

Creating Dust Map of the Milky Way

Enmanuel Hernandez

March 4, 2024

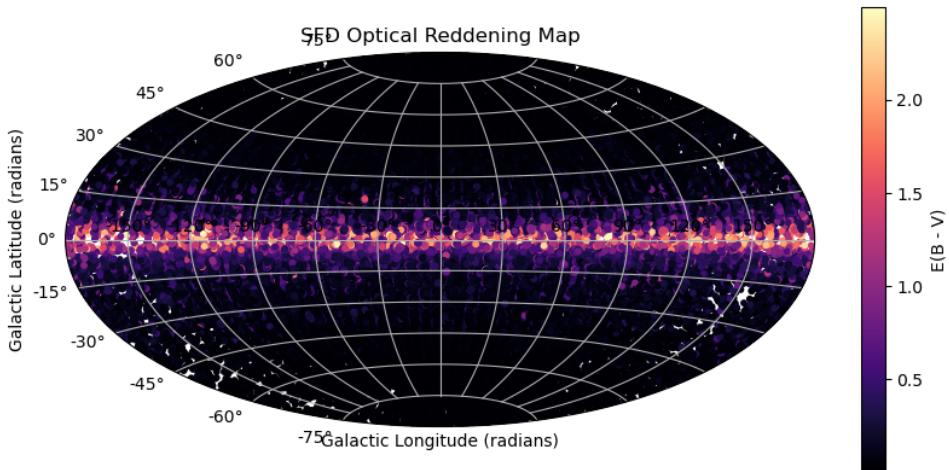


Figure 1: SFD Dust Map of the Milky Way colored with extinction levels

1 Introduction

In this paper, we delve into the world of RR Lyrae stars and their role in unraveling the mysteries of Galactic dust, utilizing data from the groundbreaking Gaia mission. RR Lyrae stars, pulsating variables found primarily in old stellar populations, serve as crucial standard candles for measuring distances in the Milky Way and beyond. Their characteristic light curves and well-defined period-luminosity relation make them invaluable tools for probing the structure and evolution of our Galaxy. The Gaia mission, with its unprecedented astrometric precision, has revolutionized our understanding of the Milky Way, providing a detailed map of the positions, motions, and luminosities of over a billion stars. By combining Gaia's data with observations of RR Lyrae stars, we can trace the distribution of interstellar dust, which scatters and absorbs starlight, affecting our measurements of stellar properties. Understanding the interplay between RR Lyrae stars and Galactic dust is essential for refining our models of Galactic structure and improving the accuracy of astronomical distance measurements.

Our primary tools will include periodograms for analyzing the periodicity of these stars, and Fourier decomposition to model their light curves. We will employ ADQL queries to access the wealth of data available in the Gaia DR3 and eDR3 catalogs. Our analysis will involve both linear and nonlinear optimization techniques to fit models to the data, as well as Markov chain Monte Carlo and Bayesian modeling for robust statistical inference. Throughout the lab, we will use Python to interact with web databases, perform our analyses, and create visualizations to interpret our results.

2 Methods

2.1 RR Lyrae Catalog Download

We start by querying the Gaia DR3 database to retrieve the light curves of RR Lyrae stars. We use the Astroquery library to download the first 100 rows of the gaiadr3.varrrlyrae catalog, which

contains stars with a measured fundamental pulsation frequency and more than 40 clean epochs in the G-band. We then extract the G-band magnitude and its error from the downloaded data, plot the light curve for one of the stars, and display the magnitude uncertainties as error bars on the plot.

```
query = """
SELECT TOP 100
    solution_id ,
    source_id ,
    pf ,
    pf_error ,
    num_clean_epochs_g
FROM
    gaiadr3.vari_rrlyrae
WHERE
    pf IS NOT NULL AND num_clean_epochs_g > 40
"""

```

2.2 Light Curve Retrieval

We are now downloading the raw light curves for the 100 RR Lyrae stars from the Gaia DR3 catalog. Using the Gaia's load data function, we retrieve the epoch photometry data, which includes G-band magnitudes, fluxes, and their errors, for each star. We then plot the light curve for one of these stars, displaying the G-band magnitude versus time with error bars representing the magnitude uncertainties.

```
retrieval_type = 'EPOCH_PHOTOMETRY'
data_structure = 'COMBINED'
data_release = 'Gaia_DR3'
datalink = Gaia.load_data(ids=results['source_id'], data_release = data_release, retr
dl_keys = [inp for inp in datalink.keys()]
dl_key = 'EPOCH_PHOTOMETRY_COMBINED.xml'
product = datalink[dl_key][0]
product_tb = product.to_table() # Exporting to Astropy Table object.
```

2.3 Period and Mean Magnitude Estimation

We then start estimating the period and mean G-band magnitude of 100 light curves using a Lomb-Scargle periodogram, which is a common method for analyzing unevenly spaced time series data to detect periodic signals. We first convert the observed flux to magnitudes, the flux magnitude equation is given by:

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

where m is the magnitude, F is the flux, and C is a constant that depends on the specific photometric system being used. Then we use the periodogram to find the frequency with the highest power, which corresponds to the best estimate of the period. For each light curve, we calculate the mean flux and convert it back to the mean magnitude, taking into account that magnitudes are logarithmic quantities. Finally, we plot the periodogram for one of the light curves, highlighting the estimated period, and print the estimated period and mean G-band magnitude for that light curve.

```
time = np.array([row['time'] for row in specific_source_data])
mag = np.array([row['mag'] for row in specific_source_data])
# Converting magnitude to relative flux
flux = 10**(-0.4 * (mag - np.median(mag)))
# Lomb-Scargle Periodogram to estimate the period
ls = LombScargle(time, flux)
frequency, power = ls.autopower(minimum_frequency = 1, maximum_frequency = (1/0.3))
best_frequency = frequency[np.argmax(power)]
```

```
best_period = 1 / best_frequency
```

2.4 Period Comparison

We compare the periods we computed from the 100 light curves to the values reported in the variablyrae catalog by querying the Gaia DR3 database. The query selects the top 100 rows where a fundamental pulsation frequency (pf) has been measured and more than 40 clean epochs were obtained in the G-band. We then execute the query asynchronously using Astroquery, retrieve the results, and print the first 100 rows to compare with our computed periods.

```
query2 = """
SELECT TOP 100
    source_id ,
    pf ,
    int_average_g ,
    num_clean_epochs_g
FROM
    gaiadr3.vari_rrlyrae
WHERE
    pf IS NOT NULL AND num_clean_epochs_g > 40
"""
```

2.5 Series Representation for a Star

We first collect data for the star with Gaia DR3 id = 5817567360327589632 and filter it for the G-band. We then compute the periodogram to find the best period for the light curve. Using this period, we phase the light curve and plot it for Fourier series representations with different numbers of terms ($K = 1, 3, 5, 7, 9$). For each K , we fit the phased light curve with the Fourier series,

$$f(t) = A_0 + \sum_{k=1}^K [a_k \sin(k\omega t) + b_k \cos(k\omega t)],$$

where A_0 , a_k , and b_k are to-be-determined constants (a_k and b_k are arrays that are K terms long), and $\omega = 2\pi/P$ is the angular frequency. Plot the fit, and calculate the residuals between the original data and the fit. Finally, we visualize the original phased light curve, the Fourier fits, and the residuals in a single plot for each K , with the magnitude axis inverted as is customary for astronomical magnitude plots.

2.6 Optimal Number of Terms

We then aim to find the optimal number of terms, K , for a Fourier series that best fits our data. We split the data into training (80 percent) and validation (20 percent) sets. For each K value ranging from 1 to 25, we fit a Fourier series to the training data and predict the fluxes for both the training and validation sets. We then calculate X Squared/Ndata for both sets. Finally, we plot the parameters for each K value to visually assess the fit and determine the appropriate value of K for our dataset.

```
# Splitting the data: 80% for training , 20% for validation
times_train , times_val , fluxes_train , fluxes_val = train_test_split(times , fluxes , test_size = .2)
# Known period P and angular frequency w
P = .5 # Determined from Lomb-Scargle periodogram
omega = 2 * np.pi / P
# Fourier series a FUNCTION
def fourier_series(t , A0 , *coeffs):
    K = len(coeffs) // 2
    result = A0
    for k in range(1 , K + 1):
        a_k , b_k = coeffs[2*k - 2] , coeffs[2*k - 1]
        result += a_k * np.sin(k * omega * t) + b_k * np.cos(k * omega * t)
```

```

    return result
# Fitting the Fourier series for a range of K values and calculate Chi-Squared/N_data

```

2.7 Future Magnitude Prediction

We then utilize a Fourier series model with 10 terms ($K = 10$) to predict the expected magnitude of an RR Lyrae star 10 days after the last observed data point from the Gaia archive. We calculate the angular frequency (ω) based on the period of the star, and then construct the Fourier series using the determined coefficients. We plot the last few observed data points from Gaia, the extrapolation over the next 12 days, and mark the predicted magnitude exactly 10 days after the last observation with a green dashed line. The first plot provides a general view, while the second plot zooms in for a closer look at the extrapolation and the predicted point. This approach allows us to estimate the future brightness of the star based on its past light curve behavior.

```

# Fourier series a FUNCTION using k = 10
def fourier_series(t, A0, *coeffs):
    K = 10
    result = A0
    for k in range(1, K + 1):
        a_k, b_k = coeffs[2*k - 2], coeffs[2*k - 1]
        result += a_k * np.sin(k * omega * t) + b_k * np.cos(k * omega * t)
    return result

```

2.8 Mean Magnitude Estimation with Fourier Model

We then calculate the mean magnitude for each of the first 100 RR Lyrae stars using the same Fourier model. We first estimate the period using a Lomb-Scargle periodogram, then construct a design matrix for the Fourier series with the best number of harmonics, and solve for the coefficients using least squares. Using these coefficients, we calculate the fitted fluxes, from which we obtain the mean flux and convert it to mean magnitude. We then compare these estimates with the mean magnitudes from part 3 and the Gaia catalog, and plot both the estimates and the residuals to analyze the accuracy of our Fourier model.

2.9 Comparing RRab and RRc Light Curve Shapes

We then first define a function to plot phased light curves and their Fourier models for RR Lyrae stars. We then query the Gaia database to retrieve light curves for the top three RR Lyrae stars in both the RRc and RRab variability classes, ensuring they have mean G-band magnitudes brighter than 15 and more than 80 clean epochs in the G-band. For each star, we use the Lomb-Scargle method to estimate the period of the light curve and fit a Fourier series with a chosen number of terms ($K=10$) to model the periodic variation. Finally, we plot the observed light curves and their corresponding Fourier models, comparing the shapes of the light curves between the RRc and RRab classes to understand the differences in their pulsation modes. This analysis helps us explore the characteristics of different subclasses of RR Lyrae stars and their variability patterns.

```

# Query for RRc and RRab stars
for classification in [ 'RRc', 'RRab' ]:
    query = f """
        SELECT TOP 3
            source_id,
            pf,
            int_average_g
        FROM
            gaiadr3.vari_rrlyrae
        WHERE
            best_classification = '{classification}'
            AND int_average_g < 15
            AND num_clean_epochs_g > 80
    """

```

```

ORDER BY
    int_average_g ASC
"""

```

2.10 RR Lyrae Period-Luminosity

We then are specifically targeting RR Lyrae stars classified as 'RRab' within the Milky Way to establish a period-luminosity relation. We utilize the Gaia catalog to select stars that meet our criteria for reliable distance measurements and minimal interstellar dust interference. Our ADQL query filters these stars based on a parallax over error ratio greater than 5, ensuring accurate distance estimates. We also exclude stars within the Galactic disk by requiring an absolute Galactic latitude greater than 30 degrees. Additionally, we only consider stars with a parallax greater than 0.25 mas to focus on relatively nearby objects. This results in a refined sample of approximately 600 RR Lyrae stars suitable for our analysis.

```

# ADQL query to select RR Lyrae stars with specific criteria
query = """
SELECT
    source_id,
    ra,
    dec,
    parallax,
    phot_g_mean_mag,
    phot_bp_mean_mag,
    phot_rp_mean_mag,
    b,
    l,
    parallax_error,
    parallax_over_error
FROM
    gaiadr3.gaia_source
JOIN
    gaiadr3.vari_rrlyrae
USING (source_id)
WHERE
    parallax_over_error > 5
    AND abs(b) > 30
    AND parallax > 0.25
    AND best_classification = 'RRab'
"""

```

2.11 Distance Estimates with Galactic Structure Prior

We are now seeking an estimate of distance that incorporates our understanding of the Galactic structure, using a prior based on parallax measurements. This approach is detailed in Bailer-Jones et al. 2020, and the data can be accessed through the Gaia archive using the external.gaiadr3 distance catalog. Our ADQL query retrieves geometric distance estimates and uncertainties for RR Lyrae stars classified as 'RRab', ensuring they have accurate parallax measurements, are located away from the Galactic disk, and are within 4 kpc.

```

# ADQL query to select geometric distances for the RR Lyrae stars
query = """
SELECT
    gaia.source_id,
    rrlyrae.best_classification,
    gaia.ra,
    gaia.dec,
    gaia.parallax,
    gaia.phot_g_mean_mag,

```

```

gaia.phot_bp_mean_mag,
gaia.phot_rp_mean_mag,
dist.r_med_geo AS distance_median,
dist.r_lo_geo AS distance_lower,
dist.r_hi_geo AS distance_upper
FROM
    gaiadr3.gaia_source AS gaia
JOIN
    gaiadr3.vari_rrlyrae AS rrlyrae
    ON gaia.source_id = rrlyrae.source_id
JOIN
    external.gaiaedr3_distance AS dist
    ON gaia.source_id = dist.source_id
WHERE
    rrlyrae.best_classification = 'RRab'
    AND gaia.parallax_over_error > 5
    AND abs(gaia.b) > 30
    AND gaia.parallax > 0.25
"""

```

2.12 Galactic Distribution of Targets

Then we first plot the distribution of RR Lyrae stars in galactic coordinates, ensuring that stars in the galactic disk are excluded as per our ADQL query. We then compare the distance estimates obtained from the Bailer-Jones catalog with naive distance estimates calculated from the inverse of the parallax, highlighting the trend that the naive method tends to underestimate distances for more distant stars. Additionally, we plot the EDR3 parallaxes against the Bailer-Jones parallaxes, including uncertainties, to assess their consistency. Lastly, we calculate the average difference between the EDR3 and Bailer-Jones parallaxes to quantify the discrepancy.

