Measuring distance to the open cluster Berkeley 51 using Be 51 #162 Cepheid data

Course Project, PH556: Astrophysics

by

Sumit Kumar Adhya

(Roll No. 23B1806)

Team: Hanle Hunters

Department of Physics

Indian Institute of Technology Bombay

Mumbai - 400076, India

Contents

1 Introduction and Objectives **2**

1.1 Target: Be #162 2

1.1.1 Target Selection and Rationale 2

1.1.2 Observations 3

2 Methods: PSF and Data Analysis **4**

2.1 PSF Photometry 4

2.2 Light Curve Fitting and Period Extraction 5

2.3 Error of Period (Using Bootstrap) 5

3 Calculations **7**

3.1 Distance Calculation 7

3.2 Error Analysis 7

4 Results **9**

Chapter 1

Introduction and Objectives

Determining the scale of the universe is a fundamental goal of astrophysics. This is built upon the "Cosmic Distance Ladder," [10] a succession of methods used to measure distances to increasingly remote objects. Near the base of this ladder, and crucial for calibrating all further steps, are the Classical Cepheid variable stars.

A Cepheid variable is a radial pulsator in the instability strip of the H-R diagram [10]. These stars rhythmically expand and contract, causing periodic changes in their brightness. This pulsation is driven by the κ-mechanism [10], which involves opacity changes in the helium ionization zones within the star [10]. Because their pulsation period is directly linked to their intrinsic luminosity, they serve as one of the most reliable "standard candles" available to astronomers.

The cornerstone of their use is the Period-Luminosity (P-L) relation, first discovered by Henrietta Swan Leavitt [13]. This tight correlation means that by simply measuring the pulsation period of a Cepheid, one can determine its absolute magnitude (intrinsic brightness). Classical (Type 1) Cepheids, like the one in this study, are young, metal-rich Population I stars. For our analysis, we adopt the P-L calibration for the r-band [2]:

$$M_r = (-2.637 \pm 0.106)(\log_{10} P - 1) + (-4.088 \pm 0.032) \tag{1.1}$$

Once the absolute magnitude (M_r) is known from the period, we can compare it to the star's measured apparent magnitude (m_r), corrected for interstellar extinction (A_r). This yields the distance modulus, $\mu_0 = m_r - M_r - A_r$, and thus the distance to the star.

In this work, we aim to apply this technique to measure the distance to the open cluster Berkeley 51. Our target is the classical Cepheid **Be 51#162**, a confirmed member of the cluster [11]. By obtaining new time-series photometry of this star, we will characterize its pulsation, derive its period, and apply the P-L relation to determine an independent distance to the cluster, which is known to host several other interesting massive stars [8].

1.1 Target: Be #162

1.1.1 Target Selection and Rationale

Two primary observing strategies were initially considered: (i) targeting Cepheids in the nearby M31 (Andromeda) galaxy, and (ii) targeting Cepheids within Galactic star clusters. A review of catalog data indicated that most Cepheids in M31 are fainter than our 20-mag limit and suffer from severe crowding and high background surface brightness. These factors would degrade photometric precision with a 1-m class instrument, rendering M31 unsuitable.

Consequently, the search was restricted to Galactic Cepheids, specifically within open clusters. Open clusters are less crowded than globular clusters, reducing the risk of blending and saturation, which is ideal for accurate point-spread-function (PSF) photometry. We queried major catalogs (SIMBAD, Pan-STARRS, VizieR) with the following criteria:

- Apparent magnitude in the 15-19 range.
- Pulsation period ≤ 10 days (to match our time allocation).
- Declination $\delta > 30°$ for adequate elevation at Hanle.
- Visibility during the 15-30 October run.
- Minimum lunar separation $> 30°$.

These filters identified the classical Cepheid **Be51#162** in the open cluster **Berkeley 51** as the highest-priority candidate. This target has a known pulsation period of ≈ 9.8 days and secure cluster membership [4]. Berkeley 51 is in a relatively sparse region of the Perseus arm, providing a well-characterized environment suitable for this study. The target meets all magnitude and visibility constraints.

1.1.2 Observations

We requested time-series imaging of Be 51#162 to measure its period and derive a distance. The observation parameters were as follows:

- **Coordinates:** $RA = 20:11:51.18$, $Dec = +34:24:47.2$ (J2000)
- **Date and Time:** '2025-10-13T15:43:19', '2025-10-15T19:02:31', '2025-10-16T14:53:04', '2025-10-19T14:21:15', '2025-10-20T13:59:25', '2025-10-23T14:38:15', '2025-10-24T15:32:45', '2025-10-26T13:00:50', '2025-10-27T17:32:41', '2025-10-28T16:39:00', '2025-10-29T16:54:14', '2025-10-30T14:37:14', '2025-10-31T15:20:41'
- **Filter:** SDSS r
- **Sequence:** 100s exposures per visit

