

Astronomy 128: Astro Data Science Lab

Lab 1: Gaia, RR Lyrae Stars, and Galactic Dust

HIEN NGUYEN^{1,2}

¹Department of Astronomy, University of California, Berkeley, CA 94720-3411, USA

²Department of Physics, University of California, Berkeley, CA 94720-7300, USA

ABSTRACT

Presented here is a 2D galactic dust map constructed from the G-band extinction of RR Lyrae stars from the Gaia Data Release 3 catalog. The extinction was computed using the intrinsic color modelled with Bayesian inference and the No-U-Turn Sampler (NUTS); all made possible due to the periodic variability of RR Lyraes.

Keywords: Periodograms; RR Lyrae; Standard candles; Leavitt Law

1. INTRODUCTION

Cosmic dust permeates the galaxy and absorbs or scatters short wavelength radiation from any cosmic sources. Long wavelength has a higher chance of passing through dust, so any light from a source that is blocked by dust will be ‘reddened’ from the perspective of the observer. As a result, dust extinction plays an important role in astronomical observations and analyses. In 1998, Schlegel, Finkbeiner, and Davis (SFD) produced a galactic dust map using extinction values in the visible band (V-band) from emission data of dust. The goal of this paper is to reproduce a galactic dust map similar to SFD, but instead using the variable stars dubbed RR Lyrae as the source for the extinction calculation. The resulting galactic dust map will not be exactly the same as the SFD map, and it will not be as accurate. Fundamentally, creating a dust map from dust emission data will always yield a better result than creating one by inferring extinction from non-dust sources.

Section 2 goes over all of the methodologies behind each steps, including the Astronomical Data Query Language (ADQL) Gaia queries, algorithms, calculations, and models. Section 3 covers the results and discussions about each step, with references and comparison to the literature. The appendix contains all error propagation formulas, and any additional data tables and plots that were deemed repetitive to the main body of the paper.

2. METHODOLOGY

RR Lyrae are variable stars that exhibit a period-luminosity relation, known as the Leavitt Law. The luminosity is observed as a magnitude value in various

photometric bands, with the G-band being the primary focus in this paper as the RR Lyrae data is obtained from the Gaia catalogs which records magnitude data in the G-band. The period is used to model the intrinsic color, and along with the observed color gives the extinction value for each RR Lyrae sources. Lomb-Scargle periodograms and fourier analysis are used to obtain estimates of RR Lyrae periods and magnitudes from raw light curves in Gaia’s DataLink service. The purpose of these estimates is to verify the reliability of the reported period and magnitude values in the Gaia catalogs. The modeling of the intrinsic color will be done using Bayesian inference and Markov Chain Monte Carlo (MCMC) sampling algorithms. The dust map itself will be created using the No-U-Turn Sampler (NUTS), an extension of the Hamiltonian Monte Carlo (HMC) sampler. To better understand the inner workings of these samplers, a Metropolis-Hastings (M-H) algorithm was constructed from scratch and used to model the period-luminosity relation for 500 RR Lyrae sources.

2.1. Querying the Gaia RR Lyrae Catalog

A processed dataset of 100 RR Lyrae stars was obtained from the Gaia Data Release 3 (GDR3) catalog using the Astronomical Data Query Language (ADQL). The first query filtered for sources with a measured fundamental pulsation frequency (pf) and more than 40 clean epochs in the G-band.

```
69  SELECT 100 *
70  FROM gaiadr3.vari_rrlyrae
71  WHERE pf IS NOT NULL
72  AND num_clean_epochs_g > 40
```

The queried dataset is the result of fitting the light curves of RR Lyrae stars, with the parameters and their derivations being documented by Clementini et al. (2016) and Clementini et al. (2023). Raw

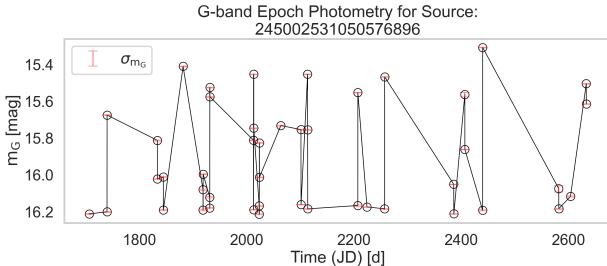


Figure 1. Epoch photometry in the G-band of source 245002531050576896. Plotted is the apparent magnitude versus time (Julian Date). The magnitude varies periodically over time. Note the inconsistent intervals over which observation was taken for the source. Magnitude uncertainties were calculated using flux error data. Since uncertainties values are small in comparison to magnitude values, horizontal caps were added for visibility.

RR Lyrae light curves were downloaded from the Gaia DataLink service using `Gaia.load_data` from the `astroquery.gaia.Gaia` package. Source IDs for each RR Lyrae were cross-matched to ensure that the raw light curves from DataLink correspond to the queried results from the `vari_rrlyrae` catalog. Apparent magnitude uncertainties values were calculated for the raw light curves by propagating errors (section B.1). A light curve for a single source is shown in Fig. 1.

2.2. Estimation of Period and Mean Magnitude

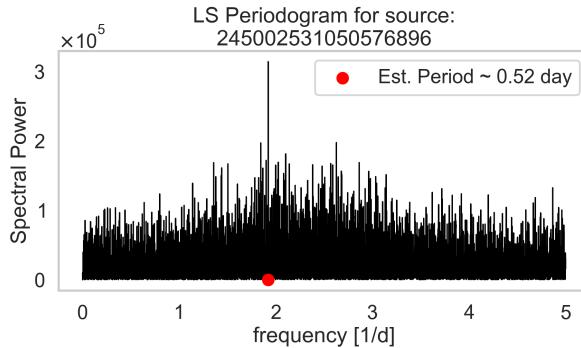


Figure 2. A LS periodogram for the same source from Fig. 1. The largest peak in spectral power corresponds to the fundamental pulsation frequency, which was inverted to estimate the period in the source's magnitude variability. Calculated magnitude uncertainties were included in the LS calculation, thus the spectral power is a dimensionless quantity.

According to Clementini et al. (2016), RR Lyrae stars are classified with a period (in days) of $0.2 \leq P \leq 1$, which translates to a frequency range (cycles per day) $1 \leq \nu \leq 5$. Applying this physical constraint to the

data, the fundamental pulsation period was estimated using a Lomb-Scargle (LS) periodogram with the package `astropy.timeseries.LombScargle`. The frequencies at the peak spectral power were extracted and inverted to estimate the period of each light curve. The mean magnitudes were estimated by calculating the means in flux space and converting to magnitude space. The relationship between the apparent magnitude and the flux is given by

$$m = -2.5 \log_{10} (F) , \quad (1)$$

so the mean magnitude can be calculated with

$$m_{\text{mean}} = -2.5 \log_{10} \left(\frac{\sum 10^{-m/2.5}}{N} \right) , \quad (2)$$

where N is the number of data points. The problem with estimating the mean magnitude this way is that the data points were not observed in a consistent time interval. Better estimates can be determined by performing the same calculations, but instead over a single pulsation period with Fourier series representation.

2.3. Fourier Analysis of Light Curves

2.3.1. Fourier Decomposition of Light Curves

The apparent magnitude of the RR Lyrae sources can be decomposed with the Fourier series

$$m_G(t) = A_0 + \sum_{k=1}^K a_k \sin(k\omega t) + b_k \cos(k\omega t) , \quad (3)$$

where A_0 , a_k , and b_k are constant coefficients, and $\omega = 2\pi/P$ is the angular frequency. With the period obtained using the estimation method described in section 2.2, the angular frequency ω can be treated as a known quantity. The Fourier representation can then be rewritten as a linear matrix equation $y = \mathbf{X}\beta$, where y is an array of measured magnitude, \mathbf{X} is a matrix constructed from known quantities as cosine and sine terms, and β is an array of the unknown coefficients. In full matrix notation, \mathbf{X} is a $(2K+1) \times K$ matrix

$$\begin{pmatrix} 1 & \sin(\omega t) & \cos(\omega t) & \cdots & \sin(k\omega t) & \cos(k\omega t) \\ \vdots & & & & \ddots & \vdots \\ 1 & \sin(\omega t) & \cos(\omega t) & \cdots & \sin(k\omega t) & \cos(k\omega t) \end{pmatrix} \quad (4)$$

and β is a $1 \times (2K+1)$ array

$$\begin{pmatrix} A_0 & a_1 & b_1 & \cdots & a_k & b_k \end{pmatrix}^\top \quad (5)$$

Using the estimated period computed from the Lomb-Scargle periodogram, a Fourier representation of the apparent magnitude can be constructed. The python implementation of Fourier decomposition used the least square package `np.linalg.lstsq` to solve for the coefficient array β . An example of Fourier decomposition was

performed with an RR Lyrae with GDR3 source_id = 5817567360327589632 for $K = 1, 3, 5, 7, 9$ (see Fig. 3 and 4).

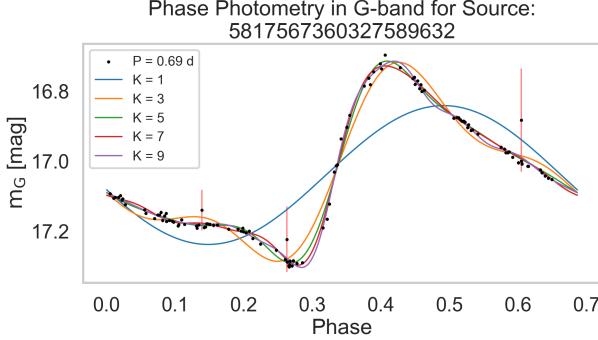


Figure 3. A phase photometry plot in the G-band for source 5817567360327589632, with Fourier representations of the apparent magnitude for $K = 1, 3, 5, 7$, and 9 overplotted. In phase space the sinusoidal variation in the magnitude is evident despite the inconsistent time intervals from data collection. Magnitude error bars were constructed as outlined in Sec. B.1.