2.13 Period vs. Absolute Magnitude Plot

We then perform a query to the Gaia database to retrieve information on RR Lyrae stars, including their periods, parallaxes, and magnitudes. We join the gaiadr3.gaia_source, gaiadr3.varirrlyrae, and external.gaiaedr3distance tables to obtain the necessary data, specifically focusing on RR Lyrae stars classified as 'RRab' with specific criteria on parallax and galactic latitude. We then calculate the absolute G-band magnitude for each star using the Bailer-Jones distances, taking into account the distance uncertainties. The absolute magnitude M is given by $M = m - 5(\log_{10} d + 5)$, where m is the apparent magnitude and d is the distance in parsecs. Finally, we plot the period versus absolute G-band magnitude for the RR Lyrae stars, including error bars for both period and magnitude, and invert the y-axis to display the magnitude in the conventional astronomical way.

```

# ADQL query with geometric distances for the RR Lyrae stars with period info
query = """
SELECT
    gaia.source_id,
    gaia.ra,
    gaia.dec,
    rrlyrae.best_classification,
    gaia.parallax,
    gaia.phot_g_mean_mag,
    gaia.phot_bp_mean_mag,
    gaia.phot_rp_mean_mag,
    gaia.astrometric_chi2_al,
    gaia.astrometric_n_good_obs_al,
    gaia.phot_bp_rp_excess_factor,
    dist.r_med_geo AS distance_median,
    dist.r_lo_geo AS distance_lower,
    dist.r_hi_geo AS distance_upper,

```

```

rrlyrae . pf AS period ,
rrlyrae . pf_error AS period_error ,
rrlyrae . int_average_g
FROM
    gaiadr3 . gaia_source AS gaia
JOIN
    gaiadr3 . vari_rrlyrae AS rrlyrae
    ON gaia . source_id = rrlyrae . source_id
JOIN
    external . gaiaedr3_distance AS dist
    ON gaia . source_id = dist . source_id
WHERE
    rrlyrae . best_classification = 'RRab'
    AND gaia . parallax_over_error > 5
    AND abs(gaia . b) > 30
    AND gaia . parallax > 0.25
"""

```

2.14 Scatter Reduction with Quality Cuts

We initially observe that most RR Lyrae stars have similar absolute magnitudes, but some exhibit significant scatter due to incorrect parallax measurements. To address this, we apply quality cuts based on Equations C1 and C2 from Lindegren et al. 2018, C1:

$$\left(\frac{\chi^2}{\nu} \right)^{\frac{1}{2}} < 1.2 \times \max(1, \exp(-0.2(G - 19.5))) . \quad (2)$$

C2:

$$1.0 + 0.015(\text{GBP} - \text{GRP})^2 < \text{phot_bp_rp_excess_factor} < 1.3 + 0.06(\text{GBP} - \text{GRP})^2 \quad (3)$$

which involve conditions on the photbp rpexcessfactor and astrometricchi2al parameters. These cuts aim to filter out stars with potentially unreliable data. After applying these additional criteria, we query the Gaia database for RR Lyrae stars that meet these refined conditions, focusing on the RRab subclass. Finally, we plot the period-luminosity relation again, incorporating error bars for both period and distance, and observe whether the scatter in the data has decreased, indicating an improvement in the reliability of our sample.

```

# ADQL query with geometric distances for the RR Lyrae stars with period info AND A da
query = """
SELECT
    gaia . source_id ,
    gaia . ra ,
    rrlyrae . best_classification ,
    gaia . dec ,
    gaia . parallax ,
    gaia . phot_g_mean_mag ,
    gaia . phot_bp_mean_mag ,
    gaia . phot_rp_mean_mag ,
    gaia . astrometric_chi2_al ,
    gaia . astrometric_n_good_obs_al ,
    gaia . phot_bp_rp_excess_factor ,
    dist . r_med_geo AS distance_median ,
    dist . r_lo_geo AS distance_lower ,
    dist . r_hi_geo AS distance_upper ,
    rrlyrae . pf AS period ,
    rrlyrae . pf_error AS period_error ,
    rrlyrae . int_average_g
FROM

```

```

gaiadr3.gaia_source AS gaia
JOIN
    gaiadr3.vari_rrlyrae AS rrlyrae
    ON gaia.source_id = rrlyrae.source_id
JOIN
    external.gaiadr3_distance AS dist
    ON gaia.source_id = dist.source_id
WHERE
    rrlyrae.best_classification = 'RRab',
    AND gaia.parallax_over_error > 5
    AND abs(gaia.b) > 30
    AND gaia.parallax > 0.25
    AND (1.0 + 0.015 * POWER(gaia.phot_bp_mean_mag - gaia.phot_rp_mean_mag, 2)) < gaia.phot_bp_rp_excess_factor
    AND gaia.phot_bp_rp_excess_factor < (1.3 + 0.06 * POWER(gaia.phot_bp_mean_mag - gaia.phot_rp_mean_mag, 2))
    AND POWER(gaia.astrometric_chi2_al / gaia.astrometric_n_good_obs_al, 0.5) < 1.2 *
"""
# Calculating absolute G-band magnitude using Bailer-Jones distances
distance_median = results3['distance_median']
distance_error = (results3['distance_upper'] - results3['distance_lower']) / 2
abs_mag_g = results3['int_average_g'] - 5 * np.log10(distance_median) + 5
results3['abs_mag_g'] = abs_mag_g

```

2.15 Outlier Removal

We then first refine our sample of RR Lyrae stars by applying additional criteria to remove outliers based on their absolute G-band magnitudes, in this case by observing where most of the cluster is (from .5 - 1.2). We then calculate the error in the absolute magnitudes, taking into account the uncertainties in the distance measurements. Finally, we create a scatter plot of the period versus absolute G-band magnitude for the cleaned sample, including error bars to represent the uncertainties in both the period and the magnitude.

```

results4 = results3[(results3['abs_mag_g'] > .5) & (results3['abs_mag_g'] < 1.2)]
# Continue for Plotting period vs. absolute G-band magnitude

```

2.16 Comparison with Gaia DR3 Parallaxes

Furthermore we first join our RR Lyrae data with the geometric distance estimates from the Gaia EDR3 catalog, matching them based on their source IDs. We then calculate the absolute magnitudes of these stars using their Gaia DR3 parallaxes, applying the standard formula for absolute magnitude. Finally, we plot the period-absolute magnitude relation for these stars, using the calculated absolute magnitudes and the periods from our RR Lyrae data, to visualize the relationship between these two properties.

2.17 Fitting Period-Luminosity Relation with M-H MCMC (i)

Now we're fitting a line to the period vs. absolute magnitude relation for RR Lyrae stars in the G band using DR3 data. We assume a model where the absolute magnitude (MG) is a function of the logarithm of the period, with a linear relationship and a free parameter for intrinsic scatter (sigma scatter). To estimate the parameters of this model, we implement our own Metropolis-Hastings (M-H) Markov Chain Monte Carlo (MCMC) sampler, using a Gaussian proposal distribution.

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

We test the sampler by generating 10,000 samples from a one-dimensional Gaussian distribution and comparing the resulting histogram to the true Gaussian distribution. The acceptance rate of the sampler is monitored to ensure efficient sampling, and we plot the sampled values and their logarithmic probabilities against the step number to demonstrate convergence.

```

# Gaussian PDF
def gaussian_pdf(x, mu, sigma):
    return (1 / np.sqrt(2 * np.pi * sigma**2)) * np.exp(-((mu - x)**2) / (2 * sigma**2))
# Logarithm of the Gaussian PDF
def log_gaussian_pdf(x, mu, sigma):
    return -0.5 * np.log(2 * np.pi * sigma**2) - ((mu - x)**2) / (2 * sigma**2)
# Metropolis-Hastings sampler
def mh_sampler(log_P, n_samples, proposal_width, init_params):
    samples = [init_params]
    log_probs = [log_P(init_params)]
    acceptance_count = 0
    for _ in range(n_samples):
        current_params = samples[-1]
        proposed_params = np.random.normal(current_params, proposal_width)
        log_acceptance_ratio = log_P(proposed_params) - log_P(current_params)
        if np.log(np.random.rand()) < log_acceptance_ratio:
            samples.append(proposed_params)
            log_probs.append(log_P(proposed_params))
            acceptance_count += 1
        else:
            samples.append(current_params)
    acceptance_rate = acceptance_count / n_samples
    log_probs.append(log_P(samples[-1]))
    return np.array(samples), np.array(log_probs), acceptance_rate

```

2.18 Fitting Period-Luminosity Relation with M-H MCMC (ii)

We'd like to first define the data and parameters for the period-luminosity relation of RR Lyrae stars. We then construct a likelihood function that models the absolute magnitudes as a function of the logarithm of the periods from Gaia Catalog, with a Gaussian scatter characterized by a parameter σ scatter. We also define a prior function that incorporates our prior knowledge about the parameters, including their means and standard deviations. We then combine the likelihood and prior to form a posterior function, which we sample using a 3D Metropolis-Hastings algorithm. This sampler generates a sequence of parameter values, where each new value is proposed based on the current value and accepted with a probability that depends on the ratio of the posterior probabilities. We aim for an acceptance rate of about 0.5 to ensure efficient sampling as always. Finally, we use the corner package to visualize the posterior distribution of the parameters, and we randomly select 50 samples from the posterior to plot the corresponding model predictions for the period-luminosity relation. This allows us to see the range of plausible models given the data and our prior knowledge.

```

# Likelihood function
def log_likelihood(params, periods, abs_mags, abs_mag_err):
    a, b, ln_sigma_scatter = params
    sigma_scatter = np.exp(ln_sigma_scatter)
    model_mags = a * np.log10(periods) + b
    total_sigma = np.sqrt(sigma_scatter**2 + abs_mag_err**2)
    return -0.5 * np.sum(((abs_mags - model_mags) / total_sigma) ** 2)

prior_a_mean = -0.4
prior_a_sd = 0.1
prior_b_mean = 0.5
prior_b_sd = 0.1
# Prior function with prior knowledge of parameters
def log_prior(params):
    a, b, ln_sigma_scatter = params
    sigma_scatter = np.exp(ln_sigma_scatter)
    if sigma_scatter <= 0:
        return -np.inf
    return -0.5 * ((a - prior_a_mean) / prior_a_sd)**2 \
        - 0.5 * ((b - prior_b_mean) / prior_b_sd)**2 \

```

```

-0.5 * ((ln_sigma_scatter - np.log(0.191)) / 0.01)**2
# Posterior function
def log_posterior(params, periods, abs_mags, abs_mag_err):
    return log_prior(params) + log_likelihood(params, periods, abs_mags, abs_mag_err)
# Metropolis-Hastings sampler for the three-dimensional parameter space
def mh_sampler_3d(data, n_samples, proposal_width, init_params):
    samples = [init_params]
    acceptance_count = 0
    for _ in range(n_samples):
        current_params = samples[-1]
        proposed_params = np.random.normal(current_params, proposal_width)
        acceptance_ratio = np.exp(log_posterior(proposed_params, *data) - log_posterior(
            current_params, *data))
        if np.random.rand() < acceptance_ratio:
            samples.append(proposed_params)
            acceptance_count += 1
        else:
            samples.append(current_params)
    acceptance_rate = acceptance_count / n_samples
    return np.array(samples), acceptance_rate

```

2.19 Fitting Period-Luminosity Relation with Model PYMC

Now we fit the period-luminosity relation of RR Lyrae stars using the No-U-Turn Sampler (NUTS), a Hamiltonian Monte Carlo sampler provided by PyMC. We define a likelihood function that calculates the log-likelihood of our model given the parameters, periods, absolute magnitudes, and their uncertainties. We then set up our Bayesian model with priors for the slope (m), intercept (b), and the logarithm of the scatter (logsigma_scatter) of the period-luminosity relation. The model also includes a potential term that incorporates our likelihood function. We use the NUTS sampler to draw samples from the posterior distribution of our parameters. Finally, we visualize the results using the corner plot to show the posterior distributions and relationships between parameters, and we use ArviZ to plot the trace of the sampling process and the posterior distributions of each parameter.

```

# Likelihood function compatible with PyMC3
def log_likelihood(params, periods, abs_mags, abs_mag_err):
    a, b, log_sigma_scatter = params
    sigma_scatter = pm.math.exp(log_sigma_scatter)
    model_mags = a * pm.math.log(periods) + b
    total_sigma = pm.math.sqrt(sigma_scatter**2 + abs_mag_err**2)
    return -0.5 * pm.math.sum(((abs_mags - model_mags) / total_sigma) ** 2)
with pm.Model() as model:
    # Priors for parameters
    a = pm.Normal('a', mu=prior_a_mean, sigma=prior_a_sd)
    b = pm.Normal('b', mu=prior_b_mean, sigma=prior_b_sd)
    log_sigma_scatter = pm.Normal('log_sigma_scatter', mu=np.log(prior_sigma_scatter_mean),
        sigma=prior_sigma_scatter_sd)
    # Expected absolute magnitude (theoretical model)
    expected_abs_mag = a * pm.math.log(periods) + b
    # Likelihood (sampling distribution) of observations
    Y_obs = pm.Normal('Y_obs', mu=expected_abs_mag, sigma=pm.math.exp(log_sigma_scatter))
    # Potential term for the likelihood function
    potential = pm.Potential('likelihood', log_likelihood([a, b, log_sigma_scatter]),
        total_sigma)
    # Drawing samples from the posterior using the NUTS sampler
    trace = pm.sample(draws=1000, tune=500, target_accept=0.95)

```

2.20 Fitting Period-Luminosity Relation with Explicit Likelihood PYMC

Next we define priors for the slope (m), intercept (b), and the logarithm of the intrinsic scatter (logsigma_scatter) of the relation. The expected absolute magnitude is modeled as a linear function

of the logarithm of the period, with a normal likelihood function that accounts for both the intrinsic scatter and the measurement uncertainties in the magnitudes. We use the No-U-Turn Sampler (NUTS) to draw samples from the posterior distribution and visualize the results using corner plots, trace plots, and posterior plots.

with pm.Model() as model:

```
# Priors for parameters
a = pm.Normal('a', mu=0, sigma=10)
b = pm.Normal('b', mu=0, sigma=10)
log_sigma_scatter = pm.Normal('log_sigma_scatter', mu=np.log(0.2), sigma=.01)
# Expected absolute magnitude (theoretical model)
expected_abs_mag = a * pm.math.log(periods) + b
# Likelihood (sampling distribution) of observations
Y_obs = pm.Normal('Y_obs', mu=expected_abs_mag, sigma=pm.math.sqrt(pm.math.exp(log_sigma_scatter)))
# Drawing samples from the posterior using the NUTS sampler
traceii = pm.sample(draws=1000, tune=500, target_accept=0.95)
```

2.21 Corner Plots of ALL 3 Models

Finally we compare the posterior constraints obtained from three different approaches to fitting the period-luminosity relation of RR Lyrae stars using Markov chain Monte Carlo (MCMC) methods. We use the corner package to create a corner plot that visualizes the posterior distributions of the parameters (slope m, intercept b, and log of the scatter lns) for each model. The blue plot represents the results from a model where the parameters are directly specified, the green plot represents the results from a model with an explicitly defined likelihood function, and the red plot represents the results from using the Metropolis-Hastings algorithm.