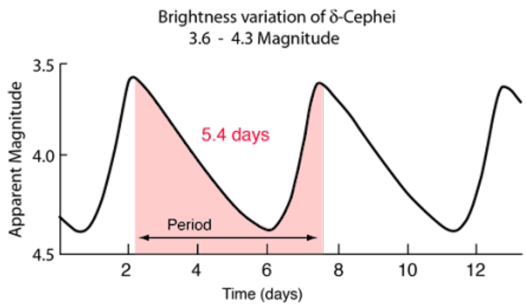

Figure 1.1: Brightness Variation of a Cepheid variable
Image credit: http://hyperphysics.phy-astr.gsu.edu/hbase/Astro/cepheid.html

Chapter 2

Methods: PSF and Data Analysis

2.1 PSF Photometry

To get the brightness measurement for our Cepheid, we first had to process the raw FITS images. The real challenge here is that our target is in a crowded open cluster, meaning other stars are very close to it and their light can overlap. The best way to handle this is with **Point Spread Function (PSF)** photometry. This technique isn't as simple as just drawing a circle; instead, it involves building a precise model of what a star "looks like" in our images, and then fitting that model to our target to measure its brightness.

Our photometric pipeline, which we built in a Python Jupyter Notebook using the `astropy` and `photutils` libraries, was a step-by-step process:

1. **Data Ingestion:** First, we had to loop through all the FITS image files. For each one, we'd open the file, read the 2D pixel data, and—most importantly—pull the time of observation (in Modified Julian Date, or MJD) from the FITS header. This MJD timestamp is what we'd eventually plot our brightness against.

2. **Background Estimation:** Images from a telescope are never truly black. There's always a background sky level. We estimated this for each image by calculating a sigma-clipped median of all the pixel values. Using a sigma-clipped median is key, as it ignores the very bright pixels from the stars themselves and gives a robust measure of just the background "glow."

3. **Building the PSF Model:** This is the core of the technique. We used a group of well-behaved relatively isolated stars to build an average, high-signal-to-noise model of the star shape (the PSF) for each image. This model accounts for all the distortions from the atmosphere and telescope optics at that specific moment.

4. **Identifying Target and References:** We then created a list of coordinates for our target star (Be 51 #162) and a separate list for our "comp" or reference stars. We chose a set of around 50 non-variable stars in the same field of view to act as our stable reference.

5. **Fitting and Measuring Flux:** With the PSF model and star list, we used `photutils` to perform the photometry. For every star in our list (target and references) on every single image, the code would fit our PSF model at that star's location. The output of this fit isn't just a count, but a precise calculation of the total flux (or brightness) coming from that star, separating its light from the background and any nearby faint stars.

6. **Differential Photometry:** This is the step that makes the data usable. The raw flux of a star can change wildly from one image to the next, simply because of thin clouds or changing atmospheric transparency. To fix this, we performed *differential photometry.* For

each image, we created a stable "ensemble" flux by summing up the flux of our reference stars. We then calculated our target's magnitude *relative* to this stable ensemble. This trick cancels out all the atmospheric effects, leaving us with only the true, intrinsic variation of the Cepheid.

7. **Light Curve Generation:** Finally, we took the resulting differential magnitude for our target from every image and plotted it against the MJD we extracted back in step 1. The resulting graph is the *light curve*, and it's the final product of our photometric analysis. This time-series plot, showing the clear "rise and fall" of the Cepheid's brightness, is the raw material we fed directly into the period-finding analysis.

2.2 Light Curve Fitting and Period Extraction

Once we had our final light curve (a set of times, magnitudes, and magnitude errors), the next crucial step was to find the star's pulsation period. A quick look at our data shows that the observation times are not evenly spaced; we have data from different nights with gaps in between. This means we can't use a standard Fast Fourier Transform (FFT).

The correct tool for this job is the **Lomb-Scargle Periodogram**, a powerful algorithm designed specifically for finding periodic signals in unevenly-sampled data. We used the `LombScargle` implementation from the `astropy.timeseries` library.

Our first step was to prepare the data. We converted the observation timestamps into a simpler NumPy array, `t`, representing the time in days since the first observation. We also loaded our magnitudes and their corresponding photometric errors into arrays called `mag` and `dmag`.

With the data ready, we initialized the `LombScargle` object. Crucially, we passed it all three arrays: `t`, `mag`, and `dmag`. By providing the `dmag` error array, we're telling the algorithm to perform a weighted analysis—it pays more attention to the data points we are more certain of (those with small error bars) and is more skeptical of noisier measurements.

We then used the `autopower` function to scan for potential periods. We limited our search to a sensible range of frequencies, from 0.05 to 1.0 cycles/day (corresponding to periods between 1 and 20 days), which is the expected range for a Cepheid of this type. This function returned an array of frequencies and their corresponding "power" (a measure of how well a sine wave at that frequency fits the data).

Finding the best period was then straightforward: we just found the frequency that produced the highest power peak using `np.argmax(power)`. The period, of course, is just the inverse of this best frequency ($P = 1/f$). To visually confirm our result, we used the `ls.model` function to plot the best-fit sine wave over our original data points, which showed an excellent match.