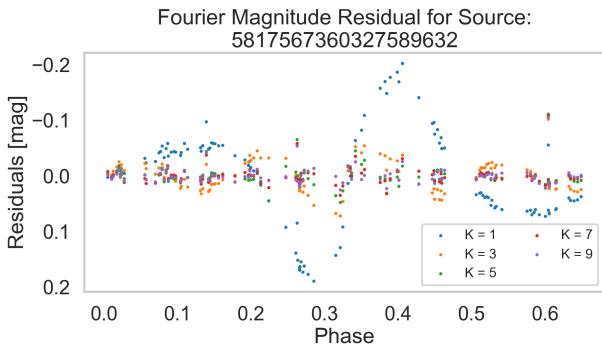


Figure 4. A phase photometry plot in the G-band of the residual between the observed magnitude and Fourier constructed magnitude of the same source from Fig. 3. For $K > 3$, the residuals converge on zero. This is in agreement with the cross-validation test which shows that the optimal range for K is around 5 to 9. In phase space, it can be observed that the residuals for small K are sinusoidal.

2.3.2. Cross-Validation: Determination of Optimal K

To determine the optimal amount of terms to use in a Fourier series representation, a cross-validation test can be performed by splitting the dataset into a training set and a testing set, then comparing their computed χ^2 values. 20% of the observed

data (time, magnitude, and magnitude uncertainty) were set aside as the cross-validation or test set and the other 80% were designated as the training set. The splitting of the data was implemented using the `sklearn.model_selection.train_test_split` package.

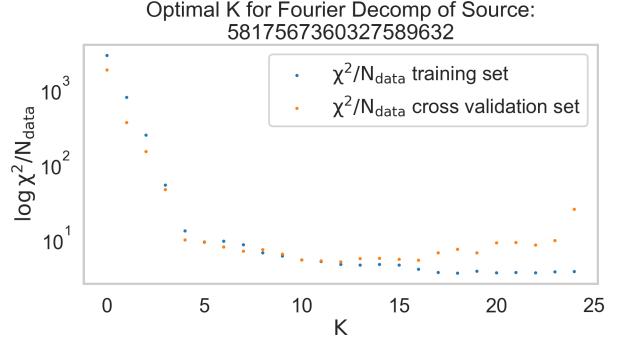


Figure 5. A plot of the χ^2 (normalized by number of data points) for the Fourier magnitudes between training and cross-validation data sets. The optimal range for K is $5 < K < 8$, with the optimal K being 7.

The Fourier model, $y = \mathbf{X}\beta$, was implemented with the training magnitude set to produce a model for the apparent magnitude, ie the training coefficients. The apparent magnitude was then computed using the training time set and the training coefficients to produce training set Fourier magnitudes. The χ^2 value was then computed for the training set with the equation:

$$\frac{\chi^2}{N_{\text{data}}} = \sum \left(\frac{m_{\text{data}} - m_{\text{fourier}}}{\sigma_m} \right)^2 \frac{1}{N_{\text{data}}}, \quad (6)$$

where the number of data N_{data} and the magnitude uncertainty σ_m are also from the training set. The process is then repeated with the testing set, with the exception of the coefficients which were computed from the training set. The optimal range of K occur for K values where the χ^2 for the cross-validation set is less than the χ^2 of the training set, and the optimal value for K corresponds to the minimum χ^2 value of the testing set within that range. From Fig. 5, the optimal K for that particular source appears to be 7.

2.4. Estimating Absolute Magnitude Using Bailer-Jones Distances

The methods of estimating period and mean magnitude and the fourier analysis of light curves, as described previously in sections 2.2 and 2.3, and the results in sections 3.1 and 3.3, should demonstrate the reliability of the period and apparent magnitude values in the Gaia catalog. The mean apparent magnitude of a source can be used to determine its absolute magnitude

$$M = m - 5 \log_{10}(d) + 5 \quad (7)$$

if the distance (in parsecs) to the source is known. The Gaia catalog contains parallax data, which is related to distance by

$$d \approx \frac{1}{\varpi} \quad (8)$$

where ϖ is the parallax in arcseconds (as). However, recall that equation 8 uses the small angle approximation. In actuality, the relationship between parallax and distance is non-linear. With smaller parallax measurements (as is the case with far away sources), the uncertainty in the calculated distance becomes significantly large. Therefore, a simple inversion of Gaia parallaxes would not be an accurate nor precise method to determine the distance to these RR Lyrae sources. Bailer-Jones et al. (2021) used Bayesian inference to determine two types of distances, one based on parallax and a direction-dependent prior on distance, and a second that includes color and apparent magnitude data. For the purposes of this paper, the geometric distances will be referenced. The galactic structure model that was used to construct the geometric distance prior assumes the spatial density of stars exponentially decreases with respect to the Sun according to some direction-dependent length scale. The likelihood function is a Gaussian that is dependent on a parallax zero point. The posterior probability distribution was then computed numerically with the help of Markov Chain Monte Carlo samplers. From the posteriors, geometric distances can then be inferred.

2.4.1. Querying for Off-Disk RR Lyraes

It is instructive to see the difference between the distance estimate from naive inversion of Gaia parallaxes to that of Bailer-Jones distance estimates. To this end, the following query for off-disk RR Lyrae sources was used:

```
WHERE parallax_over_error >= 10
AND ABS(b) >= 30
AND parallax >= 0.25 .
```

The first line ensures that parallax errors were $\leq 20\%$. The second line filters for stars that are above or below the galactic disk using galactic longitude $|b| \geq 30$ degrees. The last line selects for sources that are nearby at 4 kilo-parsecs (kpc) to minimize extinction effects. The naive inversion of distance to convert to parallax was used to implement the last condition

$$d \leq \frac{1}{\varpi} \rightarrow \varpi \geq \frac{1}{d} = \frac{10^3}{4 \cdot 10^3} \text{ mas} \rightarrow \varpi \geq 0.25 \text{ mas} \quad (9)$$

where mas stands for milli-arcseconds. Since the vari_rrlyrae does not contain parallax data, a cross-match of source_id was also implemented to join with the gaia_source catalog.

The Bailer-Jones distances can be obtained from the external gaiaedr3_distance catalog with the same

cross matching of source_id. The columns of interest in the Bailer-Jones data are r_med_geo, r_hi_geo, and r_lo_geo. From the discussion of Bayesian inference in section A, r_med_geo is the median of the geometric distance posterior, r_hi_geo is the upper 84th percentile or the upper standard deviation and r_lo_geo is the lower 16th percentile or the lower standard deviation of the geometric distance posterior.

2.4.2. Computing Distance from GDR3 Parallaxes and Parallaxes from Bailer-Jones Distances

The distance was estimated from GDR3 parallaxes with equation 8. The distance uncertainty can then be calculated with error propagation, see section B.2 and equation B6. The uncertainty in the Bailer-Jones distances was determined using symmetrized errors $(r_{hi_geo} - r_{low_geo})/2$, since these two values are symmetric about r_med_geo.

A Bailer-Jones parallax can be derived from the Bailer-Jones distances by inverting equation 8. The uncertainty in Bailer-Jones parallax can be calculated in a similar manner to the uncertainty in the GDR3 distances discussed previously (see equation B7).

2.4.3. Calculating Absolute Magnitudes

The absolute magnitude for these off-disk RR Lyrae sources were calculated using equation 7 alongside distances from either GDR3 parallaxes or the Bailer-Jones estimates. The uncertainty in the absolute magnitude is once again computed using error propagation (equation B8).

2.4.4. Quality Cuts for Off-Disk RR Lyraes

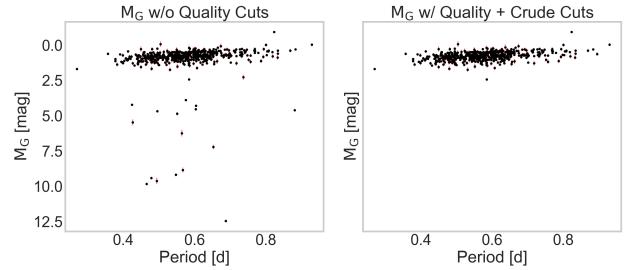


Figure 6. Absolute magnitude vs period plots highlighting the effectiveness of the applied quality cuts. The left plot does not have any quality cuts applied, but there is an absolute-magnitude-period relation present in the data. The outlier data points are due to large parallax measurement errors. The right plot includes quality cuts from Lindegren et al. (2018) and a crude cut with absolute magnitude value less than 3.

From figure 6, note that there is a clear period-absolute-magnitude relation. Some of the estimate absolute magnitudes deviate from the median, as shown

in the left plot of the figure. The outlier points are due to large parallax measurement errors. To remove these bad sources, quality cuts from Lindegren et al. (2018) were applied to the query in section 2.4.1:

```

255 WHERE phot_bp_mean_flux_over_error > 10
256 AND phot_rp_mean_flux_over_error > 10
257 AND astrometric_excess_noise < 1
258 AND phot_bp_rp_excess_factor
259     < 1.3 + 0.06*POWER(BP_RP, 2)
260 AND phot_bp_rp_excess_factor
261     > 1.0 + 0.015*POWER(BP_RP, 2)

```

These quality cuts filtered out most of the outliers, with the few remaining being filtered out by a crude absolute magnitude cut of `int_average_g - 5*LOG10(r_med_geo) + 5 < 3.0`.