2.22 Deriving Period-Color Relation

We then aim to derive a period-color relation for RR Lyrae stars using Gaia bands. We calculate the color uncertainties by using the uncertainties in BP and RP flux and create a PyMC model to fit the relationship between the logarithm of the period and the GBP - GRP color. We define priors for the intercept (alpha), slope (beta), and the intrinsic scatter (sigma) of the relation. The expected color is modeled as a linear function of the logarithm of the period, and we use a normal likelihood with a standard deviation that accounts for both the intrinsic scatter and the color uncertainties. Finally, we sample from the posterior distribution using PyMC's sampling function and plot the posterior distributions of the model parameters.

with pm.Model() as model:

```
# Priors
alpha = pm.Normal('alpha', mu=0, sigma=1)
beta = pm.Normal('beta', mu=0, sigma=1)
sigma = pm.HalfNormal('sigma', sigma=1)
# Expected color
expected_color = alpha + beta * np.log10(periods)
# Standard deviation for the likelihood
sigma_total = pm.math.sqrt(sigma**2 + color_err**2)
# Likelihood
pm.Normal('obs', mu=expected_color, sigma=sigma_total, observed=gbp_minus_grp)
# Sampling
trace = pm.sample(2000, return_inferencedata=True)
```

2.23 Full Gaia RR Lyrae Catalog Download

We then create an ADQL query to download the complete Gaia RR Lyrae catalog, which includes stars with imprecise parallaxes and stars located at low Galactic latitudes, and then cross-matched it with the Gaia source catalog. The query selects various parameters such as mean magnitudes, Galactic coordinates, flux over error, metallicity, and absorption, and the results are obtained asynchronously and printed for inspection.

```

query3 = """
SELECT
    rrlyrae.source_id,
    source.source_id,
    source.phot_bp_mean_mag,
    source.phot_rp_mean_mag,
    source.l,
    source.b,
    source.phot_g_mean_flux_over_error,
    source.phot_bp_rp_excess_factor,
    source.bp_rp,
    source.mh_gspphot,
    rrlyrae.metalllicity,
    rrlyrae.G_absorption,
    rrlyrae.pf
FROM
    gaiadr3.vari_rrlyrae AS rrlyrae
JOIN
    gaiadr3.gaia_source AS source
ON
    rrlyrae.source_id = source.source_id
"""

```

2.24 Calculating Color Excess and G-band Extinction

We calculate the color excess, E(GBP-GRP), for all RR Lyrae stars in the Gaia catalog by subtracting the intrinsic color (a model-dependent value) from the observed color (measured from Gaia data). Based on our previously created model, The intrinsic color is modeled as a function of the pulsation period, with a base value of 0.67 and a slope of 0.11 with respect to the logarithm of the period. Using the color excess, we then calculate the G-band extinction, AG, for each star, assuming a fixed ratio RG = 2.0 between AG and the color excess.

$$R_G \equiv \frac{A_G}{E(\text{GBP} - \text{GRP})} = 2.0 \quad (4)$$

We add the calculated color excess and G-band extinction as new columns to the DataFrame. Finally, we display the updated DataFrame with the source ID, color excess, and G-band extinction for each RR Lyrae star.

```

# Calculating color excess
color_excess = GBP_observed - GRP_observed - GBP_GRP_intrinsic
# Calculating G-band extinction
RG = 2.0
AG = RG * color_excess

```

2.25 Plotting 2-D Map of Color Excess

We then are plotting a two-dimensional map of the color excess E(GBP - GRP) as a function of Galactic longitude (l) and latitude (b) for RR Lyrae stars. We use an Aitoff projection, which is a common way to display data on a spherical surface, such as the sky, in a two-dimensional plot. The Galactic longitude values are adjusted to be between -180 and 180 degrees and then converted to radians, along with the latitude values, for the Aitoff projection. Each RR Lyrae star is represented as a semi-transparent point on the plot, colored according to its AG, with a color bar indicating the scale of the color excess.

2.26 Cleaning Reddening Map

Next we first apply quality cuts based on the signal-to-noise ratio (SNR) and BP/RP excess factor to remove objects with potentially incorrect reddening values based on Brown, A.G.A., Vallenari,

A., Prusti, T., de Bruijne, J.H.J., et al. (2018). We then filter the results to only include objects with color excess values between 0 and 2.5 since they seem to average there. After applying these cuts, we create a DataFrame to store the color excess and Galactic coordinates of the remaining objects. Finally, we plot these objects on an Aitoff projection map, coloring each point by its color excess value to visualize the large-scale structure and distribution of RR Lyrae stars in the Milky Way.

```
SNR_cut = 10 # SNR cut
excess_factor_cut = 1.3 + 0.06 * results3['bp_rp']**2 # cut based on BP/RP excess
# Applying the cuts
quality_mask = (results3['phot_g_mean_flux_over_error'] > SNR_cut) & \
                (results3['phot_bp_rp_excess_factor'] < excess_factor_cut)
# Further Cuts
filtered_df = df1[(df1['ebv'] >= 0) & (df1['ebv'] <= 2.5)]
```

2.27 Comparison with SFD Dust Map

Finally we compare our attenuation map derived from RR Lyrae stars to the widely used SFD dust map created by Schlegel, Finkbeiner, and Davis in 1998. We use the dustmaps Python package to query the SFD map, which provides the optical reddening values, E(B - V), for given coordinates. We create a DataFrame to store these values along with the corresponding Galactic coordinates, filtering out any extreme values. Using the matplotlib library, we then plot these reddening values on an Aitoff projection to visualize the distribution of dust in the Milky Way as seen from our vantage point. The resulting map helps us understand how the dust distribution affects the apparent brightness and color of celestial objects, such as RR Lyrae stars, and allows us to compare our findings with the established SFD map.

```
import dustmaps.sfd
dustmaps.sfd.fetch()
from dustmaps.sfd import SFDQuery
from astropy.coordinates import SkyCoord
import astropy.units as u
coords = SkyCoord(l_filtered, b_filtered, unit = u.rad, frame='galactic')
# Query SFD dust map
sfd = SFDQuery()
ebv_sfd = sfd(coords)
# Create DataFrame with color excess and Galactic coordinates
# Then we plot the SFD map
```

3 Results

solution_id	source_id	...	pf_error	num_clean_epochs_g
		...	d	
375316653866487565	245002531050576896	...	5.9112863e-06	45
375316653866487565	245504251951140864	...	3.948115e-06	62
375316653866487565	245823861938360064	...	3.0187648e-06	59
375316653866487565	246756973652292992	...	2.743857e-06	55
375316653866487565	359112665277840512	...	2.8541435e-06	51
375316653866487565	414003442535159936	...	3.1032864e-06	82
375316653866487565	414081061185873792	...	3.1364334e-06	82
375316653866487565	414152842984521856	...	1.611218e-05	69
375316653866487565	414254032403091072	...	4.310351e-06	62
375316653866487565	415000566440590976	...	3.5996763e-06	54

Figure 2: RR Lyrae Catalog Download

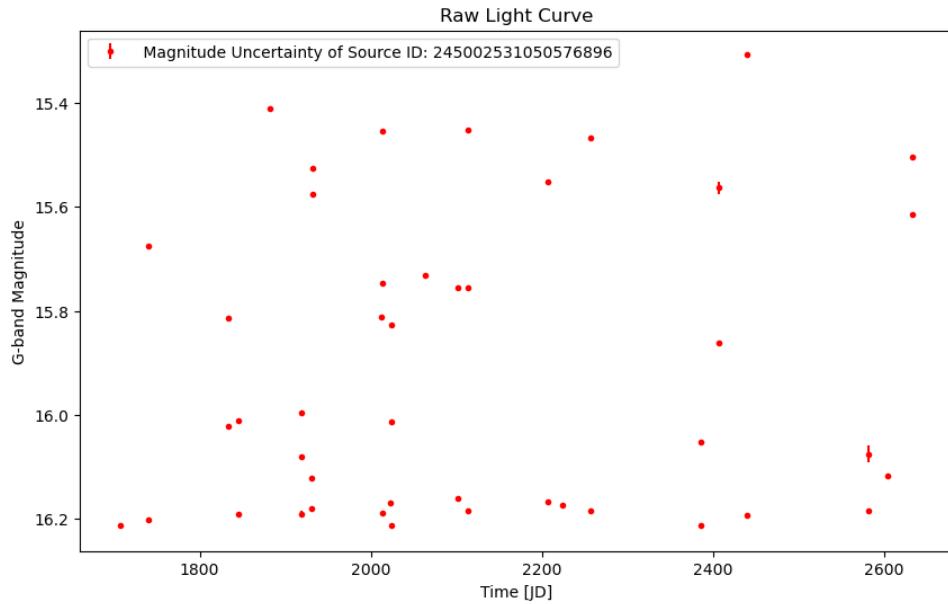


Figure 3: Light Curve Retrieval

source_id	period	mean_mag
245002531050576896	0.520523521474987	15.18301067221589
245504251951140864	0.507122814036343	17.281735936946724
245823861938360064	0.47878235225173055	18.63596097410463
246756973652292992	0.4892813604125664	11.62507330364803
359112665277840512	0.5017813817486655	15.226085501771326
414003442535159936	0.4748484168380481	17.599867930011502
414081061185873792	0.5889178839814972	17.90363990807587
414152842984521856	0.31270764648903143	19.36335646260194
414254032403091072	0.4181262880713784	15.84968218388812
415000566440590976	0.6173559783178713	14.094800208586326
...
2195316153319624192	0.5484572475982226	14.879922034409276
2195648132814029312	0.5014840484513697	18.370538513056275
2195750146876812032	0.6206873153169392	18.53278568050199
2195878209919982464	0.5666407002179531	16.94714496142207
2195880752540525952	0.749057367144842	15.879329642574119
2195937888486787840	0.4863057085026562	17.589360450079496
2195958504330260224	0.4954112054328647	18.130907126584233
2195979635570512640	0.5278730757333602	17.90201157020354
2196012968813801984	0.6058516037470081	18.71864889825055
2196067703877567232	0.5756964704182067	16.00124030217345
2196103884681824384	0.5288970152745514	18.77267077508623

Length = 100 rows

Figure 4: Period and Mean Magnitude Estimation 1

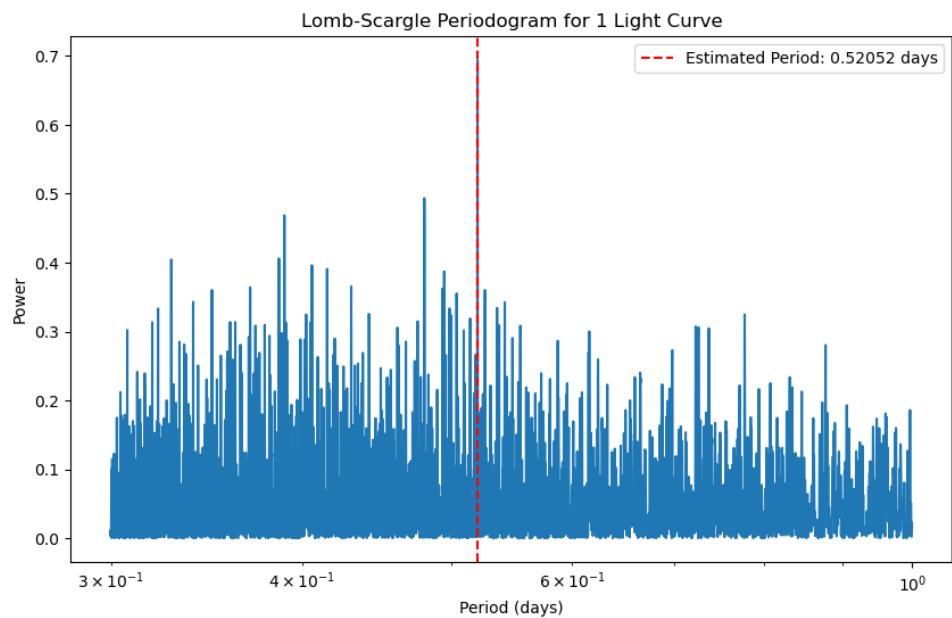


Figure 5: Period and Mean Magnitude Estimation 2 : Estimated Period: 0.520523521474987 days
Estimated Mean G-band Magnitude: 15.870377539271471

source_id	pf d	int_average_g mag
245002531050576896	0.5205123049162365	15.929145
245504251951140864	0.5071152853230051	18.00016
245823861938360064	0.47876357277336673	19.339872
246756973652292992	0.4893027405323637	12.362016
359112665277840512	0.5017802473708718	15.8938055
414003442535159936	0.474867501878603	18.356619
414081061185873792	0.5889203707366553	18.591648
414152842984521856	0.36438054616789506	20.067835
414254032403091072	0.4181043701719148	16.543129
415000566440590976	0.617382983500351	14.803384
...
2195648132814029312	0.5014868734816325	19.063015
2195750146876812032	0.6206472688259573	19.209133
2195878209919982464	0.5666440156365732	17.61679
2195880752540525952	0.7490394860439972	16.554865
2195937888486787840	0.4862684163875203	18.359108
2195958504330260224	0.4954183688956163	18.813147
2195979635570512640	0.5278834017524597	18.581533
2196012968813801984	0.6058409525264999	19.423979
2196067703877567232	0.5756898931290219	16.69619
2196103884681824384	0.5289135029029828	19.488836
Length = 100 rows		
source_id	period	mean_mag
245002531050576896	0.520523521474987	15.1830106722
245504251951140864	0.507122814036343	17.28173593694
245823861938360064	0.47878235225173055	18.6359609741
246756973652292992	0.4892813604125664	11.6250733036
359112665277840512	0.5017813817486655	15.22608550177
414003442535159936	0.4748484168380481	17.59986793001
414081061185873792	0.5889178839814972	17.9036399080
414152842984521856	0.31270764648903143	19.3633564626
414254032403091072	0.4181262880713784	15.8496821838
415000566440590976	0.6173559783178713	14.09480020858
...
2195316153319624192	0.5484572475982226	14.87992203440
2195648132814029312	0.5014840484513697	18.37053851305
2195750146876812032	0.6206873153169392	18.5327856805
2195878209919982464	0.5666407002179531	16.9471449614
2195880752540525952	0.749057367144842	15.87932964257
2195937888486787840	0.4863057085026562	17.58936045007
2195958504330260224	0.4954112054328647	18.13090712658
2195979635570512640	0.5278730757333602	17.9020115702