2.3 Error of Period (Using Bootstrap)

Just finding a single number for the period isn't enough for good science; we absolutely needed to know our uncertainty. The best way to get a robust, trustworthy error bar from a small, unevenly-sampled dataset like ours is to use a **bootstrap analysis**.

The idea behind bootstrapping is to simulate what our result might look like if we had collected our data on slightly different nights. We did this by re-running our entire period analysis 1,000 times.

In each of those 1,000 iterations, we created a new "fake" dataset by randomly sampling from our own original data points. The key is that we sampled with replacement, meaning a single bootstrap dataset might have (for example) two copies of our third data point and zero copies of our seventh. This `np.random.choice` function gave us a new set of arrays: `t_boot`, `mag_boot`, and `dmag_boot`.

We then ran the entire Lomb-Scargle periodogram analysis (initializing, finding the power, and getting the peak) on this new, slightly-different resampled dataset. This gave us a "best period" for that specific iteration, which we saved to a list.

After the loop finished, we had a list of 1,000 different "best periods." This list represents the distribution of results we could expect given our data's inherent noise and sampling. To get our final one-sigma error, we simply calculated the standard deviation (`np.std`) of this list. This standard deviation, σ_P, is our final, empirically-derived uncertainty for the period.

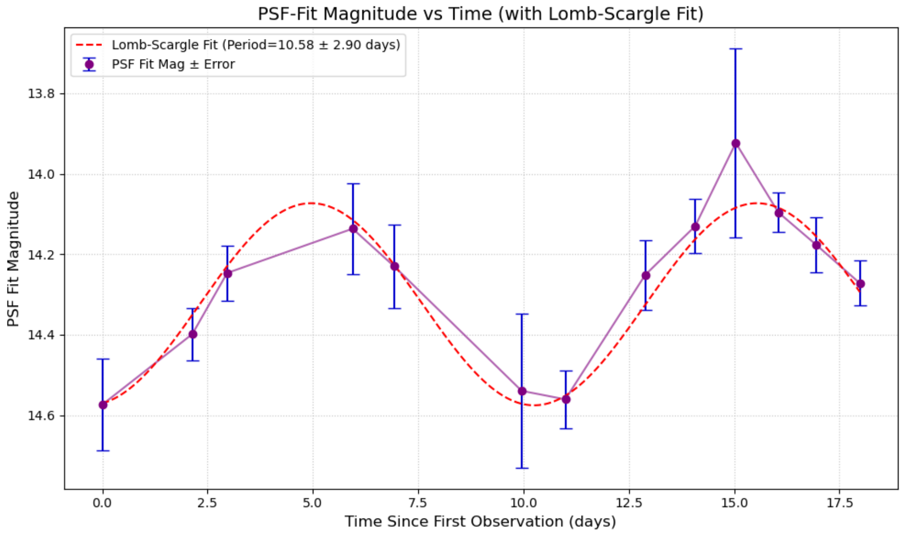

Figure 2.1: Lomb-Scargle Fit

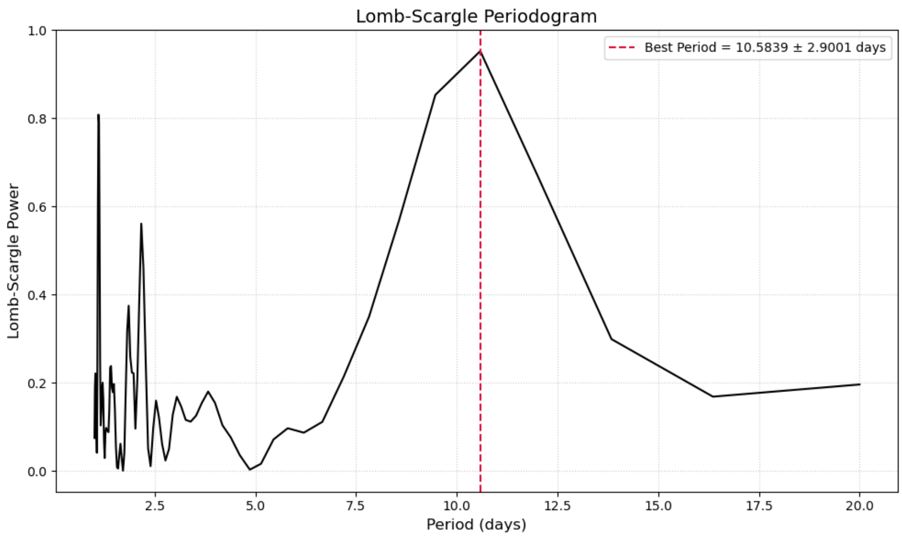

Figure 2.2: Lomb-Scargle Periodogram

Chapter 3

Calculations

3.1 Distance Calculation

From our photometric analysis, obtaining the light-curve followed by Lomb-Scargle fitting we obtain the period to be:

$$P = 10.58 \text{ days} \tag{3.1}$$

Plugging this into the P-L relation for r-band: [2]

$$M_r = a_r(\log_{10} P - 1) + b_r \tag{3.2}$$

where $a_r = -2.637$ and $b_r = -4.088$ [2], we obtain the absolute magnitude:

$$M_r = -4.153 \tag{3.3}$$

We take the error weighted mean of the observed apparent magnitudes to get:

$$m_r = 14.255 \tag{3.4}$$

The total extinction A_r is calculated from the reddening $E(B-V)$ using the relation [3]:

$$A_r = C_r \times E(B-V) \tag{3.5}$$

where $C_r = 2.285$ [5] and $E(B-V) = 1.79$ [4], which gives:

$$A_r = 4.09 \tag{3.6}$$

Now plugging all these values in the formula for distance [6]:

$$d = 10^{\frac{m_r - M_r - A_r + 5}{5}} \text{ pc} \tag{3.7}$$

we get:

$$d = 7.305 \text{ kpc} \tag{3.8}$$

3.2 Error Analysis

Using the expressions we have referred to so far and the standard error propagation formula, we obtain the following relations for the errors.

The fractional error in the distance, σ_d/d, is derived from Eq. 3.7 and is given by:

$$\frac{\sigma_d}{d} = \frac{\ln(10)}{5}\sqrt{\sigma_{m_r}^2 + \sigma_{M_r}^2 + \sigma_{A_r}^2} \tag{3.9}$$

To find the total error, we must first find the uncertainties in M_r and A_r. The uncertainty $\sigma^2_{A_r}$, from Eq. 3.5, is:

$$\sigma^2_{A_r} = C_r^2 \sigma^2_{E(B-V)} + (E(B-V))^2 \sigma^2_{C_r} \tag{3.10}$$

The uncertainty $\sigma^2_{M_r}$, from the P-L relation in Eq. 3.2, propagates from the errors in the coefficients (a_r, b_r) and the period (P):

$$\sigma^2_{M_r} = \sigma^2_b + (\log_{10} P - 1)^2 \sigma^2_a + \left(\frac{a_r}{P \ln 10}\right)^2 \sigma^2_P \tag{3.11}$$

Plugging in the numerical values for these uncertainties: $\sigma_a = 0.106$ [2], $\sigma_b = 0.032$ [2], $\sigma_P = 2.90$ days(from Bootstrap), $\sigma_{E(B-V)} = 0.09$ [4] and $\sigma_{m_r} = 0.021$ (using error weights):

$$\sigma_d = 1.269 \text{ kpc} \tag{3.12}$$

$$\text{Error} \approx 17\% \tag{3.13}$$

Chapter 4

Results

We thus obtain the distance to the star cluster **Be #162** to be:

$$d = (7.305 \pm 1.269) \text{ kpc} \tag{4.1}$$

with a 17% error. Results in literature have been reported to be around [7]: $5.3^{+1.0}_{-0.8}$ kpc, which is very close to and is contained within the error bars of our result.

Bibliography

The following github repo contains the code file.

[1] http://github.com/Kuttus4444/Astro4444

[2] https://arxiv.org/abs/2306.06326

[3] https://astronomy.swin.edu.au/cosmos/i/interstellar+reddening

[4] https://arxiv.org/pdf/1805.05663

[5] https://irsa.ipac.caltech.edu/applications/DUST/

[6] https://lco.global/spacebook/distance/what-is-distance-modulus/

[7] https://academic.oup.com/mnras/article/478/3/3825/4996801

[8] https://ui.adsabs.harvard.edu/abs/2018MNRAS.477.2976N/abstract

[9] Wikipedia contributors. "Period-luminosity relation". *Wikipedia, The Free Encyclopedia*. https://en.wikipedia.org/wiki/Period-luminosity_relation (accessed November 13, 2025).

[10] Wikipedia contributors. "Cepheid variable". *Wikipedia, The Free Encyclopedia*. https://en.wikipedia.org/wiki/Cepheid_variable (accessed November 13, 2025).

[11] SIMBAD database. "ASASSN-V J201151.18+342447.2". http://simbad.u-strasbg.fr/simbad/sim-id?Ident=Be+51+%23162 (accessed November 13, 2025).

[12] Udalski, A. et al. (1999). "The Optical Gravitational Lensing Experiment. Cepheids in the Magellanic Clouds. IV. The P-L Relation in the BVRI Bands for Cepheids in IC 1613". *Acta Astronomica*, 49, 437.

[13] Leavitt, H. S. & Pickering, E. C. (1912). "Periods of 25 Variable Stars in the Small Magellanic Cloud". *Harvard College Observatory Circular*, 173, 1-3.