2.5. Markov Chain Monte Carlo Sampler Model Fitting

Three Markov Chain Monte Carlo (MCMC) fitting methods were used to obtain the intrinsic absolute-magnitude-period relation: one using a homebrewed Metropolis-Hastings (M-H) MCMC sampler, and two using the no-U-turn (NUT) Hamiltonian Monte Carlo sampler (Hoffman & Gelman (2011)).

The model used to fit the absolute-magnitude-period relation is assumed to be linear in log space

$$M_G = a \log_{10} \left(\frac{P}{\text{day}} \right) + b \quad (10)$$

where a and b are free parameters that are produced in the posterior probability distribution (see section A). The fitting methods allow for intrinsic scatter in the absolute-magnitude-period relation. This means that without measurement uncertainties, the absolute magnitude M_G at period P follows a Gaussian distribution with variance $\sigma_{\text{scatter}}^2$.

2.5.1. Metropolis-Hastings (M-H) MCMC Sampler

A custom M-H MCMC sampler was programmed using a Gaussian proposal distribution. Diagnostic plots to verify that the M-H MCMC sampler works can be viewed in section C. To fit the absolute magnitude and period data, the likelihood probability distribution $\mathcal{L} \equiv p(\vec{d} | a, b, \sigma_{\text{scatter}})$, where \vec{d} is the multidimensional data array, is defined to be

$$\mathcal{L} = \prod_i \frac{1}{\sqrt{2\pi\sigma_{\text{total}}^2}} \exp \left[\frac{-(M_{Gi} - M_G)^2}{2\sigma_{\text{total}}^2} \right] \quad (11)$$

where $\sigma_{\text{total}} \equiv \sqrt{\sigma_i^2 + \sigma_{\text{scatter}}^2}$, σ_i is the absolute magnitude uncertainty (section B.4), M_{Gi} is the absolute magnitude calculated from Bailer-Jones geometric distances in section 3.4, and $M_G = M_G(P, a, b)$ is the linear model assumed in equation 10. What is defined above is

actually the total likelihood, hence the product of probability distributions. To simplify the computation, the likelihood can be converted into log space, which turns the product into a sum

$$\mathcal{L} = \sum_i \log \left(\frac{1}{\sqrt{2\pi\sigma_{\text{total}}^2}} \right) + \frac{-(M_{Gi} - M_G)^2}{2\sigma_{\text{total}}^2} \quad (12)$$

The M-H MCMC algorithm was implemented as follows:

- Create an initial guess for the fitting parameters and store as the ‘current’ parameters.
- Propose new parameter values by drawing from a normal Gaussian distribution with a random walk.
- Compute the posteriors for the proposed and current values using the likelihood function defined in equation 12
- Define the ‘acceptance fraction’ to be the ratio $\mathcal{L}_{\text{proposed}} / \mathcal{L}_{\text{current}}$
- If $\mathcal{L}_{\text{proposed}} > \mathcal{L}_{\text{current}}$, then the posterior parameters are accepted. Otherwise, the chances of accepting the new parameters is determined by the acceptance fraction.

The results of M-H MCMC sampler fitting is discussed in section 3.5.

2.5.2. Hamiltonian Monte Carlo (HMC) and the No-U-Turn Sampler (NUTS)

The Hamiltonian Monte Carlo (HMC) is an alternative MCMC algorithm that suppresses the inefficient random walk behaviors of MCMC algorithms such as Metropolis-Hastings in section 2.5.1. Instead, HMC simulates Hamiltonian dynamics and uses an auxiliary ‘momentum’ variable for each parameter to speed up convergence. The No-U-Turn Sampler (NUTS), introduced by Hoffman & Gelman (2011), goes one step farther and eliminates the need to specify the number of sampling steps.

Fitting with NUTS was done by utilizing the sampler provided by the PYMC3 package. Two types of fitting were done. The first has the likelihood function explicitly implemented as a ‘potential’ term in the model using the function `pymc3.Potential`. In the second method, the model was told to expect values of M_G at a given P to have a mean value μ defined by the linear relation in equation 10, and falls under a normal Gaussian distribution with mean μ and variance $\sigma_i^2 + \sigma_{\text{scatter}}^2$. This was done with the function `pymc3.Normal`. Refer to section 3.5 for the results and discussions.

2.5.3. Cross-Matching with WISE Survey

The RR Lyrae stars featured so far were also observed by the Wide-Field Infrared Survey Explorer (WISE),

342 with the difference being that the observations were
 343 made in the near-infrared. It is of interest to cross-match
 344 sources with WISE and derive an absolute-magnitude-
 345 period relation in the near-infrared.

346 The WISE dataset is in an external catalog
 347 `allwise_original_valid` from Gaia Data Release 1.
 348 In order to query these data, it is necessary to first
 349 join the `vari_rrlyrae` and `gaia_source` catalogs with
 350 the `allwise_best_neighbour` catalog from GDR3 us-
 351 ing the `source_id` of the RR Lyras. Then the ex-
 352 ternal GDR1 `allwise_original_valid` catalog can be
 353 joined with the rest using the external catalog identifier
 354 `allwise_oid` from the `allwise_best_neighbour` cata-
 355 log. The columns of interests are `allwise_oid`, `w2mpro`,
 356 and `w2mpro_error`. The query to join the catalogs is

```
357     JOIN gaiadr3.allwise_best_neighbour
  358         USING (source_id)
  359     JOIN gaiadr1.allwise_original_valid AS aw
  360         ON allwise_best_neighbour.allwise_oid
  361             = aw.allwise_oid
```

362 The third NUTS method was used to fit the assumed
 363 model (equation 10) to the near-infrared dataset (section
 364 2.5.2), as described in section 2.5.2. Discussions on the
 365 difference between the absolute-magnitude-period rela-
 366 tion in the optical and near-infrared is in section 3.5.1.

367 2.5.4. Computing Period-Color Relation in the G-band

368 Following the methods described in the previous sec-
 369 tion, the linear model of equation 10 is modified to derive
 370 a period-color relation. The Gaia color is $G_{\text{BP}} - G_{\text{RP}}$,
 371 so the model is

$$G_{\text{BP}} - G_{\text{RP}} = a \log_{10} \left(\frac{P}{\text{day}} \right) + b \quad (13)$$

372 where a and b are once again free parameters. The color
 373 magnitude in the Gaia catalog is derived from flux in
 374 the BP and RP bands, so the uncertainty in the colors
 375 is just the sum of magnitude uncertainties in section B.1,
 376 for the BP and RP bands.

$$\sigma_{\text{BP}-\text{RP}} = \sqrt{\left(\frac{2.5}{\ln 10} \frac{\sigma_{F_{\text{BP}}}}{F_{\text{BP}}} \right)^2 + \left(\frac{2.5}{\ln 10} \frac{\sigma_{F_{\text{RP}}}}{F_{\text{RP}}} \right)^2} \quad (14)$$

377 This formula for the color uncertainty is directly im-
 378 plemented into a new Gaia query that requests for the
 379 `bp_rp` column in the `gaia_source` catalog from GDR3.
 380 The period-color relation parameters are computed us-
 381 ing PYMC3 and NUTS, as before.

382 From figures 7 and 8, the derived period-color relation
 383 is

$$G_{\text{BP}} - G_{\text{RP}} = 0.29 \log_{10} \left(\frac{P}{\text{day}} \right) + 0.72 \quad (15)$$

384 2.6. Constructing the Galactic Dust Map

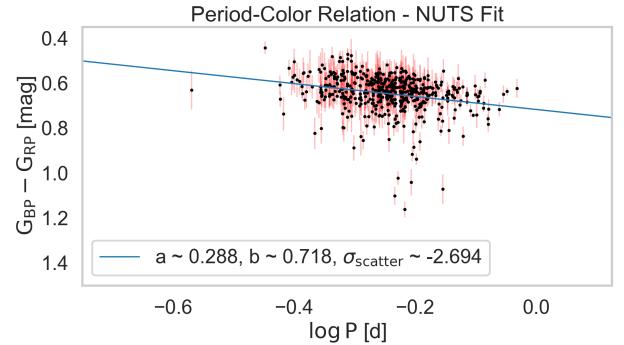


Figure 7. A period-color relation fit to the GDR3 data using NUTS.

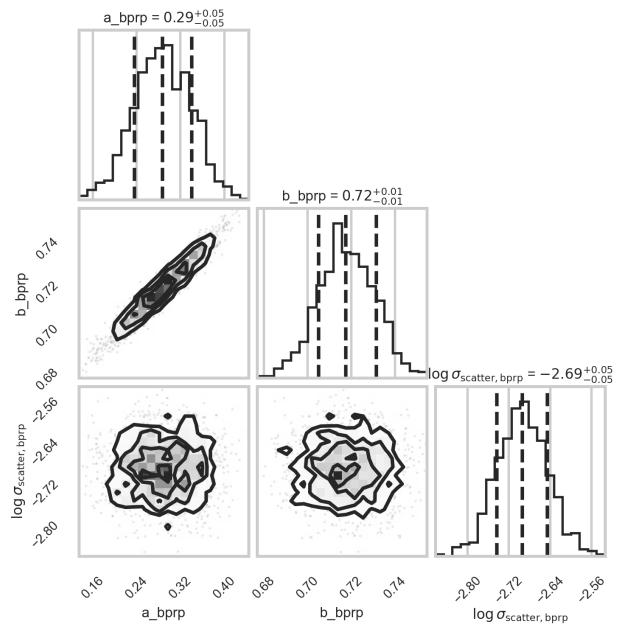


Figure 8. Corner plots of the posterior probability distributions of fitting parameters for the period-color relation of the GDR3 dataset. The quantiles shown are the lower 16th quantile and upper 84th quantile.

