusion, removing stars with $A_I > 3$, $M > -1$, or $M < -20$. This left 4386 stars, which I used to calculate the distance to the LMC using Equation 3. Since I used the median as the end approximation, the uncertainty was calculated using $\Delta x_{median} \approx \Delta x_{mean}\sqrt{\pi/2}$, taking into account the uncertainties in the P-L relationship as well as the uncertainties in the apparent magnitude measurements from OGLE.

4 Results

I found the distance to the LMC to be 41.24 ± 2.23 (kpc). This was based on the median of all distances calculated, and the distribution of these distances is shown in Figure 3. However, as shown in the histogram, there appears to be many stars with much lower distances than the others. This could mean that these stars are not in the LMC and are mislabeled by OGLE. However, further research would into the identity of these stars is necessary, so the best interpretation of the current result is that it could be an underestimate of the distance.

Regardless, this result has important implications for understanding galaxy interactions. The galactic disk of the Milky Way is only about 25 kpc across, so the size of the galaxy is comparable to its distance to the LMC (Goodwin et al., 1998). When looking at stars, the radius of the sun is 2.25e-8 pc, but the distance to its nearest

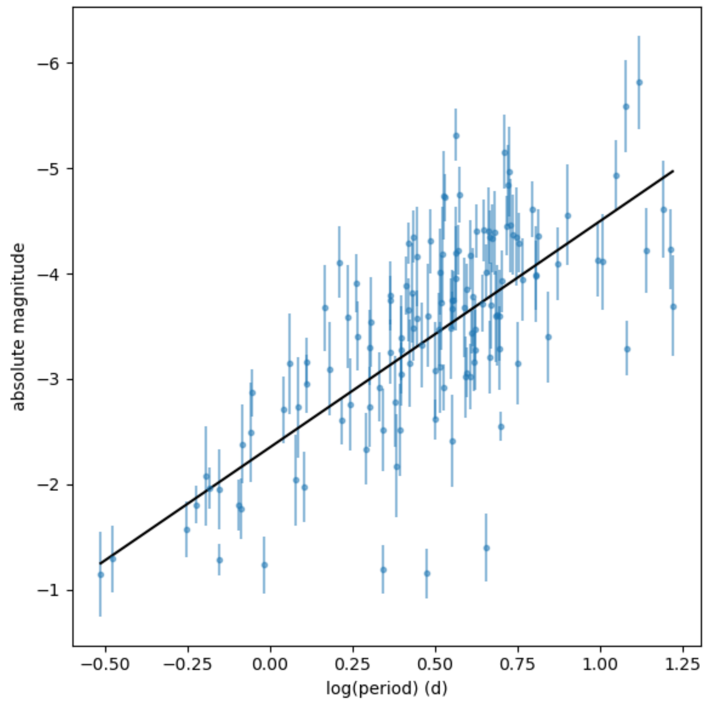

Figure 2: Graphical representation of calibrated Period-Luminosity relation. The fitted line follows equation $M = (-2.15 \pm 0.04)\log(P) + (-2.35 \pm 0.01)$.

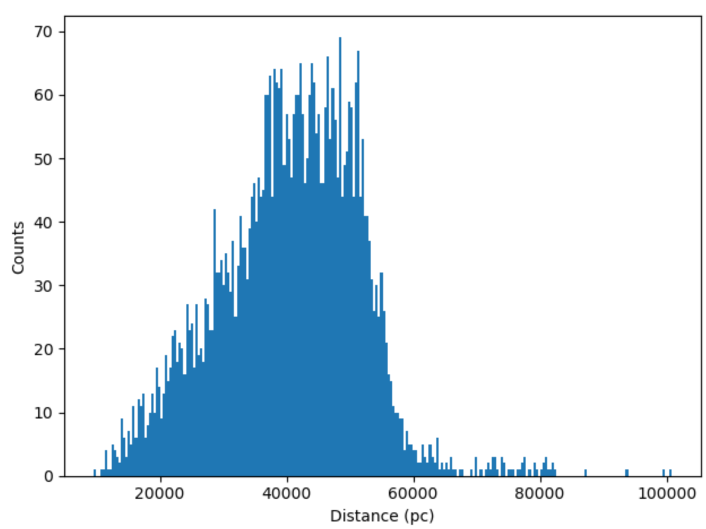

Figure 3: Distribution of distances to stars labeled as in the LMC by OGLE.

star Proxima Centauri is 1.30 pc. This means that galaxies are much closer together than stars when compared to their size, supporting the idea that galaxies could merge. We know that stars have a negligible chance of colliding unless they are in a binary, but previous observations of elliptical galaxies and irregular galaxies show evidence that galaxies are different (Pearson et al., 2019). Since this measurement shows that galaxies are relatively close compared to their size, it provides further evidence for the likelihood of galaxy collision events.

5 Conclusions

The distance to the LMC is an important step in constructing the "cosmic distance ladder." Due to the presence of many variable stars in the LMC, calculating distance using the fixed period-luminosity relationship of Cepheid variables provides an accurate measurement. Using parallax measurements from Gaia and light curves from OGLE, I calibrated the P-L relationship of Cepheids using stars in the galactic disk. Parallax angles provided the distance measurements, and the light curves provided apparent magnitudes. I then used this relationship to determine absolute magnitudes and distances to stars in the LMC. The median distance was 41.24 ± 2.23 (kpc), which is likely an underestimation due to the possibility of OGLE mislabeling closer stars as part of the LMC.

Future work should seek to fortify and extend our knowledge of distances to other galaxies. One limitation of this work is the assumption that the P-L relationship is the same for Cepheids in the galactic disk and for Cepheids in the LMC. Further research could seek to test this assumption with a calibration of Cepheids from more diverse locations. To extend this work, future research should measure distances to other galaxies. It is possible that only a few galaxies are nearby the Milky Way, meaning that galaxies tend to form clusters. Distance measurements to other galaxies could test this hypothesis and continue to build on our knowledge of the universe's structure.

References

Benedict, G. F., McArthur, B. E., Feast, M. W., et al. 2007, AJ, 133, 1810

Eyer, L., Audard, M., Holl, B., et al. 2022, arXiv, arXiv:2206.06416

Gaia Collaboration et al. 2016, A&A, 595, A1

Goodwin, S. P., Gribbin, J., & Hendry, M. A. 1998, The Observatory, 118, 201

Lanoix, P., Paturel, G., & Garnier, R. 1999, Monthly Notices of the Royal Astronomical Society, 308, 969

Leavitt, H. S. 1908, Annals of Harvard College Observatory, 60, 87

Paczynski, B., Stanek, K. Z., Udalski, A., et al. 1996, 169, 93

Pearson, W., Wang, L., Trayford, J., et al. 2019, Astronomy Astrophysics, 626

Pietrzyński, G., Graczyk, D., Gieren, W., et al. 2013, Nature, 495, 76

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

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

Vilardell, F., Jordi, C., & Ribas, I. 2007, Astronomy and Astrophysics, 473