Sumit Kumar Adhya

astro_report (3)

Course project report

Document Details

Submission ID

trn:oid:::3618:121310427

Submission Date

Nov 13, 2025, 11:33 PM GMT+5:30

Download Date

Nov 13, 2025, 11:36 PM GMT+5:30

File Name

astro_report (3).pdf

File Size

524.2 KB

11 Pages

2,535 Words

12,894 Characters

4% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Match Groups

10 Not Cited or Quoted 4%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

2% Internet sources

1% Publications

2% Submitted works (Student Papers)

Match Groups

10 Not Cited or Quoted 4%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

2% Internet sources
1% Publications
2% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	
dspace.library.iitb.ac.in		<1%
2	Student papers	
Dublin City University on 2025-08-03		<1%
3	Internet	
www.arxiv-vanity.com		<1%
4	Student papers	
University of Glamorgan on 2014-02-03		<1%
5	Publication	
C. Aerts, J. Christensen-Dalsgaard, D. W. Kurtz. "Asteroseismology", Springer Natu...		<1%
6	Student papers	
University of Bath on 2018-09-27		<1%
7	Internet	
link.springer.com		<1%
8	Student papers	
Eagle Mt-Saginaw Independent School District on 2025-03-11		<1%
9	Student papers	
University of Hertfordshire on 2024-04-18		<1%
10	Publication	
S Nösel, R Sharma, M Massi, G Cimó, M Chernyakova. "Hour timescale QPOs in th...		<1%

Measuring distance to the open cluster Berkeley 51 using Be 51 #162 Cepheid data

Course Project, PH556: Astrophysics

by

Sumit Kumar Adhya

(Roll No. 23B1806)

Team: Hanle Hunters

Department of Physics

Indian Institute of Technology Bombay

Mumbai - 400076, India

Contents

1 Introduction and Objectives **2**
- 1.1 Target: Be #162 2
  - 1.1.1 Target Selection and Rationale 2
  - 1.1.2 Observations 3

2 Methods: PSF and Data Analysis **4**
- 2.1 PSF Photometry 4
- 2.2 Light Curve Fitting and Period Extraction 5
- 2.3 Error of Period (Using Bootstrap) 5

3 Calculations **7**
- 3.1 Distance Calculation 7
- 3.2 Error Analysis 7

4 Results **9**

Chapter 1

Introduction and Objectives

Determining the scale of the universe is a fundamental goal of astrophysics. This is built upon the "Cosmic Distance Ladder," [9] a succession of methods used to measure distances to increasingly remote objects. Near the base of this ladder, and crucial for calibrating all further steps, are the Classical Cepheid variable stars.

A Cepheid variable is a radial pulsator in the instability strip of the H-R diagram [9]. These stars rhythmically expand and contract, causing periodic changes in their brightness. This pulsation is driven by the κ-mechanism [9], which involves opacity changes in the helium ionization zones within the star [9]. Because their pulsation period is directly linked to their intrinsic luminosity, they serve as one of the most reliable "standard candles" available to astronomers.

The cornerstone of their use is the Period-Luminosity (P-L) relation, first discovered by Henrietta Swan Leavitt [12]. This tight correlation means that by simply measuring the pulsation period of a Cepheid, one can determine its absolute magnitude (intrinsic brightness). Classical (Type 1) Cepheids, like the one in this study, are young, metal-rich Population I stars. For our analysis, we adopt the P-L calibration for the r-band [1]:

$$M_r = (-2.637 \pm 0.106)(\log_{10} P - 1) + (-4.088 \pm 0.032) \tag{1.1}$$

Once the absolute magnitude (M_r) is known from the period, we can compare it to the star's measured apparent magnitude (m_r), corrected for interstellar extinction (A_r). This yields the distance modulus, $\mu_0 = m_r - M_r - A_r$, and thus the distance to the star.

In this work, we aim to apply this technique to measure the distance to the open cluster Berkeley 51. Our target is the classical Cepheid **Be 51#162**, a confirmed member of the cluster [10]. By obtaining new time-series photometry of this star, we will characterize its pulsation, derive its period, and apply the P-L relation to determine an independent distance to the cluster, which is known to host several other interesting massive stars [7].

1.1 Target: Be #162

1.1.1 Target Selection and Rationale

Two primary observing strategies were initially considered: (i) targeting Cepheids in the nearby M31 (Andromeda) galaxy, and (ii) targeting Cepheids within Galactic star clusters. A review of catalog data indicated that most Cepheids in M31 are fainter than our 20-mag limit and suffer from severe crowding and high background surface brightness. These factors would degrade photometric precision with a 1-m class instrument, rendering M31 unsuitable.

Consequently, the search was restricted to Galactic Cepheids, specifically within open clusters. Open clusters are less crowded than globular clusters, reducing the risk of blending and saturation, which is ideal for accurate point-spread-function (PSF) photometry. We queried major catalogs (SIMBAD, Pan-STARRS, VizieR) with the following criteria:

- Apparent magnitude in the 15-19 range.
- Pulsation period ≤ 10 days (to match our time allocation).
- Declination $\delta > 30°$ for adequate elevation at Hanle.
- Visibility during the 15-30 October run.
- Minimum lunar separation $> 30°$.

These filters identified the classical Cepheid **Be51#162** in the open cluster **Berkeley 51** as the highest-priority candidate. This target has a known pulsation period of ≈ 9.8 days and secure cluster membership [3]. Berkeley 51 is in a relatively sparse region of the Perseus arm, providing a well-characterized environment suitable for this study. The target meets all magnitude and visibility constraints.