385 Having derived a period-color relation, it is now possi-
 386 ble construct a galactic dust map based on the G-band
 387 extinctions of RR Lyrae stars. With the period-color
 388 model in equation 15 and the observed color BP_RP in the
 389 `gaia_source` catalog, the color excess can be determined
 390 with equation $E(G_{\text{BP}} - G_{\text{RP}}) = (G_{\text{BP}} - G_{\text{RP}})_{\text{observed}} -$
 391 $(G_{\text{BP}} - G_{\text{RP}})_{\text{intrinsic}}$ where $(G_{\text{BP}} - G_{\text{RP}})_{\text{intrinsic}}$ is the
 392 derived period-color relation. The G-band extinction
 393 can then be found using

$$A_G = \frac{2.0}{E(G_{\text{BP}} - G_{\text{RP}})} \quad (16)$$

394 The quality cuts in section 2.4.4 need to be modified
 395 now that it is of interest to query for the entire RR
 396 Lyrae catalog to construct the galactic dust map. It is
 397 no longer necessary to exclude RR Lyrae stars with im-
 398 precise parallaxes, with low galactic latitude, or within
 399 4 kpc. So the following query lines can be removed

```
400 WHERE parallax_over_error >= 10
401 AND ABS(b) >= 30
402 AND parallax >= 0.25
```

403 The G-band extinction A_G was calculated using the fun-
 404 damental pulsation period pf from the `vari_rrlyrae`
 405 catalog, the color bp_rp from the `gaia_source` catalog,
 406 and the derived intrinsic period-color relation model in
 407 equation 15. The extinction A_G was then used to cre-
 408 ate a color map for a 2D Aitoff projection map of the
 409 RR Lyrae sources. To convert the galactic coordinates
 410 of the RR Lyraes to the Aitoff projection, the galactic
 411 longitude was sign inverted and both the galactic longi-
 412 tude and latitude were multiplied by $2\pi/180$. The initial
 413 result is shown in figure 23.

414 The photometric signal-to-noise and BP/RP excess
 415 quality cuts from Lindegren et al. (2018) can be relaxed
 416 in order to not exclude too many sources for the color
 417 excess calculation. These quality cuts are modified as
 418 follows

```
419 WHERE astrometric_excess_noise < 5
420 AND phot_bp_rp_excess_factor
421 < 3 + 0.06*POWER(bp_rp, 2)
422 AND phot_bp_rp_excess_factor
423 > 0.3 + 0.015*POWER(bp_rp, 2)
424 AND int_average_g - 5*LOG10(r_med_geo) + 5 < 6
```

425 The result of these modifications is shown in figure

3. RESULTS & DISCUSSIONS

3.1. Estimated Mean Magnitude and Period of RR Lyrae

426 The mean magnitude for 100 RR Lyrae light curves
 427 was estimated with the use of Fourier analysis as de-
 428 scribed in sections 2.2 and 2.3. A figure containing the
 429 data table is shown in Fig. 9. Fig. 10 is a compari-
 430 son plot of the mean magnitudes for 100 RR Lyrae light
 431 curves between GDR3, an estimate using flux in the raw
 432 light curves from the DataLink service, and the Fourier
 433 series representation with 7 terms. The optimal num-
 434 ber of Fourier terms was chosen to be 7 based on the
 435 cross-validation test performed in Sec. 2.3.2 for source
 436 581767360327589632. An alternative method would be
 437 to determine the optimal amount of Fourier terms for
 438 each source. Observe in Fig. 10 that the Fourier de-
 439 composition estimate line up well with the values cal-
 440 culated in GDR3. The naive estimates done with the
 441 DataLink raw light curves, that is calculating the mean
 442 in flux space and then converting to magnitude space,
 443 in general deviates from the mean magnitude values of
 444 both GDR3 and the Fourier series representation.

	source_id	pf_gdr3	period_est	mag_gdr3	mean_mag_est
0	245002531050576896	0.520512	0.520487	15.929145	15.870378
1	245504251951140864	0.507115	0.507130	18.000160	17.969103
2	245823861938360064	0.478764	0.478795	19.339872	19.323328
3	246756973652292992	0.489303	0.489275	12.362016	12.312440
4	359112665277840512	0.501780	0.501785	15.893806	15.913452
...
95	6061741130185585280	0.495418	0.459509	18.813147	17.996277
96	6061807062215683072	0.527883	0.642805	18.581533	19.610502
97	6061866508861866240	0.605841	8.132802	19.423979	19.863476
98	6061867986376466688	0.575690	0.766113	16.696190	16.310418
99	6061905026177458176	0.528914	0.627512	19.488836	18.525794

Figure 9. A data table comparing the mean magnitude and fundamental pulsation period calculated (via Lomb-Scargle) from 100 raw light curves, to the values in the `vari_rrlyrae` catalog, with the latter being denoted by ‘gdr3’.

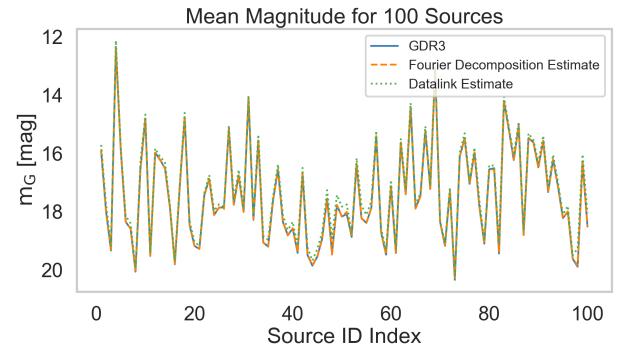


Figure 10. A plot comparing the three sets of mean magnitudes for 100 RR Lyrae light curves: Gaia Data Release 3, DataLink estimate through averaging in flux space, and Fourier decomposition with 7 terms.

448 The residuals between the GDR3 magnitude and
 449 DataLink estimated mean magnitude, and between the
 450 GDR3 magnitude and Fourier decomposition mean mag-
 451 nitude, are shown in Fig. 11. The difference between
 452 the Fourier series representation and the GDR3 mag-
 453 nitude is minimal, as can be seen by the residual plot
 454 converging on zero. The estimate done with the raw
 455 light curves from DataLink, taking the average in flux
 456 space and converting to magnitude space, again show a
 457 larger deviation from GDR3 data.

458 The difference between the estimated period and the
 459 period in the fundamental pulsation mode, as pro-
 460 vided by the Gaia catalog, is plotted in Fig. 12.
 461 The mean residual between the estimated period and
 462 the values obtained from the `vari_rrlyrae` are $0.2 \pm$
 463 0.1 d (with outliers) and 0.02 ± 0.01 d (without out-
 464 liers). The outliers are sources 4318106453828014336,

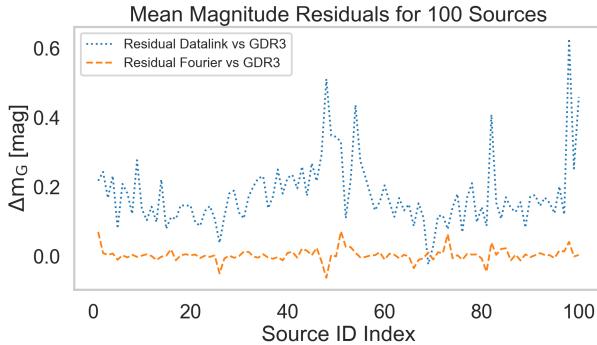


Figure 11. Residual plot of apparent magnitude in the G-band for 100 RR Lyrae sources from the GDR3 catalog. The blue curve is the residual between the DataLink mean magnitude estimate and the average G-band magnitude reported in GDR3. The orange curve compares the Fourier series representation mean magnitude to the same GDR3 G-band magnitude.

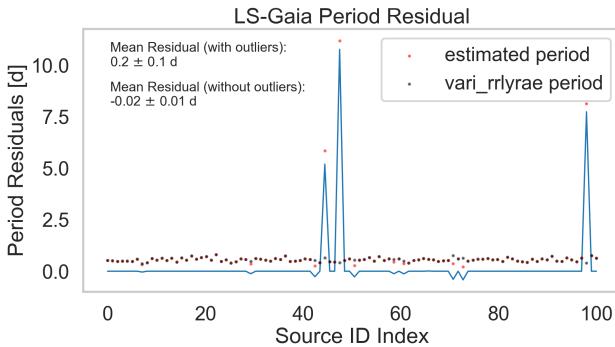


Figure 12. A residual plot between the estimated period (Lomb-Scargle) and the fundamental pulsation period obtained from Gaia’s `vari_rrlyrae` catalog. The red data points are the estimated period from DataLink and the black data points are values from the catalog. There are 3 outliers in the data, as shown by the spikes in the residual plots. Mean residual values were calculated with and without the outliers: 0.2 ± 0.1 d and -0.02 ± 0.01 d, respectively.

6061866508861866240, and 4317965200961192448. The residual plot and the data table in Fig. 9 show that the estimated values for the period are in agreement with the values from the Gaia catalog. Clementini et al. (2016) also determined the peak of the period distribution for RRab class RR Lyrae to be $P \sim 0.59$ d. Residual deviations from zero can be due to a large signal-to-noise ratio in the data, as can be seen in the Lomb-Scargle periodogram (Fig. 2).