Figure 6: Period Comparison

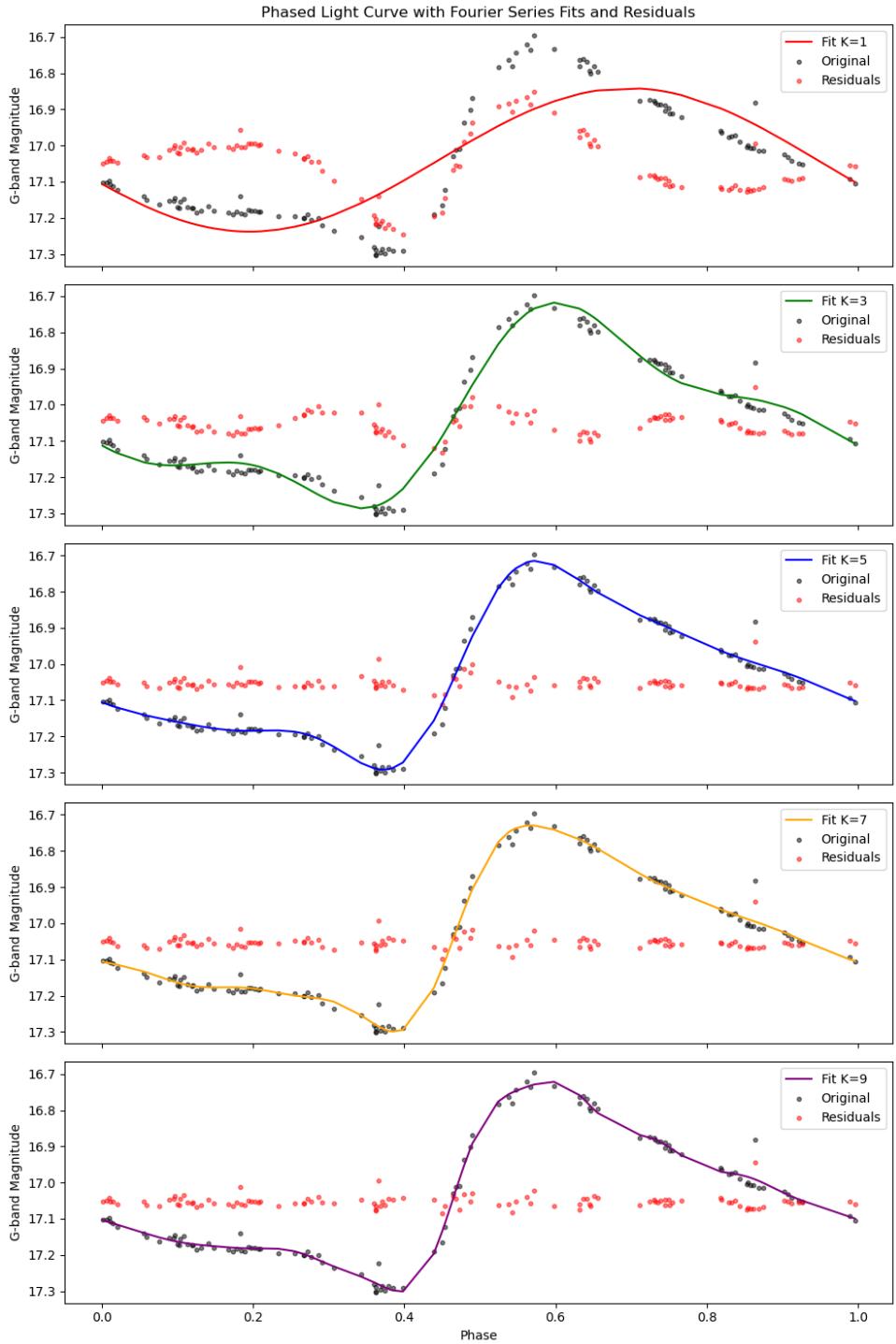


Figure 7: Series Representation for a Star. We computed various K's from 1 to 9 and saw how the curves then go from a smooth to more fitted fit

```

K=1: χ²/N_data(train): 0.00030844704191367354, χ²/N_data(val): 0.0017500576377713693
K=2: χ²/N_data(train): 0.00030966141404604715, χ²/N_data(val): 0.002020878043128776
K=3: χ²/N_data(train): 0.00031204027386028864, χ²/N_data(val): 0.0023519453052468216
K=4: χ²/N_data(train): 0.0003165585904739784, χ²/N_data(val): 0.0027881979358014993
K=5: χ²/N_data(train): 0.0003199411556840858, χ²/N_data(val): 0.0037712388532339255
K=6: χ²/N_data(train): 0.0003282592639580081, χ²/N_data(val): 0.004550296731052161
K=7: χ²/N_data(train): 0.0003354673026967185, χ²/N_data(val): 0.005331239228045281
K=8: χ²/N_data(train): 0.0003232548332450435, χ²/N_data(val): 0.007035101623750892
K=9: χ²/N_data(train): 0.00031437786796185074, χ²/N_data(val): 0.010898644202758287
K=10: χ²/N_data(train): 0.0003059198421237377, χ²/N_data(val): 0.019940060247872208
K=11: χ²/N_data(train): 0.00031069519565060454, χ²/N_data(val): inf
K=12: χ²/N_data(train): 0.0003191002618679172, χ²/N_data(val): -0.021603867530926194
K=13: χ²/N_data(train): 0.00031584538010104914, χ²/N_data(val): -0.014238012020497508
K=14: χ²/N_data(train): 0.00032325581939717164, χ²/N_data(val): -0.009932999070779528
K=15: χ²/N_data(train): 0.0003213957156472821, χ²/N_data(val): -0.010216043406190116
K=16: χ²/N_data(train): 0.00033072406154901, χ²/N_data(val): -0.013439929863721781
K=17: χ²/N_data(train): 0.00032977859911146717, χ²/N_data(val): -0.061954699611613755
K=18: χ²/N_data(train): 0.00033360446708674747, χ²/N_data(val): -0.053413333493649594
K=19: χ²/N_data(train): 0.0003433502033513378, χ²/N_data(val): -0.07980191873971561
K=20: χ²/N_data(train): 0.00033768174022795147, χ²/N_data(val): -0.021994615226266997
K=21: χ²/N_data(train): 0.000351567223598762, χ²/N_data(val): -0.045178389591084944

```

Figure 8: Optimal Number of Terms 1, The most optimal term is seen to be near K 10 for when it stops increasing/decreasing

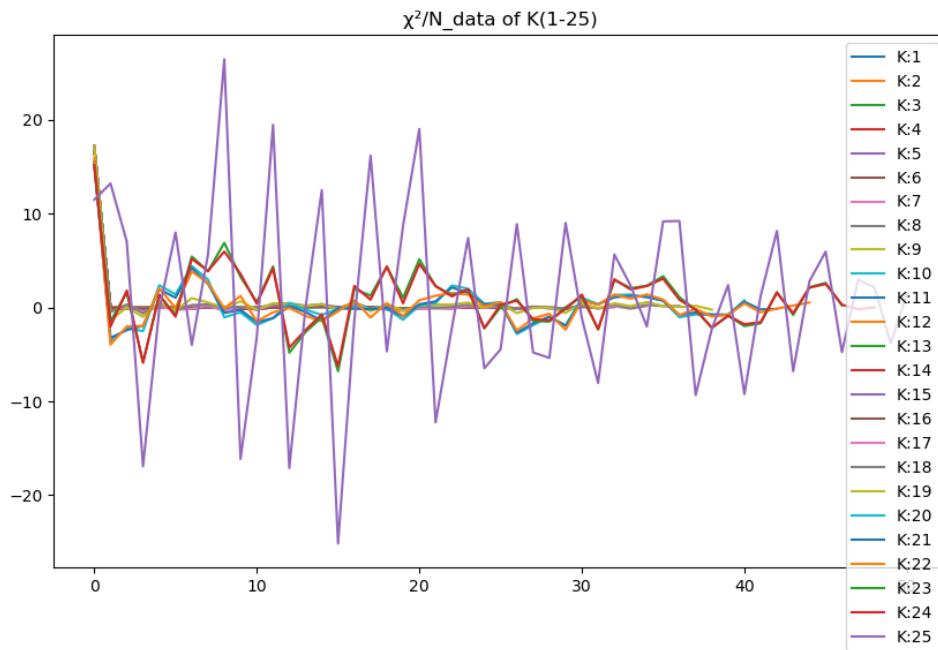


Figure 9: Optimal Number of Terms 2, Optimal K terms visualized

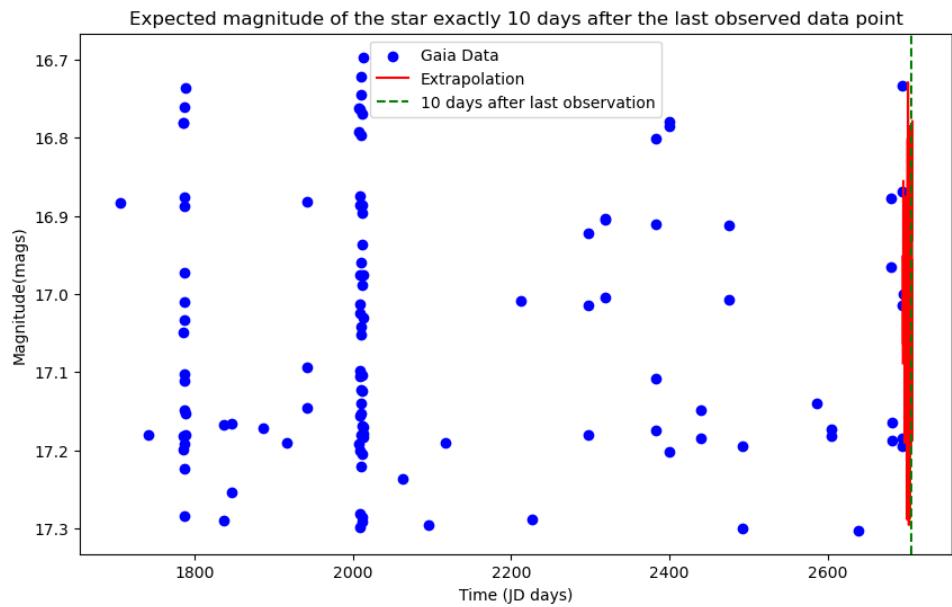


Figure 10: Future Magnitude Prediction 1

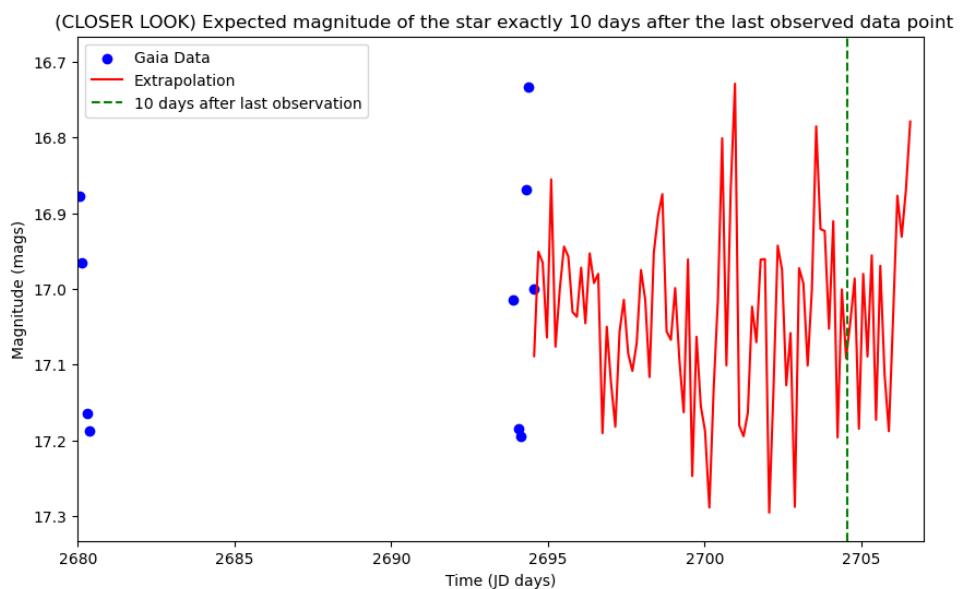


Figure 11: Future Magnitude Prediction 2; Extrapolation shows the how the star's brightness propagates in its next 10/12 days

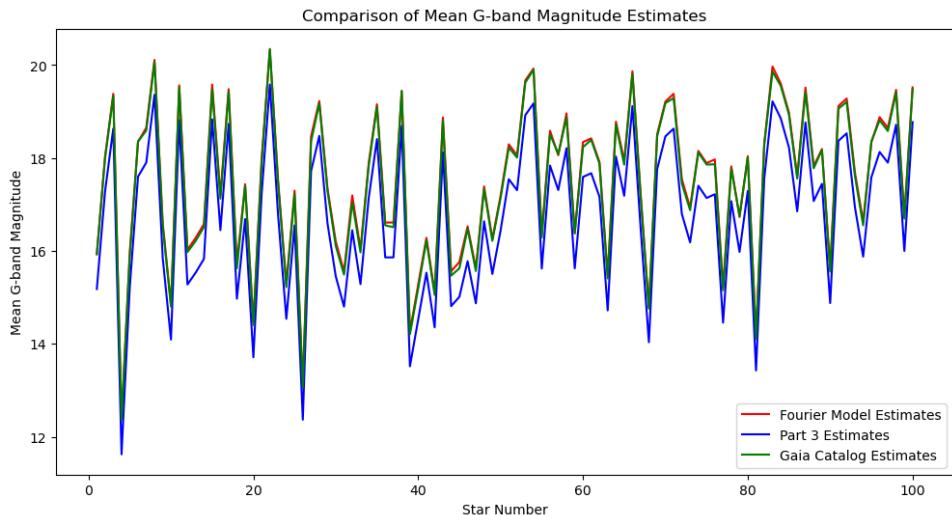


Figure 12: Mean Magnitude Estimation with Fourier Models. A visual comparison between the estimates of fourier, those in part 3 and the Gaia Catalog

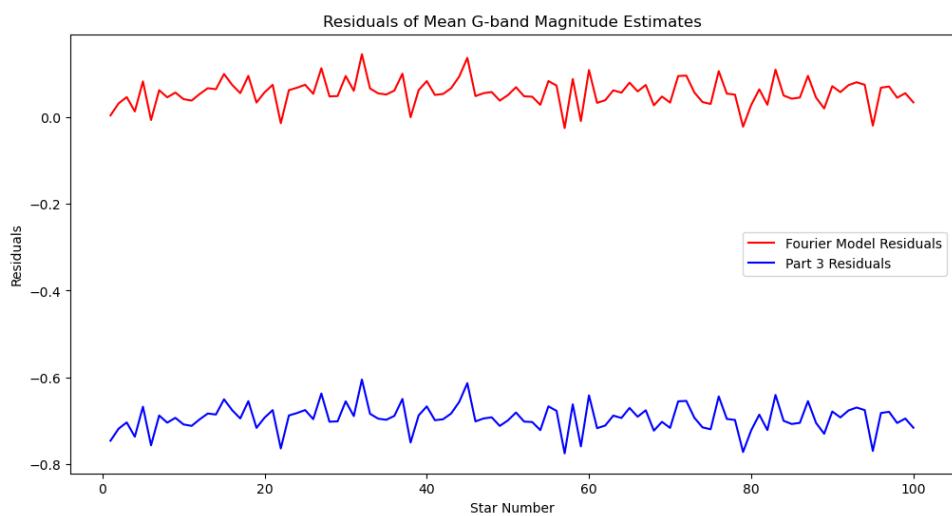


Figure 13: Mean Magnitude Estimation with Fourier Model 2, a visual comparison between the estimates of fourier and those in part 3