1.1.2 Observations

We requested time-series imaging of Be 51#162 to measure its period and derive a distance. The observation parameters were as follows:

- **Coordinates:** $RA = 20:11:51.18$, $Dec = +34:24:47.2$ (J2000)
- **Date and Time:** '2025-10-13T15:43:19', '2025-10-15T19:02:31', '2025-10-16T14:53:04', '2025-10-19T14:21:15', '2025-10-20T13:59:25', '2025-10-23T14:38:15', '2025-10-24T15:32:45', '2025-10-26T13:00:50', '2025-10-27T17:32:41', '2025-10-28T16:39:00', '2025-10-29T16:54:14', '2025-10-30T14:37:14', '2025-10-31T15:20:41'
- **Filter:** SDSS r
- **Sequence:** 100s exposures per visit

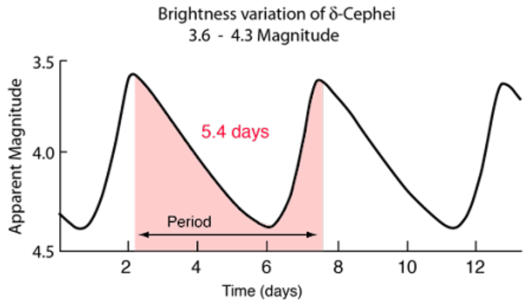

Figure 1.1: Brightness Variation of a Cepheid variable
Image credit: http://hyperphysics.phy-astr.gsu.edu/hbase/Astro/cepheid.html

Chapter 2

Methods: PSF and Data Analysis

2.1 PSF Photometry

To get the brightness measurement for our Cepheid, we first had to process the raw FITS images. The real challenge here is that our target is in a crowded open cluster, meaning other stars are very close to it and their light can overlap. The best way to handle this is with **Point Spread Function (PSF)** photometry. This technique isn't as simple as just drawing a circle; instead, it involves building a precise model of what a star "looks like" in our images, and then fitting that model to our target to measure its brightness.

Our photometric pipeline, which we built in a Python Jupyter Notebook using the `astropy` and `photutils` libraries, was a step-by-step process:

1. **Data Ingestion:** First, we had to loop through all the FITS image files. For each one, we'd open the file, read the 2D pixel data, and—most importantly—pull the time of observation (in Modified Julian Date, or MJD) from the FITS header. This MJD timestamp is what we'd eventually plot our brightness against.

2. **Background Estimation:** Images from a telescope are never truly black. There's always a background sky level. We estimated this for each image by calculating a sigma-clipped median of all the pixel values. Using a sigma-clipped median is key, as it ignores the very bright pixels from the stars themselves and gives a robust measure of just the background "glow."

3. **Building the PSF Model:** This is the core of the technique. We used a group of well-behaved relatively isolated stars to build an average, high-signal-to-noise model of the star shape (the PSF) for each image. This model accounts for all the distortions from the atmosphere and telescope optics at that specific moment.

4. **Identifying Target and References:** We then created a list of coordinates for our target star (Be 51 #162) and a separate list for our "comp" or reference stars. We chose a set of around 50 non-variable stars in the same field of view to act as our stable reference.

5. **Fitting and Measuring Flux:** With the PSF model and star list, we used `photutils` to perform the photometry. For every star in our list (target and references) on every single image, the code would fit our PSF model at that star's location. The output of this fit isn't just a count, but a precise calculation of the total flux (or brightness) coming from that star, separating its light from the background and any nearby faint stars.

6. **Differential Photometry:** This is the step that makes the data usable. The raw flux of a star can change wildly from one image to the next, simply because of thin clouds or changing atmospheric transparency. To fix this, we performed *differential photometry*. For

each image, we created a stable "ensemble" flux by summing up the flux of our reference stars. We then calculated our target's magnitude *relative* to this stable ensemble. This trick cancels out all the atmospheric effects, leaving us with only the true, intrinsic variation of the Cepheid.

7. **Light Curve Generation:** Finally, we took the resulting differential magnitude for our target from every image and plotted it against the MJD we extracted back in step 1. The resulting graph is the *light curve*, and it's the final product of our photometric analysis. This time-series plot, showing the clear "rise and fall" of the Cepheid's brightness, is the raw material we fed directly into the period-finding analysis.

2.2 Light Curve Fitting and Period Extraction

Once we had our final light curve (a set of times, magnitudes, and magnitude errors), the next crucial step was to find the star's pulsation period. A quick look at our data shows that the observation times are not evenly spaced; we have data from different nights with gaps in between. This means we can't use a standard Fast Fourier Transform (FFT).

The correct tool for this job is the **Lomb-Scargle Periodogram**, a powerful algorithm designed specifically for finding periodic signals in unevenly-sampled data. We used the `LombScargle` implementation from the `astropy.timeseries` library.

Our first step was to prepare the data. We converted the observation timestamps into a simpler NumPy array, `t`, representing the time in days since the first observation. We also loaded our magnitudes and their corresponding photometric errors into arrays called `mag` and `dmag`.

With the data ready, we initialized the `LombScargle` object. Crucially, we passed it all three arrays: `t`, `mag`, and `dmag`. By providing the `dmag` error array, we're telling the algorithm to perform a weighted analysis—it pays more attention to the data points we are more certain of (those with small error bars) and is more skeptical of noisier measurements.