3.2. Magnitude Extrapolation

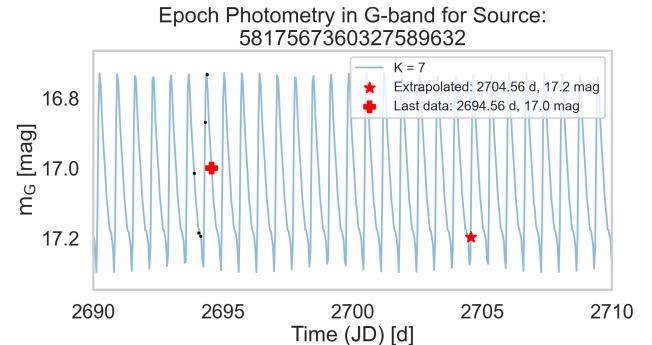


Figure 13. Extrapolated epoch photometry in the G-band for the source in Fig. 3 and 5.

Fourier decomposition of magnitude data enabled the analysis of a functional form $m_G(t)$. Since RR Lyrae magnitude fluctuations are periodic, the Fourier representation can be used to extrapolate a magnitude value to some future time t . Fourier reconstructed magnitudes are only an approximation, however. For the source 5817567360327589632 (the same source as in Fig. 3 and 5), a magnitude of 17.2 mag at 2704.56 days was extrapolated 10 days from the last recorded data point of 17.0 mag at 2694.56 days. The epoch photometry for the source is plotted in Fig. 13 along with the Fourier series representation and a few data points leading up to the final measured point.

3.3. Light Curves of Fundamental-Mode and First-Overtone Pulsators

RR Lyraes are categorized in two classes: RRab type and RRc type stars. RRab stars are known as the fundamental-mode pulsators, and RRc stars are called the first-overtone pulsators. RRab stars tend to have a period in the range of 0.3 to 1 day, while RRc have shorter periods at around 0.2 to 0.5 day. The amplitude in magnitude of RRc stars are also less than that of RRab stars (Soszyński et al. (2011) & Soszyński et al. (2014)).

A comparison between the light curves of RRab and RRc class RR Lyrae stars are shown with phase photometry in the G-band in figures 15 and 14. RRab light curves exhibit the steep rising branch leading up to the amplitude followed by a slow decline in magnitude. By plotting over 2 periods, the asymmetric RRab light curves can be contrasted to the more symmetric RRc light curves. The RRc light curves also nearly sinusoidal in appearance. RRab stars have a longer range for the period of their variability than RRc stars. In the 6 light curves that were sampled, this also appear to be the case. The estimated periods for each type of RR Lyrae are in agreement with Soszyński et al. (2011) & Soszyński et al. (2014), where RRab tend to have a period within the range 0.2 to 1 days and RRc have a

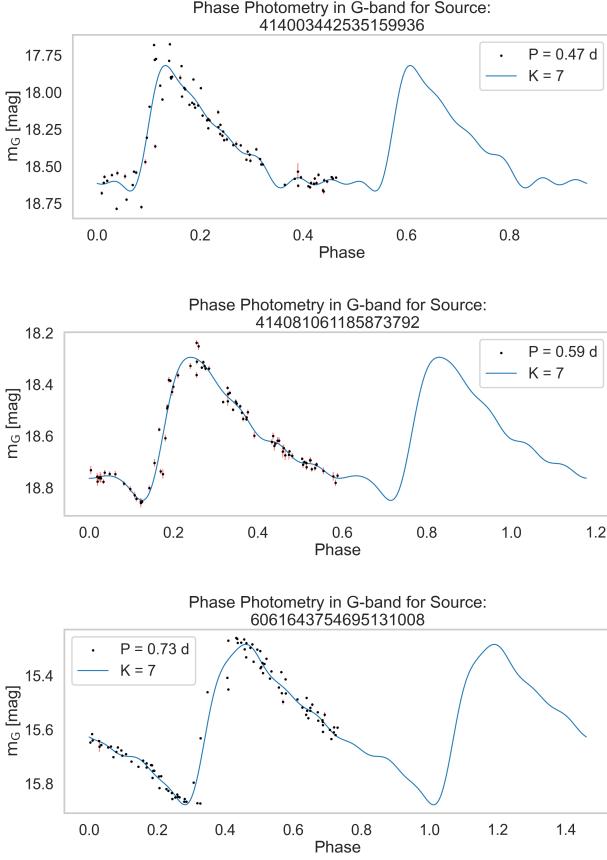


Figure 14. Phase photometry in the G-band for three fundamental-mode pulsators (RRab stars), plotted over 2 periods. The fundamental pulsation period and the number of fourier terms are listed in the legend of each plot. Observe the steep rising branch to the peak magnitude and the slow decrease that follows in all three light curves. Additionally, all 3 light curves show the presence of intrinsic scatter, which might have been caused by the modulation of amplitude/phase due to the Blazhko effect (Netzel et al. (2018)).

smaller range of 0.2 to a little over 0.5 days. RRab stars are also characterized as having larger amplitudes than RRc stars, which is also shown in the 6 light curves. Five of the 6 light curves, with the possible exception of the 3rd light curve for the RRc source, indicate the presence of intrinsic scatter, where the observed data is not well described by the fourier decomposition. The deviation between the data points and the series representation in these two light curves can attributed to the modulation of amplitude/phase by the Blazhko effect (Netzel et al. (2018)), a physical phenomenon that is still not well understood.

The optimal number of terms to use in Fourier decomposition of the magnitude for both classes of RR Lyraes was determined with the cross validation test as described in section 2.3.2. The cross validation test

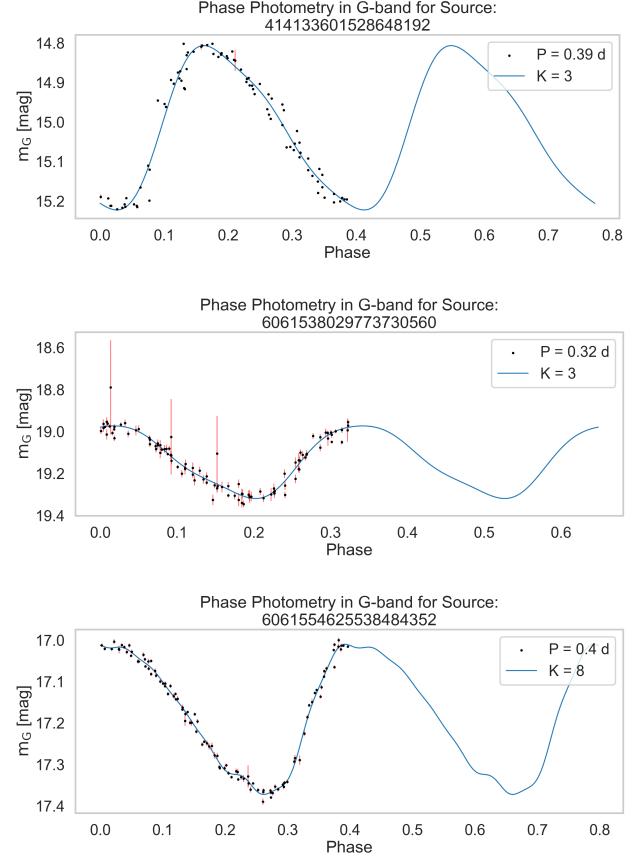


Figure 15. Phase photometry in the G-band for three First-overtone pulsators (RRc stars), plotted over 2 periods. The fundamental pulsation period and the number of fourier terms are shown in the legend of each plot. In comparison to RRab stars, the RRc light curves are more symmetric and are nearly sinusoidal. The first and second light curves also exhibit intrinsic scatter, albeit less than that of the RRab equivalent in figure 14, which again can be attributed to modulation by the Blazhko effect.

found the optimal K for these RRc stars to be 3, 3, and 8. The optimal number of fourier terms for the three RRab sources turned out to be overfitting. So the optimal number of terms for Fourier decomposition was chosen to be 7 based on a prior cross validation test for another RRab source (see figure 5).

3.4. Absolute Magnitude Estimate: Gaia DR3 vs. Bailer-Jones

The plots in figure 16 compares the distance estimates computed by inverting the GDR3 parallaxes to the Bailer-Jones distances. The non-linear relationship between parallax and distance can be observed in that as the parallax measurements become small, the discrepancy between the distance estimated from GDR3 parallaxes and the Bailer-Jones distances become significant. The parallax plot and its residual plot show

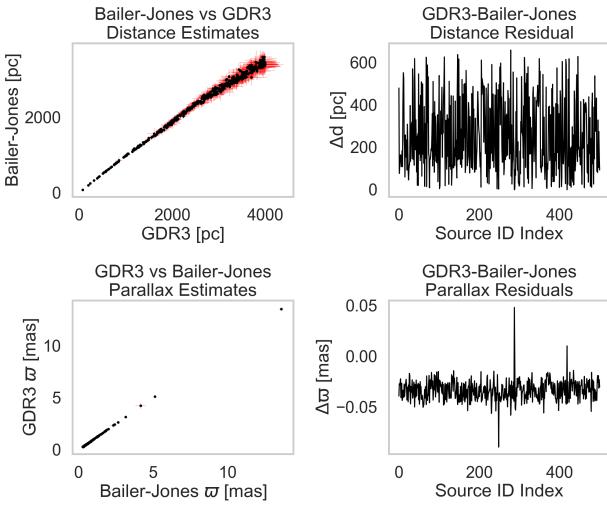


Figure 16. Plots of comparison between distance estimates computed from Gaia parallaxes and distances estimates from Bayesian inference as detailed in Bailer-Jones et al. (2021). The top left corner plot compares distances between GDR3 estimates and BJ data, with the top right plot showing the residuals in the distances. The bottom left plot compares the parallaxes between GDR3 data and BJ estimates, with the bottom right plot detailing the parallax residuals.