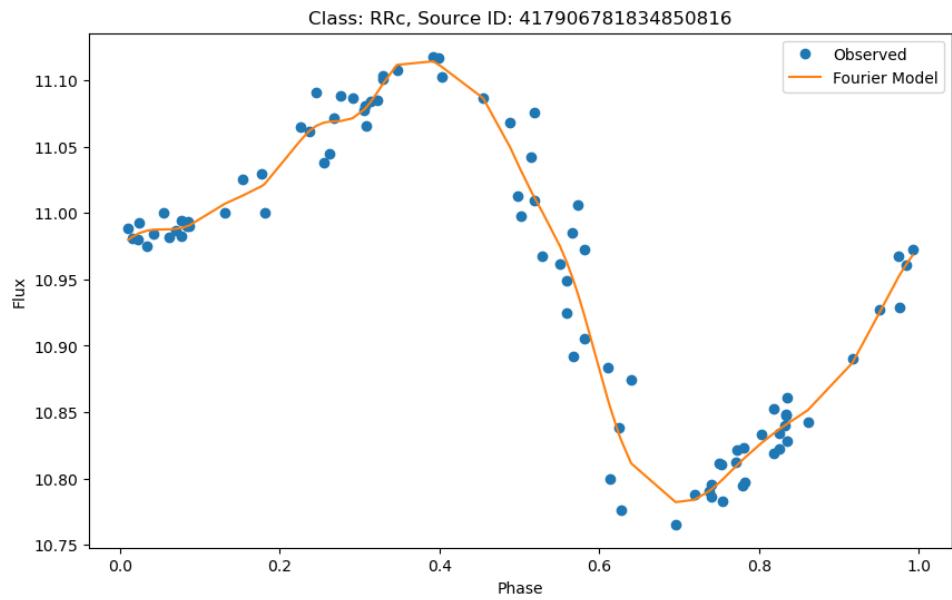


Figure 14: Sample of RRc Light Curve Shape 1

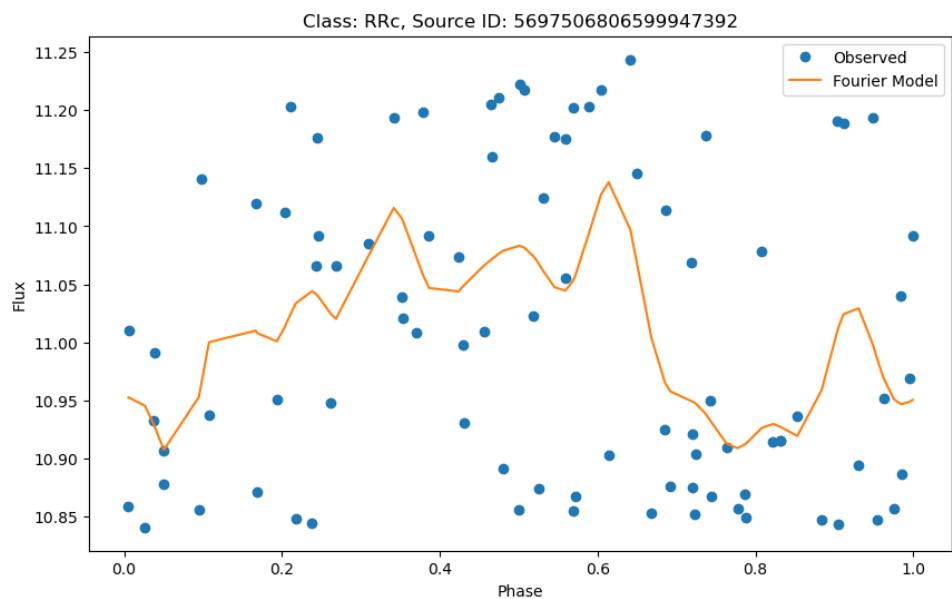


Figure 15: Sample of RRc Light Curve Shape 2

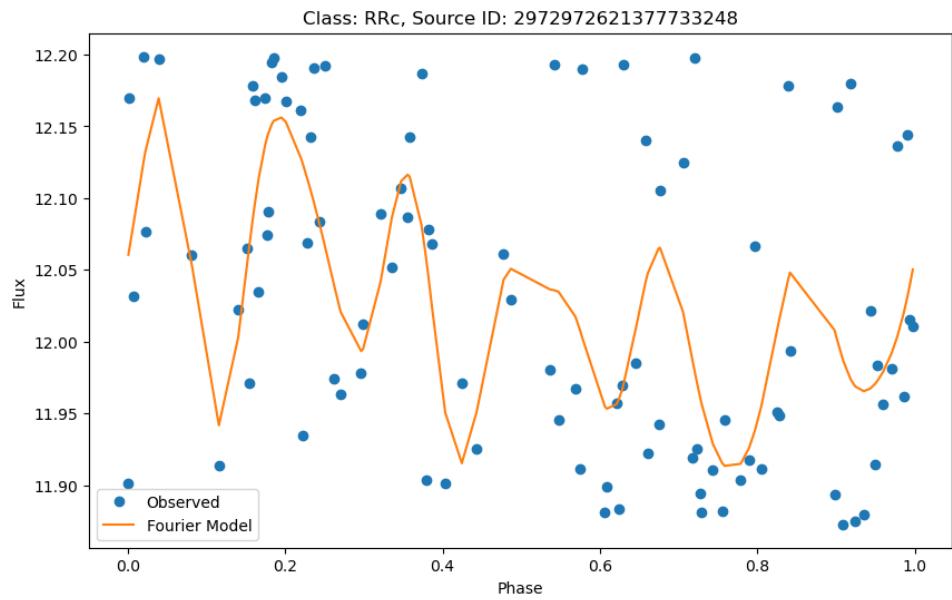


Figure 16: Sample of RRc Light Curve Shape 3

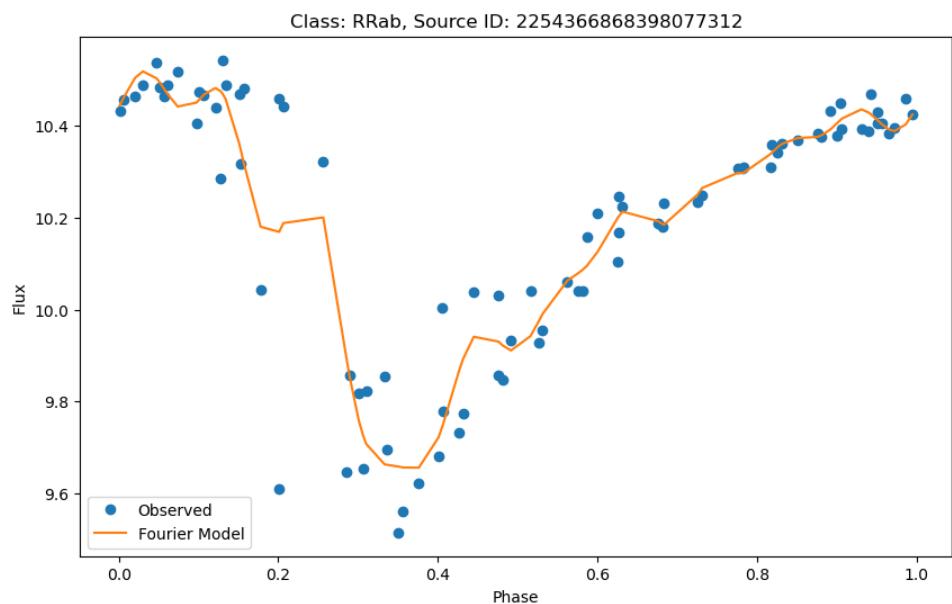


Figure 17: Sample of RRab Light Curve Shape 1

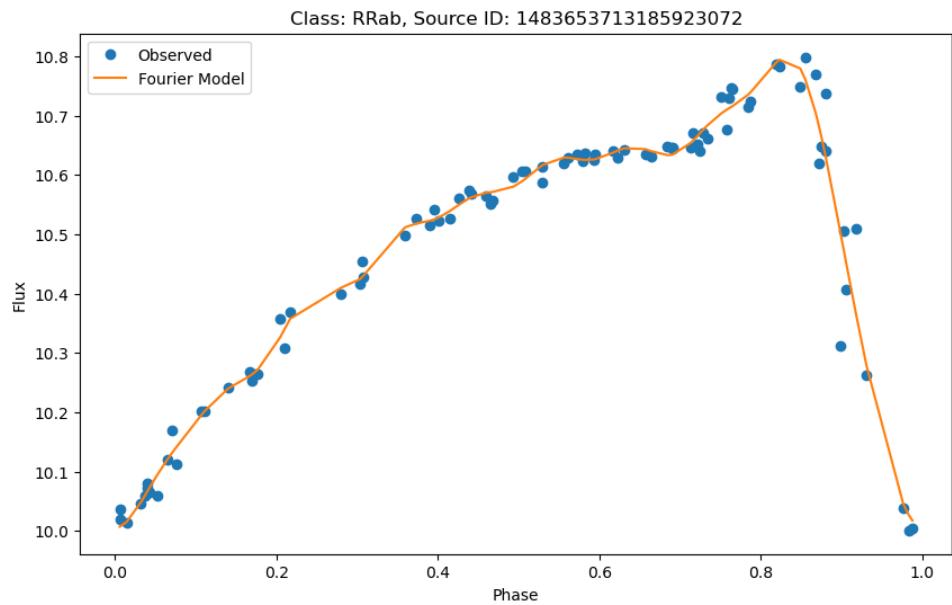


Figure 18: Sample of RRab Light Curve Shape 2

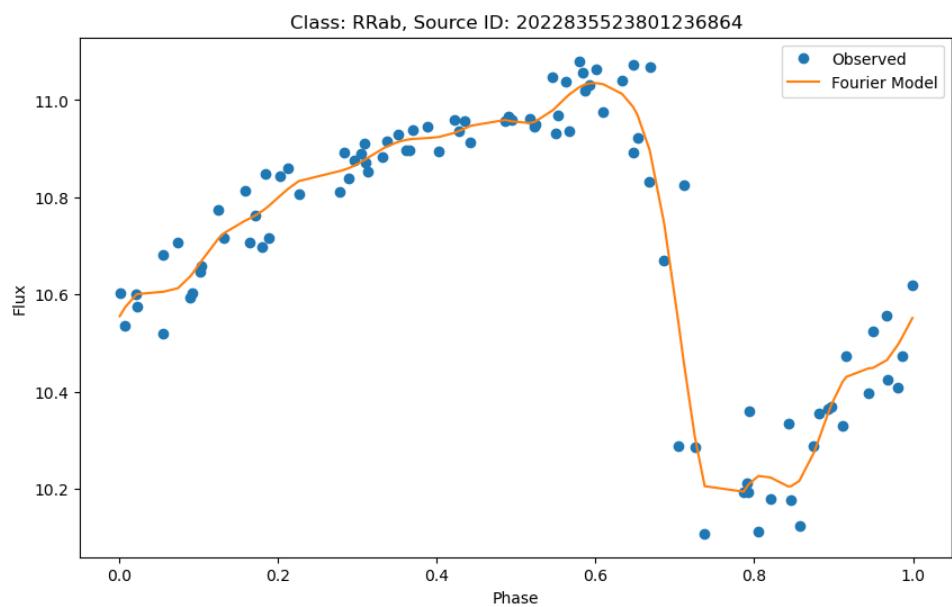


Figure 19: Sample of RRab Light Curve Shape 3

: Table length=665

source_id	ra		dec		parallax			phot_g_mean_mag	phot_bp_mean_mag	phot_rp_mean_mag	b
	deg		deg		mas	mag	mag	mag	mag	mag	deg
	int64	float64	float64	float64	float64	float32	float32	float32	float32	float32	float64
5082275006547809536	55.908656887737756	-25.069824753615666	0.26809523957285347	13.505211	13.75607	13.1515045	-51.47302371332451				
5083946363004134144	57.46167951706597	-23.93054026446269	0.594678130908292	12.15166	12.366504	11.732067	-49.8322488970572				
5094203642556959744	60.80074303875493	-19.853202386844266	0.5017024595959433	12.088892	12.2673	11.662561	-45.693940849004406				
5094426804761045248	60.04487036910923	-19.827044802745775	0.4086511433886336	12.439417	12.668287	12.050987	-46.35806582585982				
5098240155900394112	51.16015378190931	-23.57851381193244	0.36106562106170625	12.535652	12.769855	12.184289	-55.33792695831056				
46568048266412380288	82.3241822609254	-69.36958691558688	1.5453037460475805	19.322914	19.361633	18.642393	-32.46203610681037				
4654699433004977920	73.25966067452313	-71.004186243352655	8.071455328097471	19.629314	19.33239	18.841137	-35.04652054776616				
580348453620559360	136.20240348333073	5.502315372000791	0.5183501987504233	12.576814	12.770765	12.212395	32.04933996526803				
584371601026374272	133.72290623274645	6.436824593962133	0.5023216203946383	12.292535	12.54757	11.882094	30.31802198627761				
...
648476397396238976	145.86990336610856	29.45377819377066	0.49866765663757956	12.073245	12.272	11.670257	48.83708779104214				
3717352885416905984	200.5881071358682	5.88629877641593	0.4726663845615311	12.214666	12.410804	11.764416	67.52057071088163				
3719179562187828992	200.91800299036544	7.212647527259573	0.274366801687668	13.532721	13.810611	13.121407	68.66618260515463				
3727833391597367424	209.69308079142908	12.951586359270129	0.7722731001961485	10.996892	11.218989	10.577158	68.80878454163745				
373172390075245696	200.04819969340105	9.187725579160595	0.5577167596167709	11.764955	11.97401	11.333155	70.81643285109054				
3737667393532367616	197.18468166756443	13.40229367319402	0.3294025279628000	13.240041	13.431516	12.904701	75.69081408986534				
6255453672871215616	232.82829156406103	-17.5155156082623	0.2997836465215604	13.552397	13.767755	13.159515	30.87872607144671				
6257531784202384000	227.4627859037128	-18.198236722138173	0.2943880762497345	13.518623	13.730894	13.130404	33.60780617769447				
1183706283996839680	225.3872935033853	14.701559336649462	0.328909284582799	12.529523	12.754896	12.1284075	57.30213904325092				
1185942999884896640	219.59070157774215	14.415183329929095	0.28819236877379956	13.107498	13.362993	12.73848	62.05070129278371				

Figure 20: RR Lyrae Period-Luminosity Relation with Gaia Distances

best_classification	ra		dec		parallax			phot_g_mean_mag	phot_bp_mean_mag	phot_rp_mean_mag	distance_median	distance_lower	distance_upper
	deg		deg		mas	mag	mag	mag	mag	mag	pc	pc	pc
	object	float64	float64	float64	float64	float32	float32	float32	float32	float32	float32	float32	float32
RRab	55.908656887737756	-25.069824753615666	0.26809523957285347	13.505211	13.75607	13.1515045	3251.9053	3081.9224	3426.7776				
RRab	57.46167951706597	-23.93054026446269	0.594678130908292	12.15166	12.366504	11.732067	1607.3054	1558.1212	1652.5009				
RRab	60.80074303875493	-19.853202386844266	0.5017024595959433	12.088892	12.2673	11.662561	1885.8388	1809.5858	1978.6187				
RRab	60.04487036910923	-19.827044802745775	0.4086511433886336	12.439417	12.668287	12.050987	2297.7349	2213.6853	2390.4333				
RRab	51.16015378190931	-23.57851381193244	0.36106862106170625	12.535652	12.798655	12.184289	2607.2856	2500.5015	2714.123				
RRab	82.3241822609254	-69.36958691558688	1.5453037460475805	19.322914	19.361633	18.642303	691.4829	576.464	817.82166				
RRab	73.25966067452313	-71.004186243352655	8.071455328097471	16.929314	18.841337	140.70844	116.3465	116.3465	178.88231				
RRab	136.20240348333073	5.502315372000791	0.5183501987504233	12.576814	12.770765	12.212395	1787.6908	1656.3534	1955.4971				
RRab	133.72290623274645	6.436824593962133	0.5023216203946383	12.292535	12.54757	11.882094	1884.4073	1837.6722	1942.4014				
...
RRab	145.86990336610856	29.45377819377066	0.49866765663757956	12.073245	12.272	11.670257	1865.8291	1789.4633	1938.1553				
RRab	200.5881071358682	5.88629877641593	0.4726663845615311	12.214666	12.410804	11.764416	2001.7004	1925.4808	2067.964				
RRab	200.91800299036544	7.212647527259573	0.274366801687668	13.532721	13.810611	13.121407	3119.702	2956.674	3275.4878				
RRab	209.69308079142908	12.951586359270129	0.7722731001961485	10.996892	11.218989	10.577158	1232.3232	1200.33	1263.0288				
RRab	200.04819969340105	9.187725579160595	0.5577167596167709	11.764955	11.97401	11.333155	1646.2631	1596.8693	1700.4858				
RRab	197.18468166756443	13.402293687319402	0.3214025279628000	13.240041	13.431516	12.904701	2826.137	2655.234	3014.0408				
RRab	232.82629156406103	-17.5155156082623	0.29297836465215604	13.562397	13.767755	13.159195	2954.5999	2795.866	3155.7043				
RRab	227.4627859037128	-18.198236722138173	0.2943880762497345	13.518623	13.730894	13.130404	2934.6714	2811.0996	3075.7012				
RRab	225.3872935033853	14.701559336649462	0.328909284582799	12.529523	12.754896	12.1284075	2787.287	2676.745	2921.4912				
RRab	219.59070157774215	14.415183329929095	0.28819236877379956	13.107498	13.362993	12.73848	3073.7395	2917.7317	3252.363				