We then used the `autopower` function to scan for potential periods. We limited our search to a sensible range of frequencies, from 0.05 to 1.0 cycles/day (corresponding to periods between 1 and 20 days), which is the expected range for a Cepheid of this type. This function returned an array of frequencies and their corresponding "power" (a measure of how well a sine wave at that frequency fits the data).

Finding the best period was then straightforward: we just found the frequency that produced the highest power peak using `np.argmax(power)`. The period, of course, is just the inverse of this best frequency ($P = 1/f$). To visually confirm our result, we used the `ls.model` function to plot the best-fit sine wave over our original data points, which showed an excellent match.

2.3 Error of Period (Using Bootstrap)

Just finding a single number for the period isn't enough for good science; we absolutely needed to know our uncertainty. The best way to get a robust, trustworthy error bar from a small, unevenly-sampled dataset like ours is to use a **bootstrap analysis**.

The idea behind bootstrapping is to simulate what our result might look like if we had collected our data on slightly different nights. We did this by re-running our entire period analysis 1,000 times.

In each of those 1,000 iterations, we created a new "fake" dataset by randomly sampling from our own original data points. The key is that we sampled with replacement, meaning a single bootstrap dataset might have (for example) two copies of our third data point and zero copies of our seventh. This `np.random.choice` function gave us a new set of arrays: `t_boot`, `mag_boot`, and `dmag_boot`.

We then ran the entire Lomb-Scargle periodogram analysis (initializing, finding the power, and getting the peak) on this new, slightly-different resampled dataset. This gave us a "best period" for that specific iteration, which we saved to a list.

After the loop finished, we had a list of 1,000 different "best periods." This list represents the distribution of results we could expect given our data's inherent noise and sampling. To get our final one-sigma error, we simply calculated the standard deviation (`np.std`) of this list. This standard deviation, σ_P, is our final, empirically-derived uncertainty for the period.

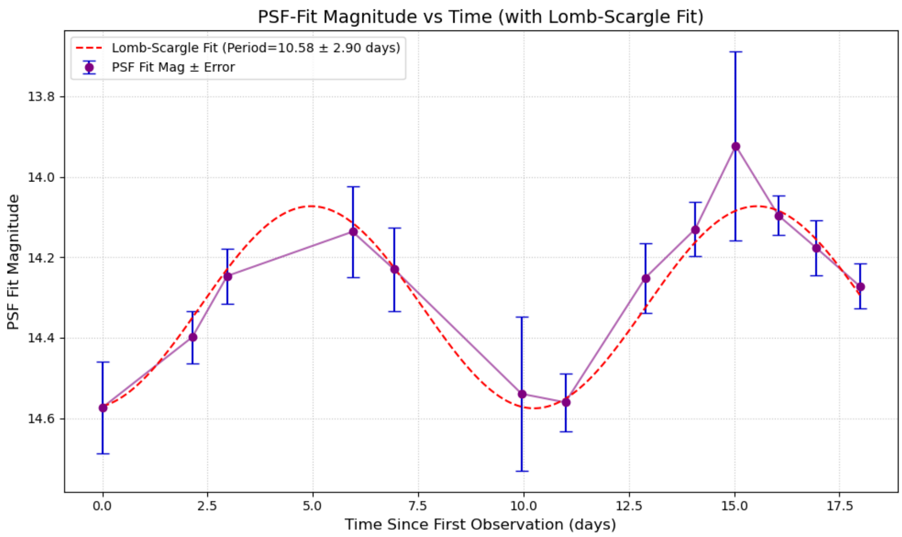

Figure 2.1: Lomb-Scargle Fit

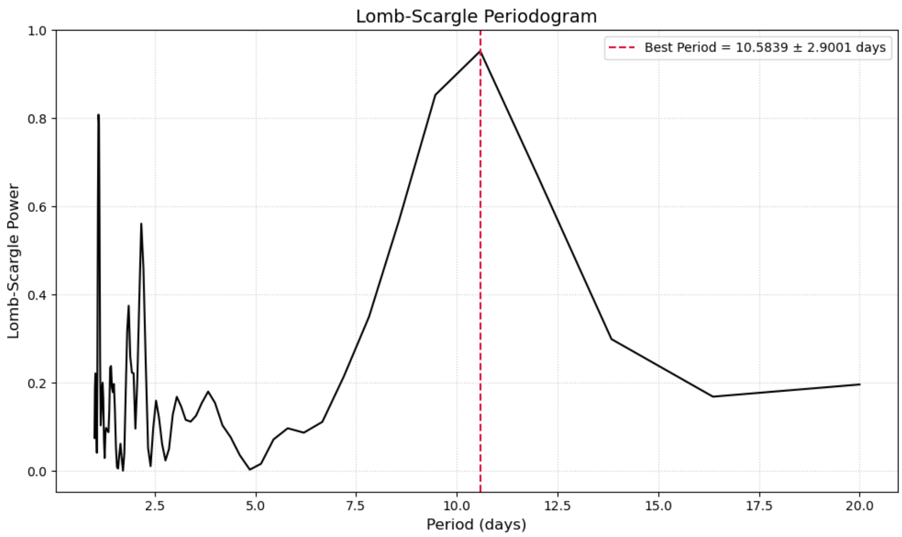

Figure 2.2: Lomb-Scargle Periodogram

Chapter 3

Calculations

3.1 Distance Calculation

From our photometric analysis, obtaining the light-curve followed by Lomb-Scargle fitting we obtain the period to be:

$$P = 10.58 \text{ days} \tag{3.1}$$

Plugging this into the P-L relation for r-band: [1]

$$M_r = a_r(\log_{10} P - 1) + b_r \tag{3.2}$$

where $a_r = -2.637$ and $b_r = -4.088$ [1], we obtain the absolute magnitude:

$$M_r = -4.153 \tag{3.3}$$

We take the error weighted mean of the observed apparent magnitudes to get:

$$m_r = 14.255 \tag{3.4}$$

The total extinction A_r is calculated from the reddening $E(B-V)$ using the relation [2]:

$$A_r = C_r \times E(B-V) \tag{3.5}$$

where $C_r = 2.285$ [4] and $E(B-V) = 1.79$ [3], which gives:

$$A_r = 4.09 \tag{3.6}$$

Now plugging all these values in the formula for distance [5]:

$$d = 10^{\frac{m_r - M_r - A_r + 5}{5}} \text{ pc} \tag{3.7}$$

we get:

$$d = 7.305 \text{ kpc} \tag{3.8}$$

3.2 Error Analysis

Using the expressions we have referred to so far and the standard error propagation formula, we obtain the following relations for the errors.

The fractional error in the distance, σ_d/d, is derived from Eq. 3.7 and is given by:

$$\frac{\sigma_d}{d} = \frac{\ln(10)}{5}\sqrt{\sigma_{m_r}^2 + \sigma_{M_r}^2 + \sigma_{A_r}^2} \tag{3.9}$$

To find the total error, we must first find the uncertainties in M_r and A_r. The uncertainty $\sigma^2_{A_r}$, from Eq. 3.5, is:

$$\sigma^2_{A_r} = C_r^2 \sigma^2_{E(B-V)} + (E(B-V))^2 \sigma^2_{C_r} \tag{3.10}$$

The uncertainty $\sigma^2_{M_r}$, from the P-L relation in Eq. 3.2, propagates from the errors in the coefficients (a_r, b_r) and the period (P):

$$\sigma^2_{M_r} = \sigma^2_b + (\log_{10} P - 1)^2 \sigma^2_a + \left(\frac{a_r}{P \ln 10}\right)^2 \sigma^2_P \tag{3.11}$$

Plugging in the numerical values for these uncertainties: $\sigma_a = 0.106$ [1], $\sigma_b = 0.032$ [1], $\sigma_P =$ 2.90 days(from Bootstrap), $\sigma_{E(B-V)} = 0.09$ [3] and $\sigma_{m_r} = 0.021$ (using error weights):

$$\sigma_d = 1.269 \text{ kpc} \tag{3.12}$$

$$\text{Error} \approx 17\% \tag{3.13}$$

Chapter 4

Results

We thus obtain the distance to the star cluster **Be #162** to be:

$$d = (7.305 \pm 1.269) \text{ kpc} \tag{4.1}$$

with a 17% error. Results in literature have been reported to be around [6]: $5.3^{+1.0}_{-0.8}$ kpc

Bibliography

[1] https://arxiv.org/abs/2306.06326

[2] https://astronomy.swin.edu.au/cosmos/i/interstellar+reddening

[3] https://arxiv.org/pdf/1805.05663

[4] https://irsa.ipac.caltech.edu/applications/DUST/

[5] https://lco.global/spacebook/distance/what-is-distance-modulus/

[6] https://academic.oup.com/mnras/article/478/3/3825/4996801

[7] https://ui.adsabs.harvard.edu/abs/2018MNRAS.477.2976N/abstract

[8] Wikipedia contributors. "Period-luminosity relation". *Wikipedia, The Free Encyclopedia*. https://en.wikipedia.org/wiki/Period-luminosity_relation (accessed November 13, 2025).

[9] Wikipedia contributors. "Cepheid variable". *Wikipedia, The Free Encyclopedia*. https://en.wikipedia.org/wiki/Cepheid_variable (accessed November 13, 2025).

[10] SIMBAD database. "ASASSN-V J201151.18+342447.2". http://simbad.u-strasbg.fr/simbad/sim-id?Ident=Be+51+%23162 (accessed November 13, 2025).

[11] Udalski, A. et al. (1999). "The Optical Gravitational Lensing Experiment. Cepheids in the Magellanic Clouds. IV. The P-L Relation in the BVRI Bands for Cepheids in IC 1613". *Acta Astronomica*, 49, 437.

[12] Leavitt, H. S. & Pickering, E. C. (1912). "Periods of 25 Variable Stars in the Small Magellanic Cloud". *Harvard College Observatory Circular*, 173, 1-3.