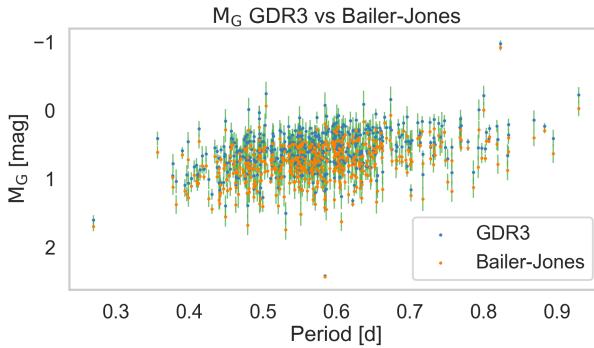


Figure 17. A plot comparing the absolute magnitude computed using the simple distance-parallax relation and GDR3 parallax data, and the absolute magnitude using Bailer-Jones distances that were calculated from a posterior probability distribution. References: Bailer-Jones et al. (2021) and references therein.

the minimal difference between GDR3 and Bailer-Jones parallaxes, with the residual converging to a small value near zero.

Figure 17 compares the two absolute magnitude vs. period relations: one using the distances computed with a simple inversion of the distance-parallax relation with parallax data from the GDR3 catalog, the other using distances determined through Bayesian inference as de-

tailed in Bailer-Jones et al. (2021). The error bars in the plot are absolute magnitude uncertainties, which were calculated using error propagation of apparent magnitude errors, period estimate errors, and distance propagated errors (sections B.1, B.2, and B.4). Both the GDR3 and Bailer-Jones sample demonstrate a clear absolute-magnitude-period relation, with some sources that deviate from the median due to inaccurate parallax measurements that have not been filtered out. From the plot, it is evident that the absolute magnitudes that used distances calculated from GDR3 parallaxes are larger in magnitude than the absolute magnitudes calculated using Bailer-Jones geometric distances. Since the difference between the two samples are not significant, in theory either distances can be used to estimate the absolute magnitudes. However, since Bailer-Jones et al. (2021) employed an algorithm and physical prior model that better corrects for the non-linear relationship between distance and parallax, it is more accurate to create the dust map using Bailer-Jones distances.

3.5. MCMC Absolute-Magnitude-Period Fitting

The method of fitting a model to a dataset using a Markov Chain Monte Carlo (MCMC) sampler provides a posterior distribution of fitting parameters, enabling the determination of how well the model fits to the data and how the parameters are correlated. Figure 17 from section 3.4 shows that there is a correlation between absolute magnitude and period, but there is still a significant spread in the data around this median. An intrinsic fit can be found by using MCMC samplers. Three methods of MCMC fitting were employed with the use of two MCMC samplers: Metropolis-Hastings and the No-U-Turn Sampler; see section 2.5 for details.

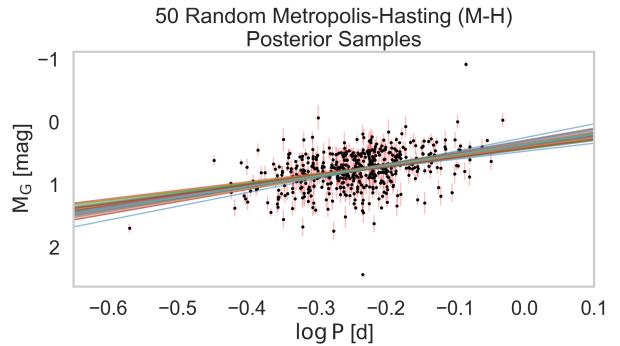


Figure 18. A plot of 50 random, independent M-H samples from the posterior over the data calculated from Bailer-Jones distances.

Starting with the Metropolis-Hastings sampler, figure 18 shows 50 random and independent posterior samples plotted over the absolute magnitudes calculated from bairer-Jones distances. In general, the models from the

591 50 posterior samples fit well to the intrinsic absolute-
 592 magnitude-period relation of the data. The spread be-
 593 tween the individual samples, as a function of period, ap-
 594 pear to widen outside of the range $-0.3 \lesssim \log P \lesssim -0.2$
 595 or $0.5 \lesssim P \lesssim 0.6$.

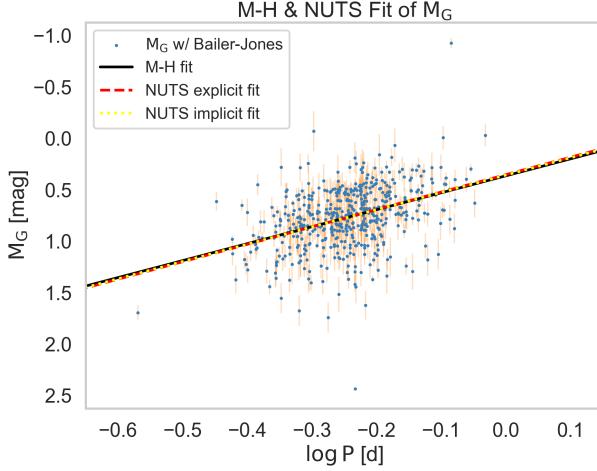


Figure 19. A plot of absolute magnitude calculated with Bailer-Jones distances, overplotted with linear model fits from the NUTS sampler. The mean posterior samples are as listed in table 1.

596 The next two fitting methods utilize the NUTS
 597 sampler. From figure 19, it is clear that there is no difference
 598 between the explicit implementation of the likelihood
 599 function (equation 12) as a potential term in the PYMC3
 600 model and the implicit implementation as described in
 601 section (2.5.2).

602 The mean values from the posterior probability dis-
 603 tributions of all three methods are listed in table 1.
 604 The corner plots in figure 20 and table 1 show that the
 605 three methods of fitting are in agreement, with the mean
 606 posteriors converging on similar values. As mentioned
 607 briefly in section 2.5, the Metropolis-Hastings algorithm
 608 is less efficient than the Hamiltonian Monte Carlo algo-
 609 rithm or its No-U-Turn sampler extension. The reason
 610 being the M-H algorithm goes through parameter space
 611 with the inefficient random walk method. Results in
 612 the rest of the paper will use NUTS to generate poste-
 613 rior probability distribution for fitting, since it does not
 614 require the manual optimization of step size Hoffman &
 615 Gelman (2011).

3.5.1. Comparison with WISE Dataset

617 The query to obtain the equivalent WISE dataset is
 618 described in section 2.5.3. Fitting of the near-infrared
 619 data was done used the PYMC3 model where the like-
 620 lihood function was implicitly defined. The posterior
 621 distribution is shown in the corner plots of figure 22. In
 622 comparison to the optical G-band posteriors (table 1,

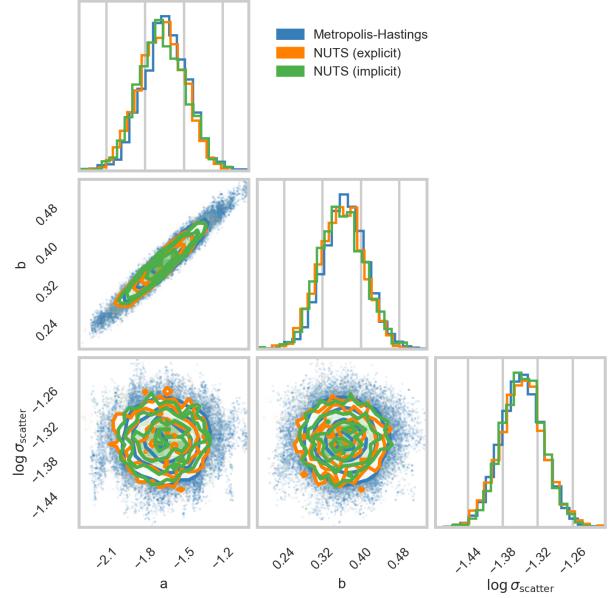


Figure 20. Corner plots of the posterior probability distributions for each of the three fitting methods. All histograms are normalized to unit area. Blue uses the M-H MCMC sampler, and both green and orange use NUTS. All three methods are in agreement for the values of the mean posteriors. The values of the mean posteriors can be found in table 1.

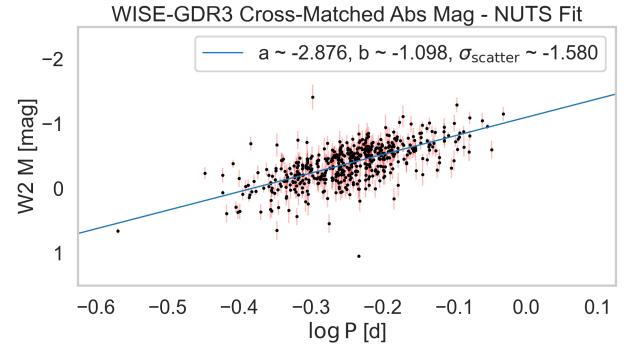


Figure 21. A plot of the absolute-magnitude-period relation in the near-infrared, using a WISE dataset that was cross-matched with sources from GDR3. The legend lists the mean posteriors that are the fitting parameters in the assumed linear log model of equation 10.