Figure 21: Distance Estimates with Galactic Structure Prior

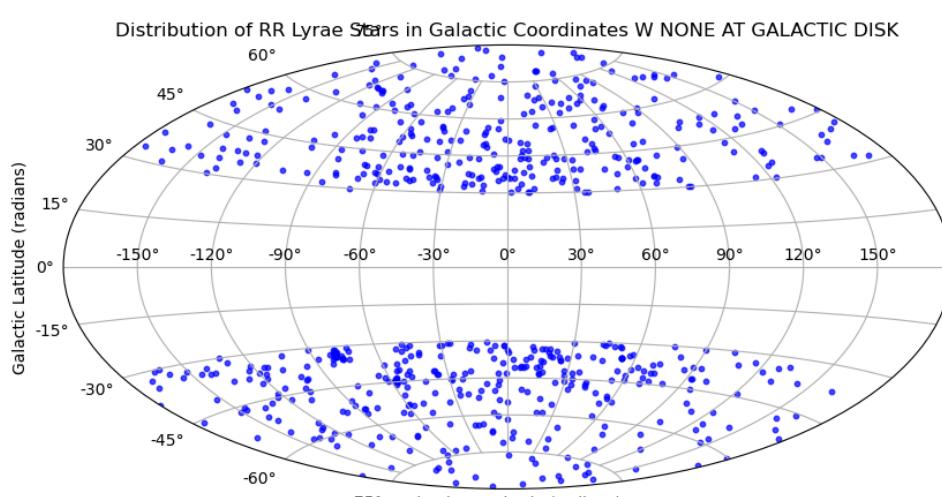


Figure 22: Galactic Distribution of Targets ; Comparison of Naive and Bailer Jones Distance Estimates in parsec

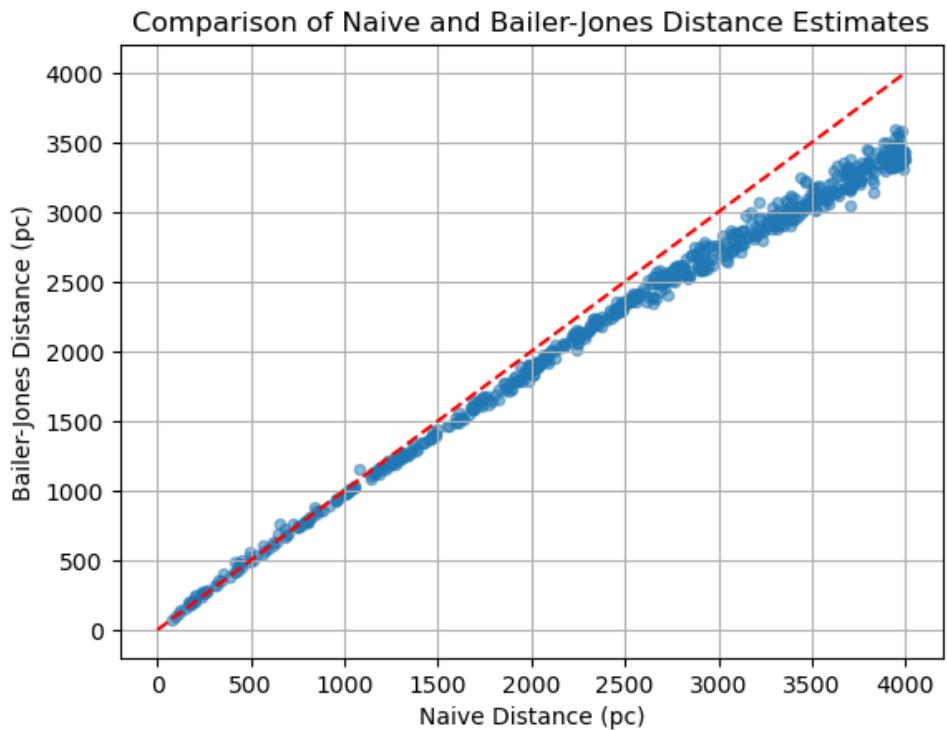


Figure 23: Galactic Distribution of Targets ; Comparison of EDR3 and Bailer Jones Parallaxes in mas

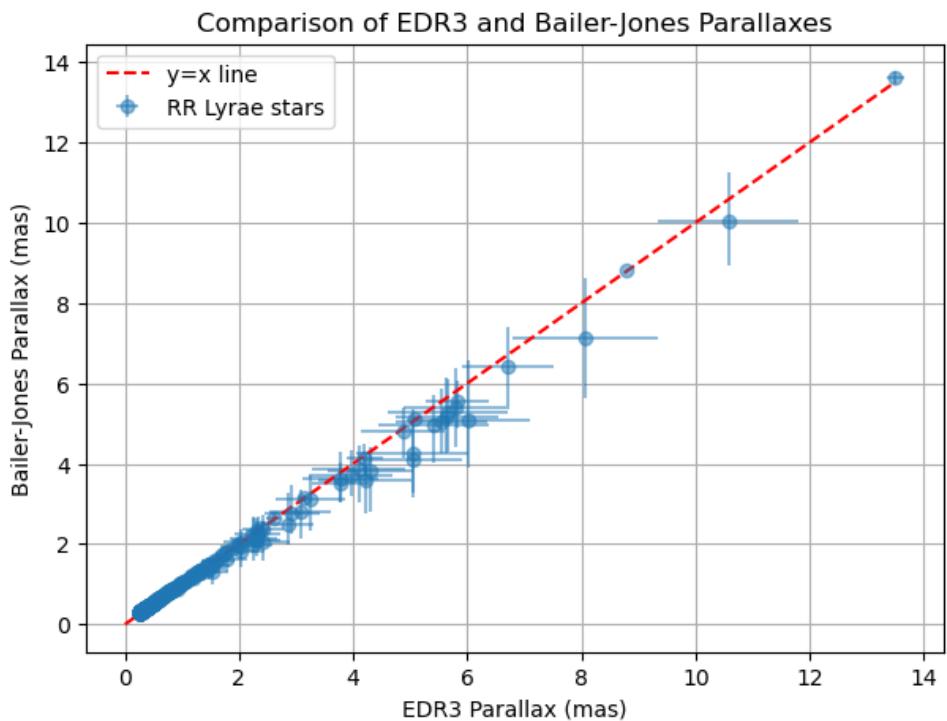


Figure 24: Galactic Distribution of Targets 3: Average difference between EDR3 and Bailer-Jones parallaxes: -0.01 mas

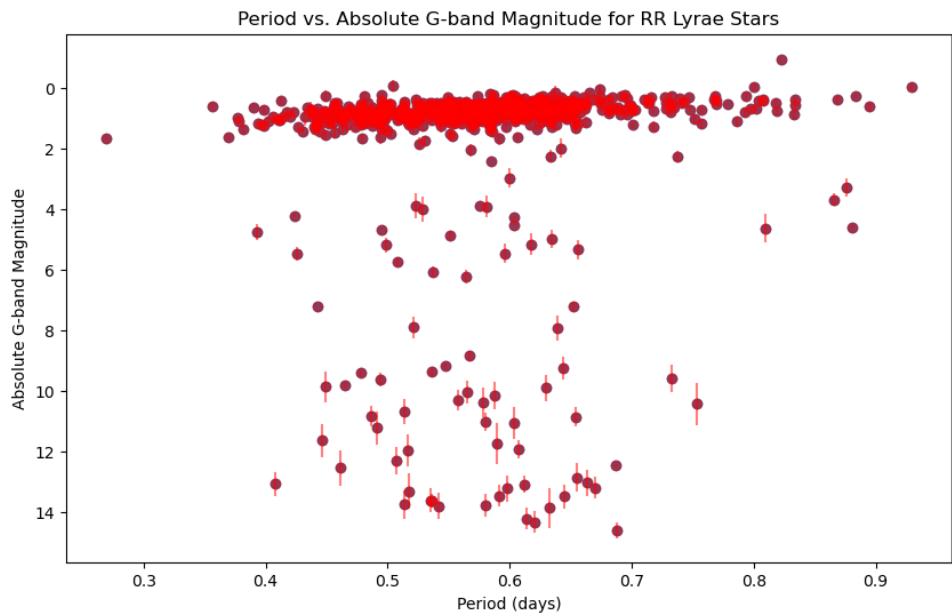


Figure 25: The plot above is a Period (in days) vs. Absolute Magnitude (in mags)

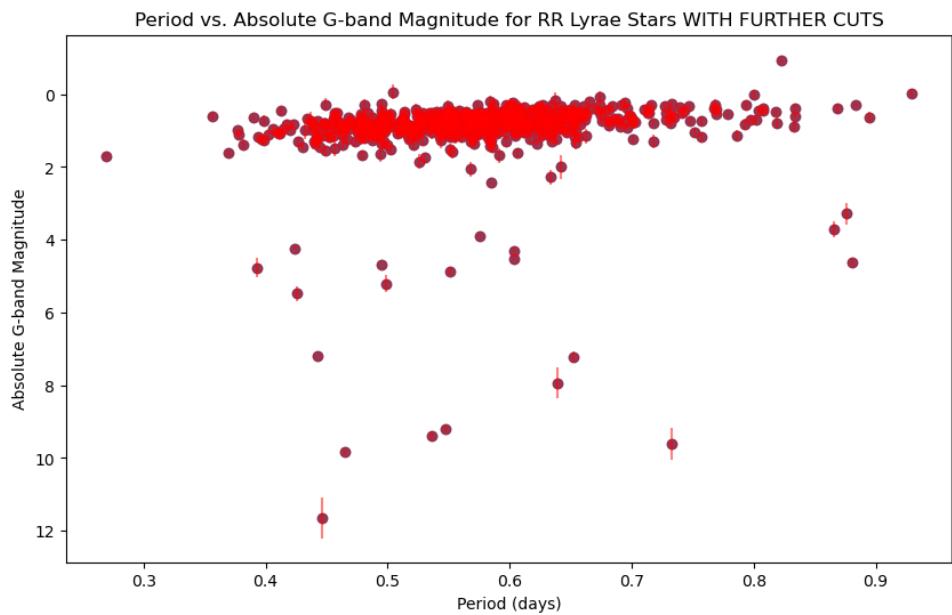


Figure 26: Scatter Reduction with Quality Cuts from figure 25

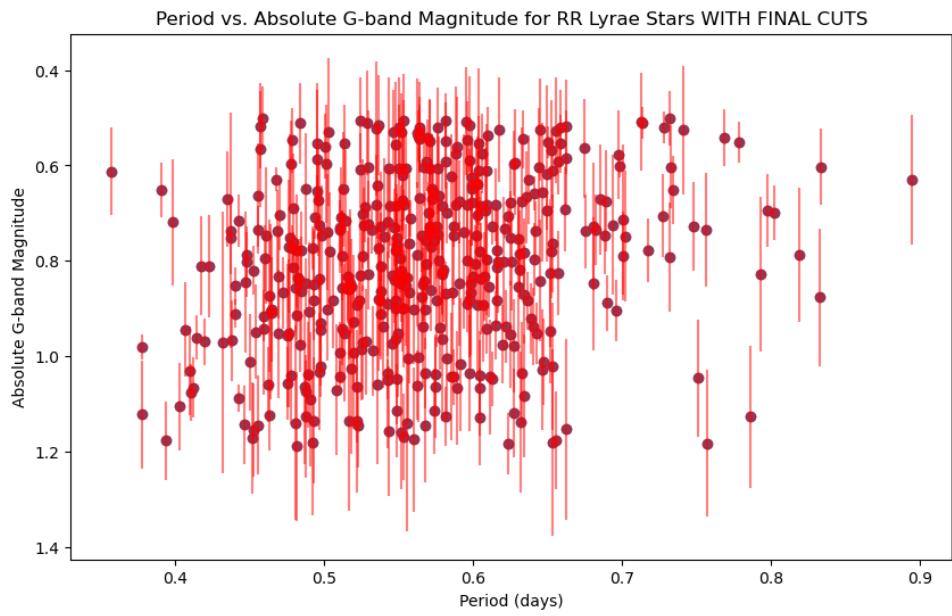


Figure 27: Removing Remaining Outliers from figure 26

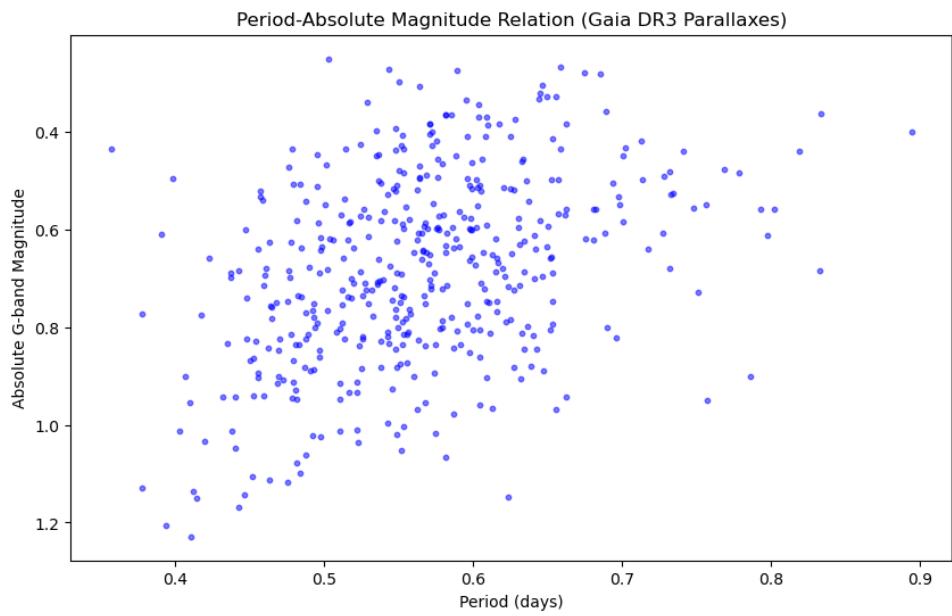


Figure 28: Comparison with Gaia DR3 Parallaxes

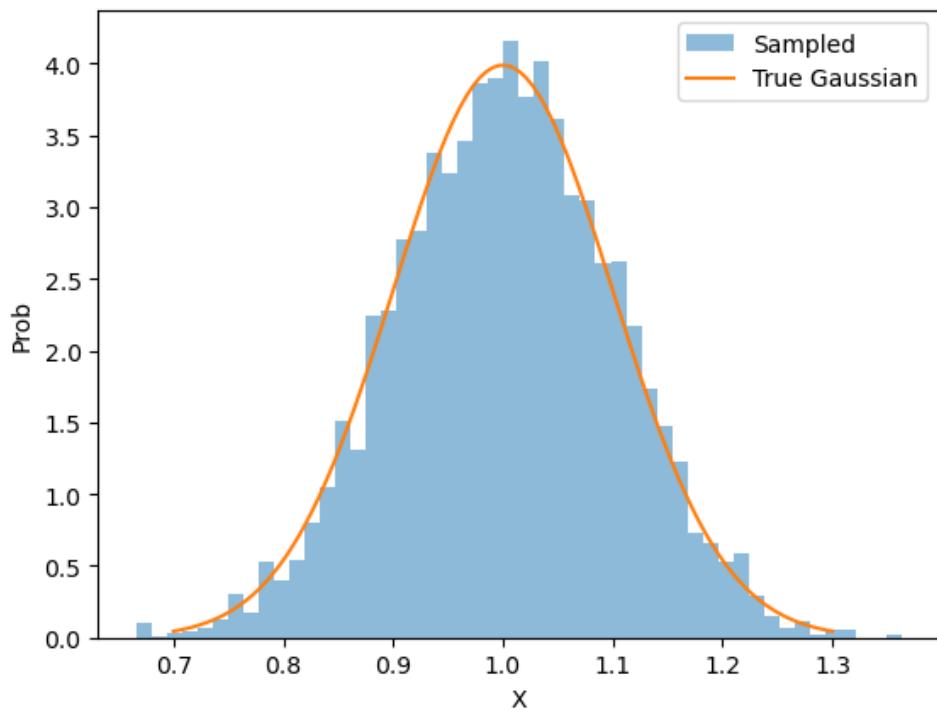


Figure 29: Fitting Period-Luminosity Relation with M-H MCMC (i) 1: Acceptance rate: 0.5141

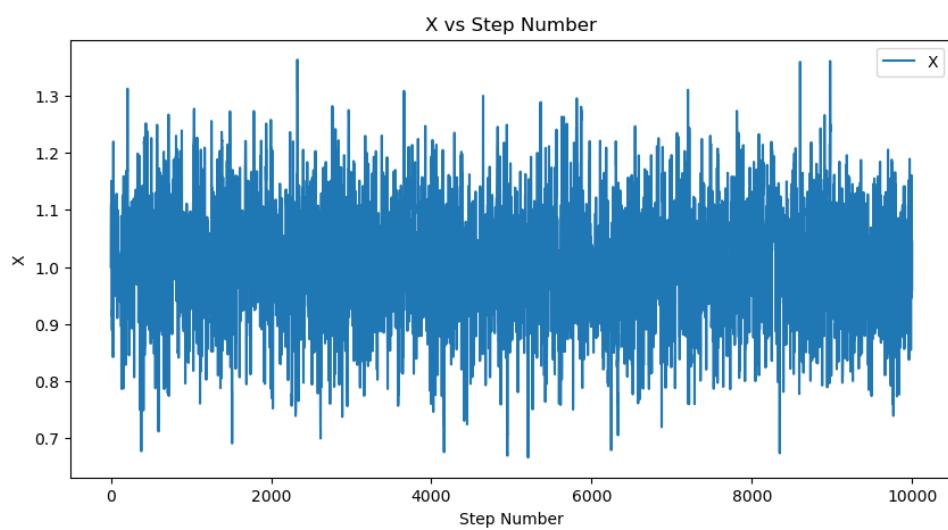


Figure 30: Fitting Period-Luminosity Relation with M-H MCMC; X vs Step Numbers 2

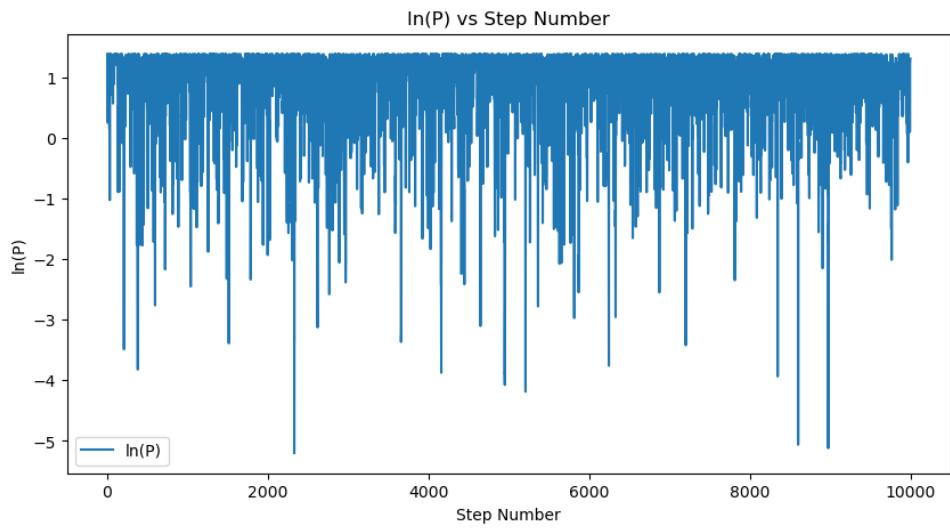


Figure 31: Fitting Period-Luminosity Relation with M-H MCMC with log probability (i) 3

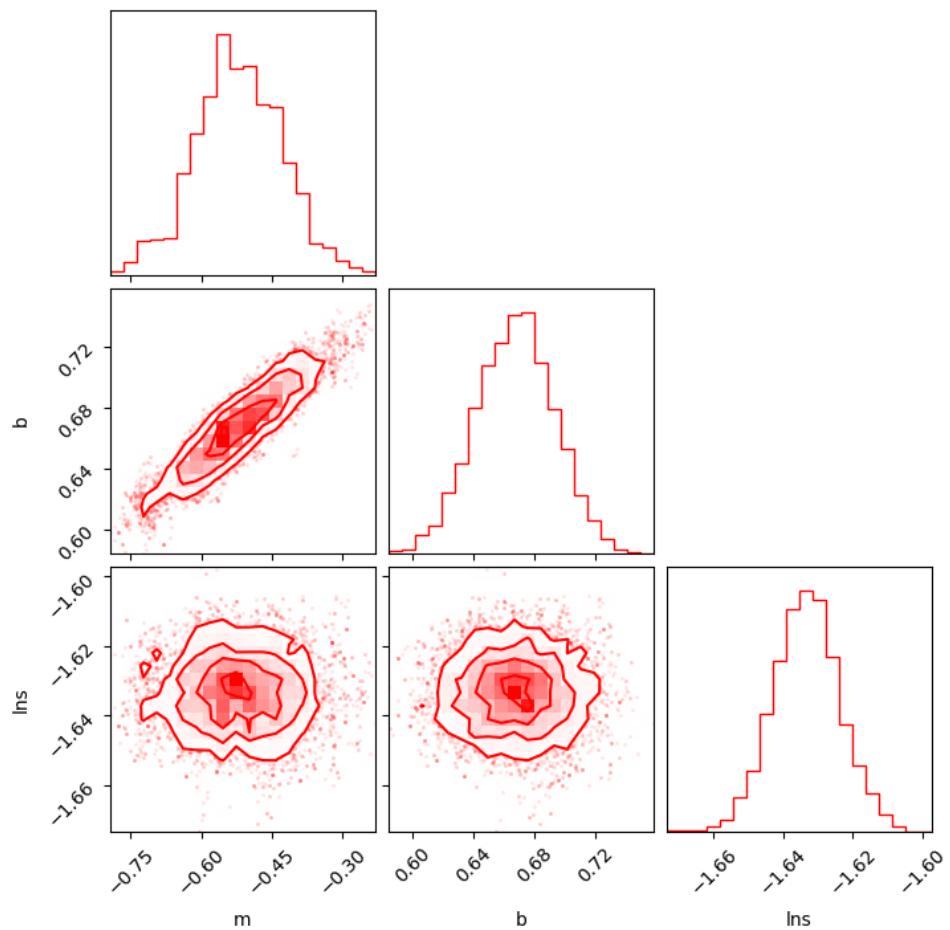


Figure 32: Fitting Period-Luminosity Relation with M-H MCMC, visualizing parameters with corner plots (i) 4: Acceptance rate: 0.5176

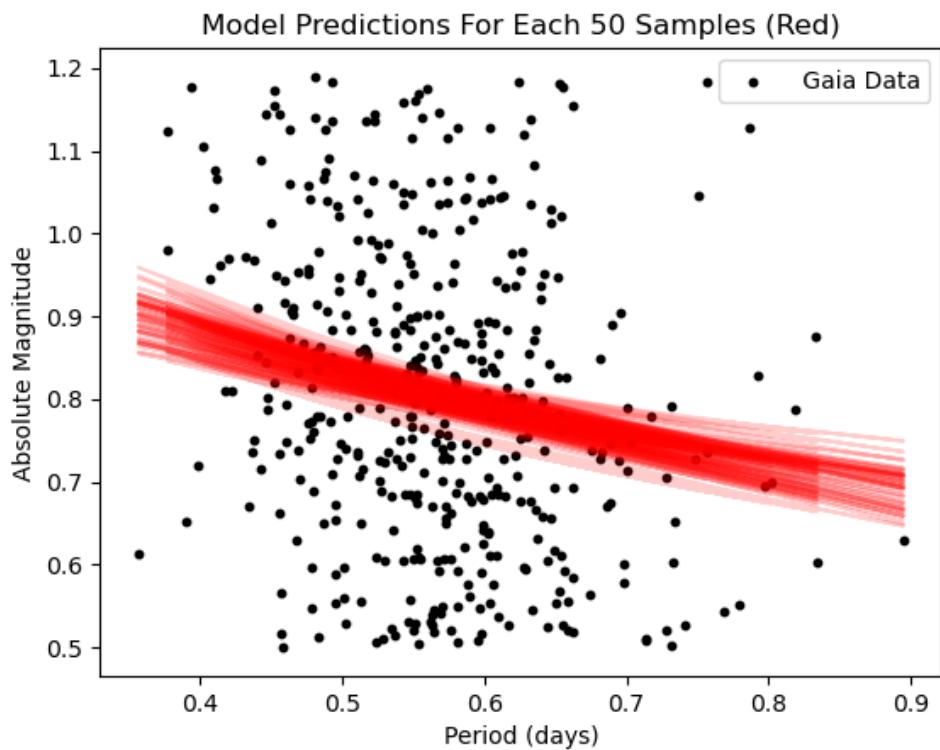


Figure 33: Period-Luminosity Relation with M-H MCMC and its predictions of period (days) vs absolute magnitude (mags) (i) 5

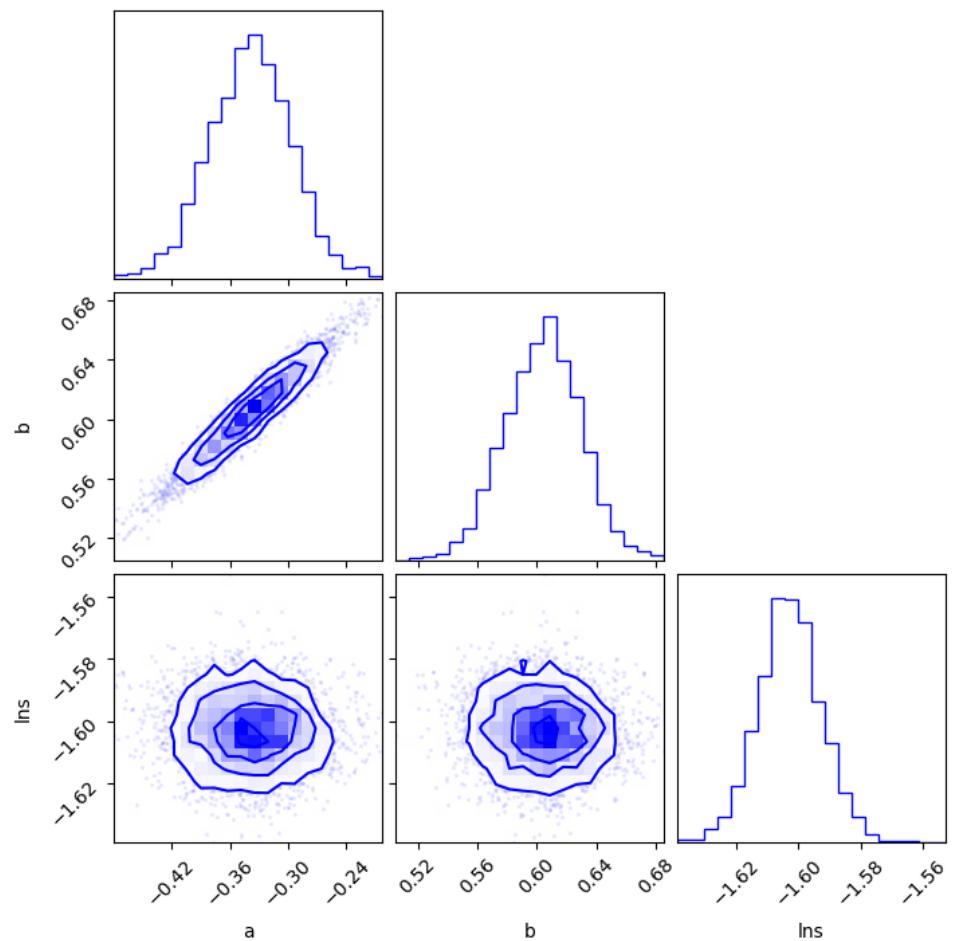


Figure 34: Fitting Period-Luminosity Relation with PYMC Model, visualizing parameters with corner plots (ii) 1

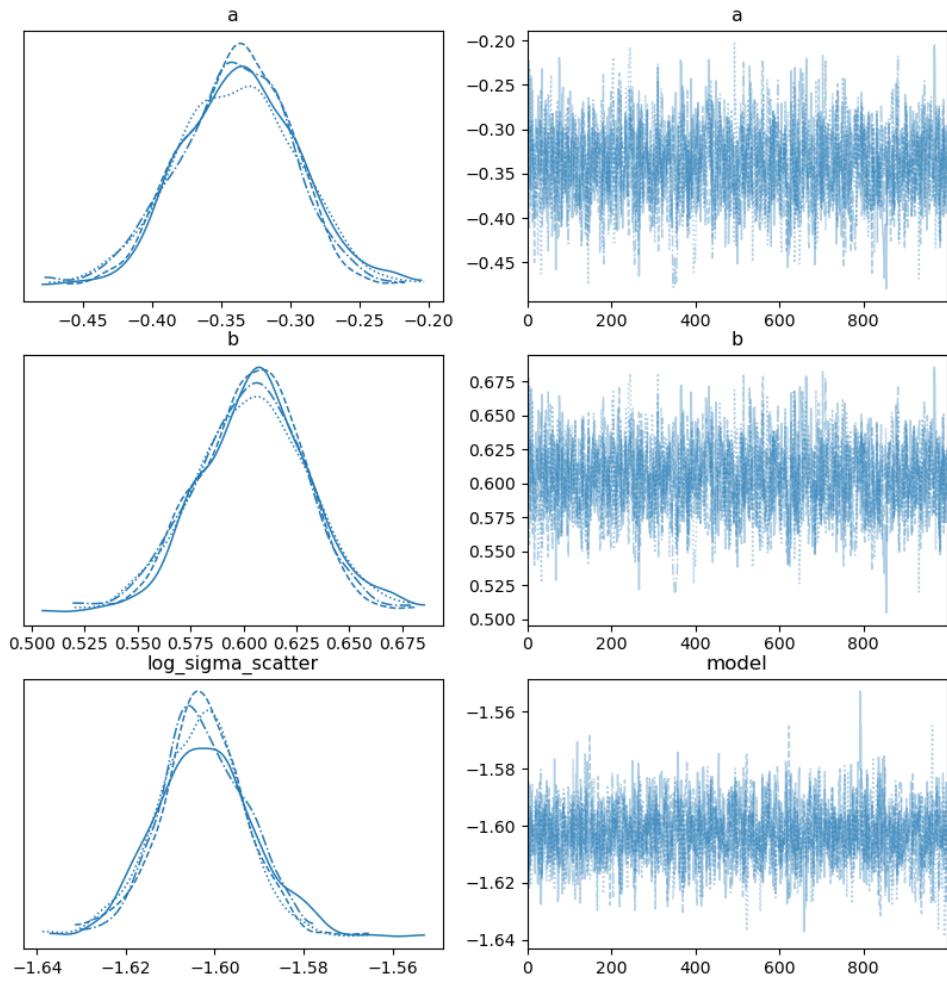


Figure 35: Period-Luminosity Relation with PYMC Model, distribution of parameters (ii) 2