623 the near-infrared model has a steeper luminosity-period
 624 relationship.

3.5.2. Comparison of Derived Period-Luminosity Relations to Literature

625 Here the derived period-luminosity relations for both
 626 the G-band, using GDR3 data, and the near-infrared,

Table 1. A table of mean values from posterior probability distributions, generated by MCMC samplers. The parameters are from the linear log space model of absolute magnitude in equation 10.

MCMC Sampler	a	b	$\log \sigma_{\text{scatter}}$
Metropolis-Hastings	-1.64 ± 0.17	0.37 ± 0.04	-1.35 ± 0.04
NUTS (explicit)	$-1.66^{+0.17}_{-0.19}$	0.36 ± 0.05	-1.35 ± 0.04
NUTS (implicit)	$-1.68^{+0.19}_{-0.17}$	0.36 ± 0.05	-1.35 ± 0.04

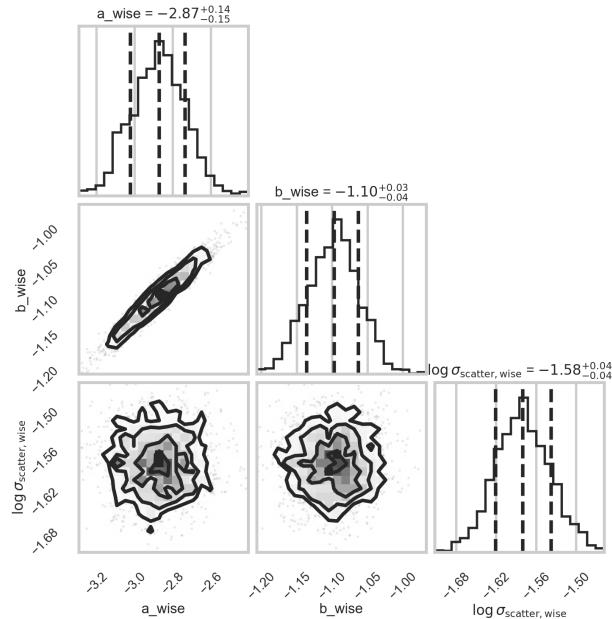


Figure 22. Corner plots of the posterior probability distribution of the fitting parameters for the absolute-magnitude-period relation of the WISE dataset. The quantiles shown are the lower 16th quantile and upper 84th quantile.

using WISE data, are compared to results from literature. The parameters for the G-band relations are shown in table 1. Beaton et al. (2018) uses the First Overtone Blue Edge (FOBE) method as a distance indicator to model their period-luminosity relation. They also accounted for mass and metallicity of the stars in their model, which were not used as parameters in equation 10. In figure 6 of Beaton’s paper, the V-band fit shows no slope. As the photometric bands progress farther into the infrared, however, the period-luminosity relation becomes progressively steep. The latter part is in agreement with the comparison to the WISE dataset that was collected in the infrared. The period-luminosity model for WISE exhibit a steeper slope than that of the G-band model, as discussed previously in section 3.5.1 and shown in figures 21 and 22. The disagreement between the G-band relations in table 1 and Beaton’s zero V-band slope can be attributed to the difference

in how the absolute magnitude was modelled. The linear model in equation 10 is much more simplistic than Beaton’s, which considered mass, metallicity, and a different period estimate for the RR Lyraes. The G-band and V-band are different photometric bands, however, they are still close enough in the optical band that one should not have a zero period-luminosity relation while the other does. Beaton’s period-luminosity model for the near-infrared for the I, J, H, and K bands, in ascending order of wavelengths, have slopes in the range $-2.5 < a < -0.5$, with the steeper slopes corresponding to the longer wavelength. The near-infrared model in figures 21 and 22 have steep slope values greater than 2.0, suggesting that the WISE dataset is past the near-infrared. This is consistent with the w2mpro magnitude used in fitting the WISE dataset, which is centered around $4.6 \mu\text{m}$ or 4600 nm in the mid-infrared (Klein & Bloom (2014)).

Figure 10 from Klein & Bloom (2014) shows the contour density plot and histogram for the slope and intercept parameter in their linear model for the period-luminosity relation in the V-band. Klein and Bloom found that the slope converged on $a \approx -0.4 \pm 0.2$ and the intercept converged on $b \approx 0.43 \pm 0.01$. The intercept is larger than the intercept found for the G-band, as shown in table 1. The slope are wildly different, however. Klein’s and Bloom’s value of $a \approx -0.4$ is in agreement with Beaton’s result that the V-band model exhibit a zero, or in this case a flat, slope. Similarly, figure 18 from Klein and Bloom shows the contour density plot and histogram for the fitting parameters in the W2 band, or the mid-infrared band centered around 4600 nm. The slope converged at $a \approx -2.25 \pm 0.15$ and the intercept converged at $b \approx -0.46 \pm 0.01$. The posterior distributions for the WISE data fit in figure 22 are not in agreement with these values.

The deviation of the slope from both literature results can be attributed to the more simplistic model used in equation 10, considering only 2 free parameters. Klein & Bloom (2014) used parallax data from Hipparcos and the Hubble Space Telescope in addition to W2 magnitude data from WISE, which might account for the difference. The RR Lyrae sources investigated are unlikely to be the same. Overall that should not affect the in-

691 trinsic period-luminosity relation, so it is unlikely that
 692 different sources can account for the differences.

693 **3.6. 2D Galactic Dust Map from G-band Extinction of
 694 RR Lyraes**

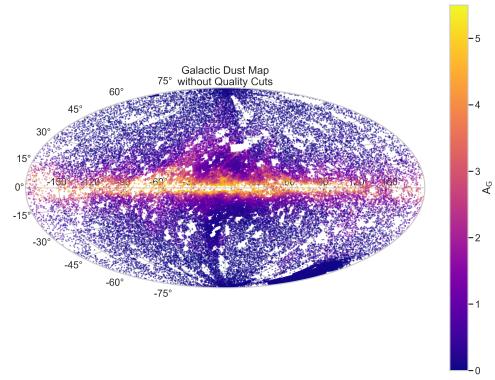


Figure 23. A 2D galactic dust map constructed from the G-band extinction A_G of RR Lyrae sources from the Gaia catalog, queried without any quality cuts on photometric signal-to-noise or BP/RP excess. The color map was created using the calculated A_G values.

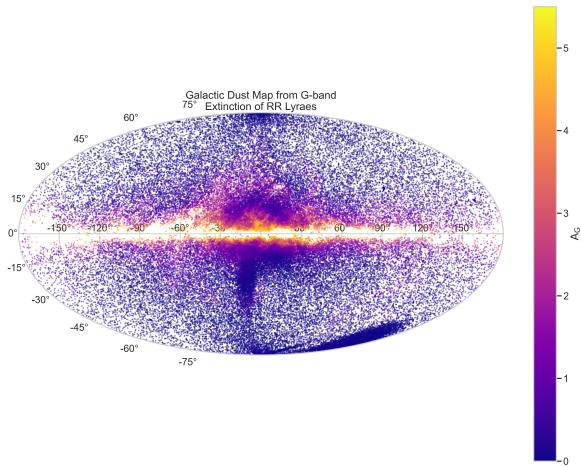


Figure 24. The same galactic dust map as in figure 23, but with quality cuts to photometric signal-to-noise and BP/RP excess to smooth out incorrect extinction values.

695 The initial dust map in figure 23 shows the presence
 696 of large-scale structure, such as the galactic disk with
 697 the most reddening from extinction. However, there are
 698 regions in which the A_G values are not consistent with
 699 the surroundings, owing to the photometric noise inac-
 700 curate color excess. To remedy this, additional quality

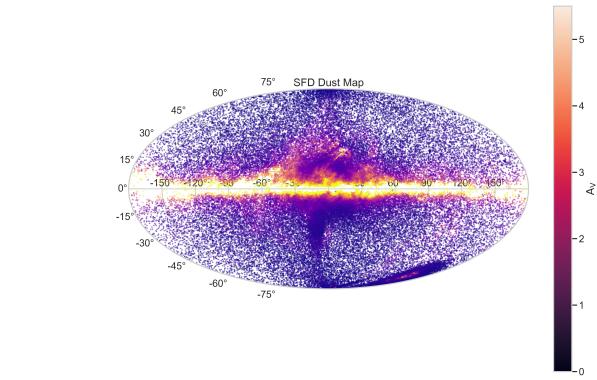


Figure 25. A galactic dust map from Schlegel, Finkbeiner, and Davis (SFD), using the same RR Lyrae galactic coordinates as figures 23 and 24. The color excess and extinction are both in the V-band and were queried using the SFD dustmap package.

701 cuts were implemented, as discussed in section 2.6. The
 702 quality cuts on photometric signal-to-noise and BP/RP
 703 excess produce the cleaned dust map in figure 24. Ob-
 704 serve that the A_G color map becomes more consistent
 705 throughout the map. The distribution of RR Lyrae stars
 706 is not uniform. Figure 25 uses color excess and extinc-
 707 tion values in the V-band of dust data as calculated by
 708 Schlegel, Finkbeiner, and Davis (Schlegel et al. (1998)).
 709 The large scale structure of the cleaned dust map pro-
 710 duced based on the G-band extinction of RR Lyraes is
 711 in good agreement with the SFD map. The SFD map
 712 has a smoother gradient of extinction values around the
 713 galactic bulge. It is to be expected that the dust map
 714 constructed from RR Lyrae color excess will not match
 715 exactly with the SFD map, since SFD was constructed
 716 using direct observation of emitted light from dust. As
 717 a consequence, the SFD map is more accurate and rep-
 718 resentative as a dust map of the galaxy.

4. CONCLUSIONS

720 The dust map produced from extinction values of RR
 721 Lyrae stars demonstrate the same overall large scale
 722 galactic structure as the dust map produced by SFD.
 723 The larger presence of extinction near the galactic bulge
 724 and disk lines up with the knowledge of dust distribu-
 725 tion in these regions of the galaxy. However, on close in-
 726 spection it is still evident that extinction values remain
 727 inconsistent between the surroundings for some regions
 728 of the dust map. There are many sources for this dis-
 729 crepancy with the SFD dust map. It can be attributed
 730 to photometric noise, which can be further reduced with
 731 more detailed analysis methods or better quality cuts in
 732 the ADQL query. The intrinsic color model used is also
 733 simplistic, with only 2 free parameters. It is possible

that following the examples of Beaton et al. (2018) and Klein & Bloom (2014) to include more free parameters that model other physical quantities, such as mass and metallicity, will help improve the extinction gradient on the local scale.

ACKNOWLEDGEMENTS

Sincerest thanks to Professor Dan Weisz, and graduate student instructors Anna Pusack and Olivia Aspegren for guidance and instruction on the conceptual

and technical components throughout this lab. Thank you to Talia Saarinen, Andrew Goh, Jason Wong, and Bradley Arias for helping me troubleshoot my python code. This work utilizes data from the European Space Agency (ESA) mission Gaia, processed by the Gaia Data Processing and Analysis Consortium (DPAC). Funding for DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral agreement.

APPENDIX

A. BAYESIAN INFERENCE

A better estimate of distance is the probabilistic approach, as described by Bailer-Jones et al. (2021), involving the use of Bayesian parameter inference. In Bayesian parameter inference, a prior function, built upon previous knowledge (e.g. a physical model), is updated to account for new data. Written in terms of probability distributions, Bayesian inference takes on the form

$$p(\theta|\mathbf{X}) \propto p(\mathbf{X}|\theta)p(\theta) \quad (\text{A1})$$

where $p(\theta)$ is the prior, $p(\mathbf{X}|\theta)$ is the likelihood function, and $p(\theta|\mathbf{X})$ is the posterior probability distribution, or the distribution of our fitting parameters (e.g. slope, intercept, uncertainty, etc.). The parameter of interest is denoted by θ , and the data is denoted by \mathbf{X} .

B. ERROR PROPAGATION

For a function $Z(A, B, \dots)$, where A, B, \dots have uncertainties $\sigma_A, \sigma_B, \dots$, the error propagation for the uncertainty in Z is given by

$$\sigma_Z = \sqrt{\left(\frac{\partial Z}{\partial A}\right)^2 \sigma_A^2 + \left(\frac{\partial Z}{\partial B}\right)^2 \sigma_B^2 + \dots} \quad (\text{B2})$$

The rest of this section will collect error propagation formulas used for this paper.

B.1. Apparent Magnitude Uncertainty

From the apparent magnitude relationship to flux

$$m = -2.5 \log_{10} F, \quad (\text{B3})$$

the uncertainty in apparent magnitude is

$$\sigma_m = \frac{2.5}{\ln 10} \frac{\sigma_F}{F}. \quad (\text{B4})$$

B.2. Distance from Parallax Uncertainty

From the distance and parallax relationship

$$d \approx 1/\varpi \quad (\text{B5})$$

the uncertainty in the distance is

$$\sigma_d^2 = \left(\frac{\partial d}{\partial \varpi}\right)^2 \sigma_\varpi^2 \rightarrow \sigma_d = \frac{1}{\varpi^2} \sigma_\varpi = d \frac{\sigma_\varpi}{\varpi} \quad (\text{B6})$$

B.3. Parallax from Distance Uncertainty

$$\sigma_\varpi = \sigma_d/d^2 \quad (\text{B7})$$

B.4. Absolute Magnitude Uncertainty

Starting from absolute magnitude equation

$$M = m - 5 \log_{10} (d) + 5 = m - 5 \log_{10} \left(\frac{d}{10} \right) \quad (\text{B8})$$

the uncertainty is

$$\sigma_M = \sqrt{\sigma_m^2 + \left(\frac{-5}{d \ln 10} \right)^2 \sigma_d^2} \quad (\text{B9})$$

C. METROPOLIS-HASTINGS MCMC SAMPLER

The sampler was used to draw 10,000 samples from a 1D Gaussian distribution

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(\mu-x)^2}{2\sigma^2}\right) \quad (\text{C10})$$

with $\mu = 1$ and $\sigma = 0.1$.

REFERENCES

- Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, AJ, 161, 147, doi: [10.3847/1538-3881/abd806](https://doi.org/10.3847/1538-3881/abd806)
- Beaton, R. L., Bono, G., Braga, V. F., et al. 2018, SSRv, 214, 113, doi: [10.1007/s11214-018-0542-1](https://doi.org/10.1007/s11214-018-0542-1)

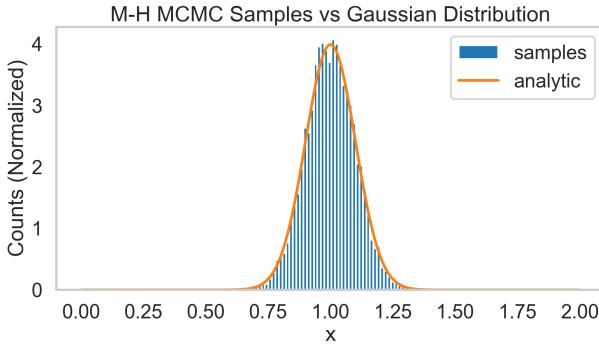


Figure 26. A histogram of the posterior probability distribution from a M-H MCMC sampler compared to the analytic Gaussian distribution. The histogram was normalized to unit area. The sampler drew 10,000 samples from a 1D Gaussian distribution with $\mu = 1$ and $\sigma = 0.1$.

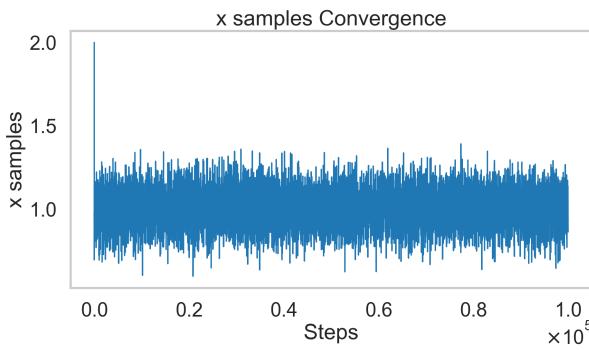


Figure 27. A plot of the x samples vs. the number of steps to qualitatively demonstrate the convergence of x to $\mu = 1$

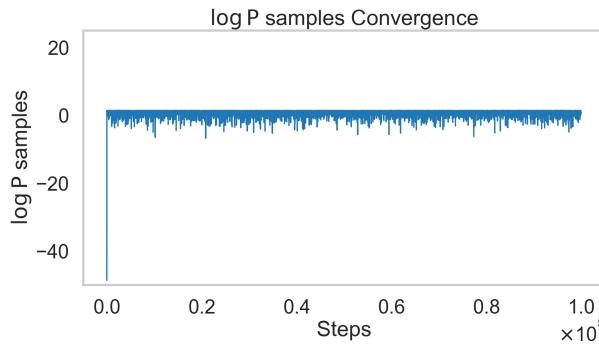


Figure 28. A plot of the log of the posterior samples $\log P$ vs. the number of steps to qualitatively demonstrate convergence.

- 792 Clementini, G., Ripepi, V., Leccia, S., et al. 2016, *A&A*,
595, A133, doi: [10.1051/0004-6361/201629583](https://doi.org/10.1051/0004-6361/201629583)
- 794 Clementini, G., Ripepi, V., Garofalo, A., et al. 2023, *A&A*,
674, A18, doi: [10.1051/0004-6361/202243964](https://doi.org/10.1051/0004-6361/202243964)
- 796 Hoffman, M. D., & Gelman, A. 2011, arXiv e-prints,
arXiv:1111.4246, doi: [10.48550/arXiv.1111.4246](https://doi.org/10.48550/arXiv.1111.4246)
- 798 Klein, C. R., & Bloom, J. S. 2014, arXiv e-prints,
arXiv:1404.4870, doi: [10.48550/arXiv.1404.4870](https://doi.org/10.48550/arXiv.1404.4870)
- 800 Lindegren, L., Hernández, J., Bombrun, A., et al. 2018,
A&A, 616, A2, doi: [10.1051/0004-6361/201832727](https://doi.org/10.1051/0004-6361/201832727)
- 802 Netzel, H., Smolec, R., Soszyński, I., & Udalski, A. 2018,
MNRAS, 480, 1229, doi: [10.1093/mnras/sty1883](https://doi.org/10.1093/mnras/sty1883)
- 804 Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*,
500, 525, doi: [10.1086/305772](https://doi.org/10.1086/305772)
- 806 Soszyński, I., Dziembowski, W. A., Udalski, A., et al. 2011,
AcA, 61, 1, doi: [10.48550/arXiv.1105.6126](https://doi.org/10.48550/arXiv.1105.6126)
- 808 Soszyński, I., Udalski, A., Szymański, M. K., et al. 2014,
AcA, 64, 177, doi: [10.48550/arXiv.1410.1542](https://doi.org/10.48550/arXiv.1410.1542)