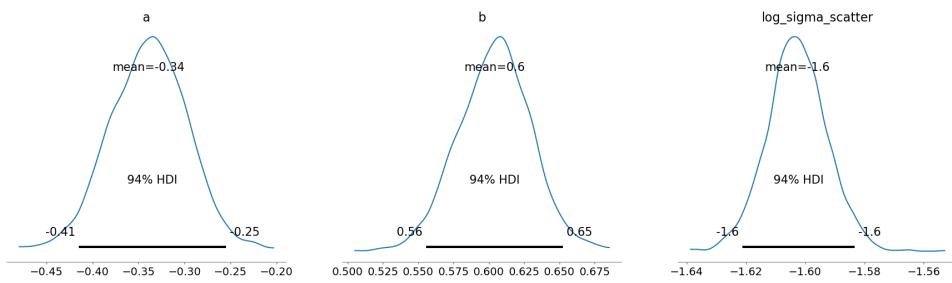


Figure 36: Period-Luminosity Relation with PYMC Model, distribution of parameters (ii) 3

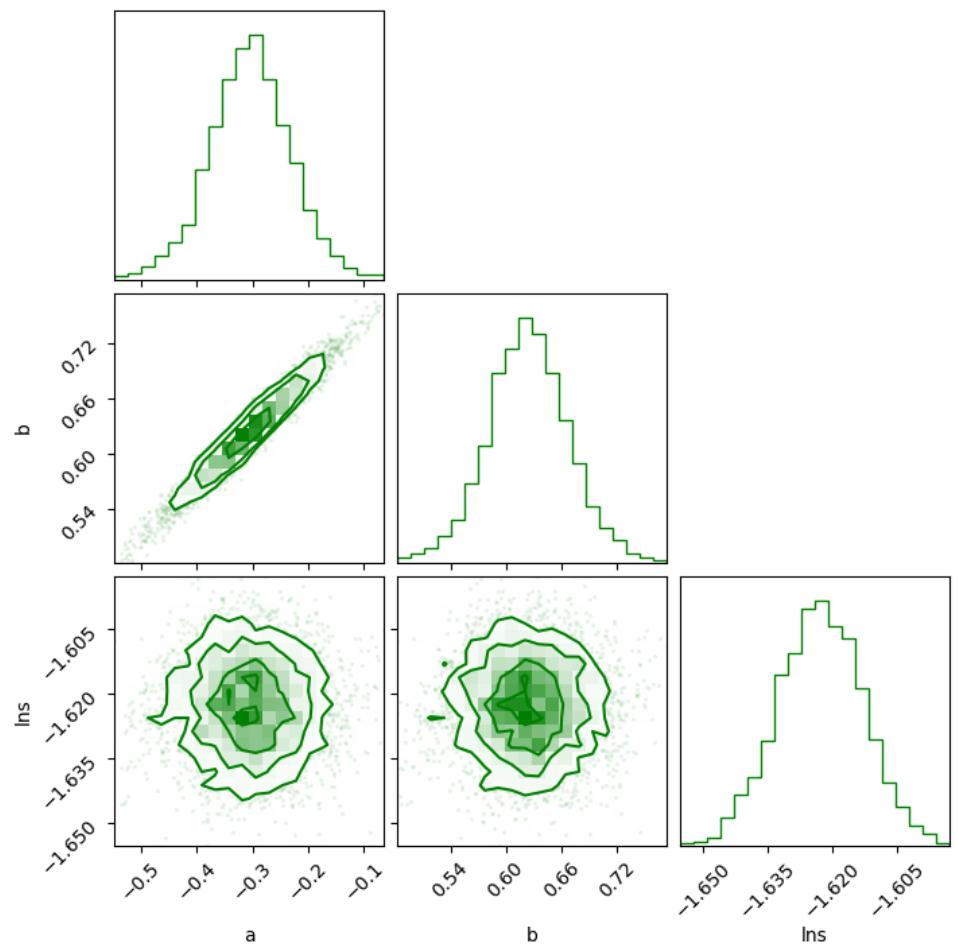


Figure 37: Fitting Period-Luminosity Relation with PYMC Model, visualizing parameters with corner plots (iii) 1

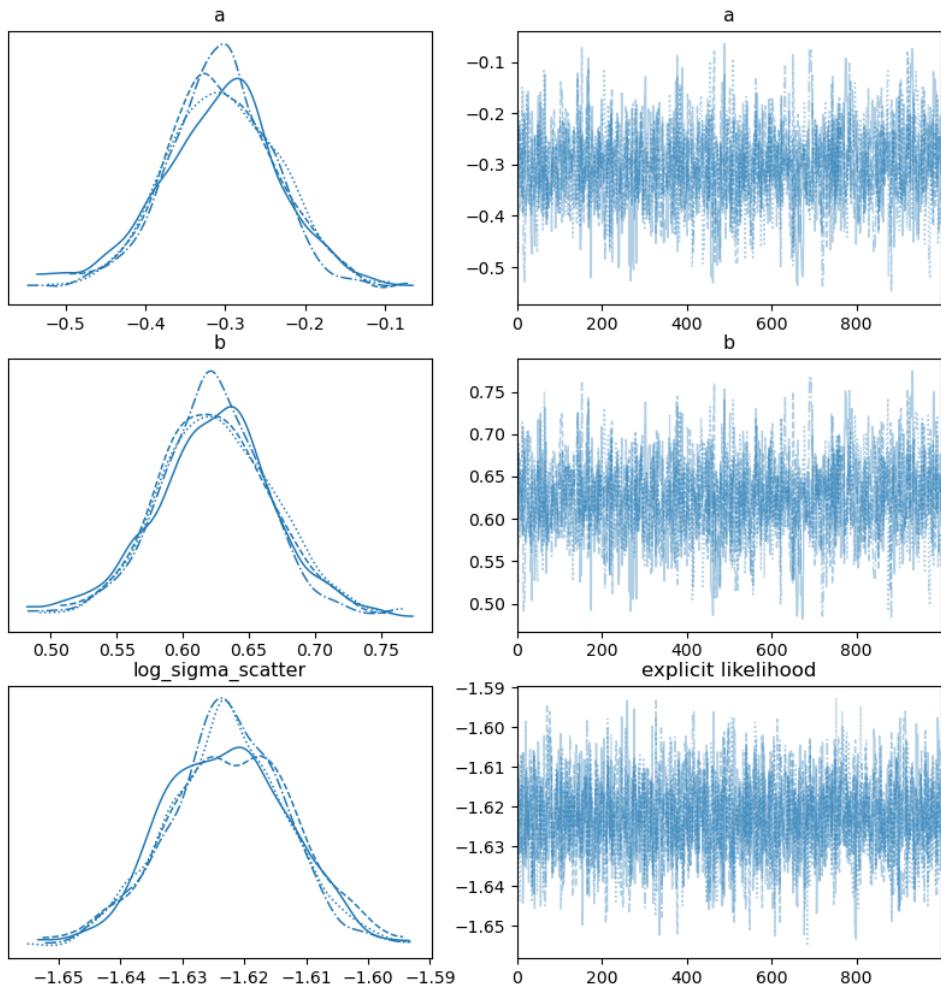


Figure 38: Period-Luminosity Relation with PYMC Model with distribution of parameters (iii) 2

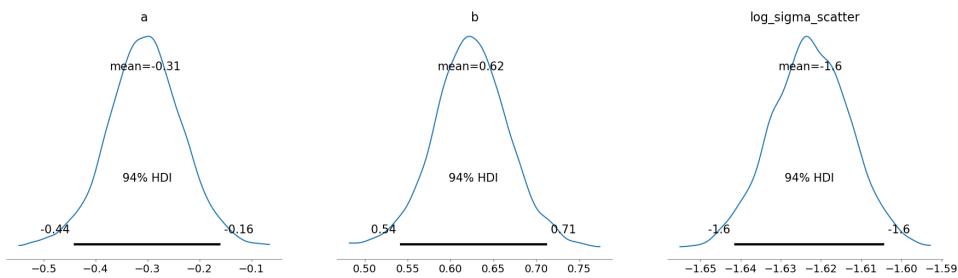


Figure 39: Period-Luminosity Relation with PYMC Model distribution of parameters (iii) 3

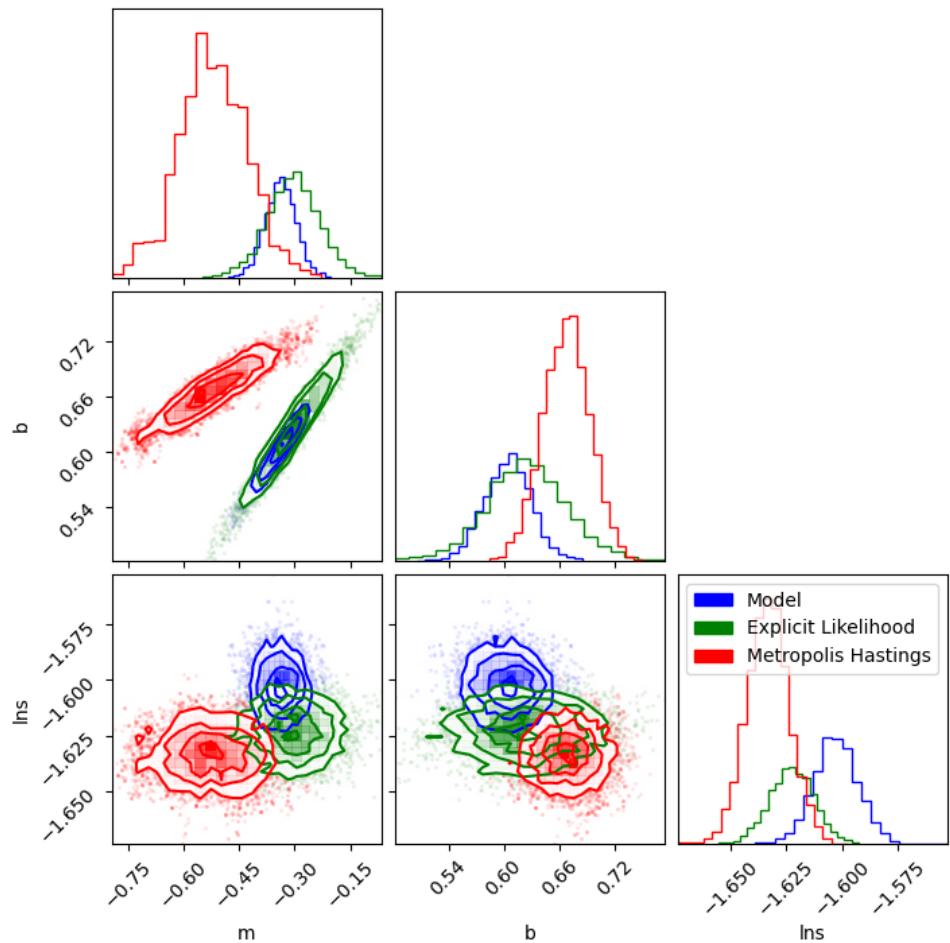


Figure 40: Fitting Period-Luminosity Relation with ALL 3 Models and their visualizations through corner plots(iii)

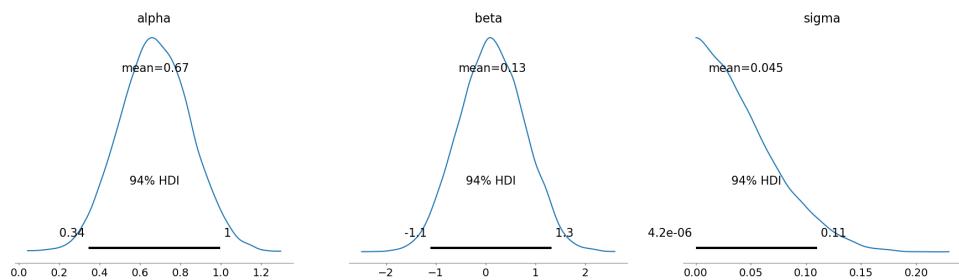


Figure 41: Deriving Period-Color Relation and the distribution of its parameters

source_id	source_id_2	...	G_absorption mag	p f d
244893954275107200	244893954275107200	...	--	--
245002531050576896	245002531050576896	...	1.122552	0.5205123049162365
245251978451433728	245251978451433728	...	--	--
245293416297967360	245293416297967360	...	--	--
245435906133523072	245435906133523072	...	--	--
245464806968150400	245464806968150400	...	--	--
245504251951140864	245504251951140864	...	2.0499203	0.5071152853230051
5712016786666075776	5712016786666075776	...	--	0.5894468159499638
245806617648715008	245806617648715008	...	--	--
245823861938360064	245823861938360064	...	1.2478467	0.47876357277336673
...
5917407173655038336	5917407173655038336	...	--	--
5917418134421212288	5917418134421212288	...	3.9412684	0.6620894476196592
5917421329878084224	5917421329878084224	...	3.8620486	0.34230985883772486
5917421845281955584	5917421845281955584	...	0.6402818	0.5329578734043147
5917430366489806080	5917430366489806080	...	--	0.5769009580921866
5917433080917061376	5917433080917061376	...	--	--
5917434730176001152	5917434730176001152	...	1.3096962	0.49554678618479203
5917438024388028928	5917438024388028928	...	--	--
5917442220595980416	5917442220595980416	...	--	0.6774738129526996
5917450741791830528	5917450741791830528	...	--	--
Length = 271779 rows				

Figure 42: Full Gaia RR Lyrae Catalog Download with needed labels

source_id	color_excess	AG	
244893954275107200	--	--	
245002531050576896	0.5318599466350564	1.0637198932701128	
245251978451433728	--	--	
245293416297967360	--	--	
245435906133523072	--	--	
245464806968150400	--	--	
245504251951140864	0.7664739345162207	1.5329478690324414	
5712016786666075776	1.1960250949787594	2.392050189957519	
245806617648715008	--	--	
245823861938360064	0.6355839213621689	1.2711678427243378	
...	
5917403497163999232	--	--	
5917407173655038336	--	--	
5917418134421212288	0.0297131031004485	0.059426206200897	
5917421329878084224	1.8271662258919763	3.6543324517839526	
5917421845281955584	0.3841134488740654	0.7682268977481308	
5917430366489806080	0.33341474049111364	0.6668294809822273	
5917433080917061376	--	--	
5917434730176001152	0.5874098817723185	1.174819763544637	
5917438024388028928	--	--	
5917442220595980416	--	--	
5917450741791830528	--	--	
Length = 271779 rows			

Figure 43: Calculation of Color Excess and G-band Extinction

source_id	color_excess	AG	G_absorption_mag
244893954275107200	--	--	--
245002531050576896	0.5318599466350564	1.0637198932701128	1.122552
245251978451433728	--	--	--
245293416297967360	--	--	--
245435906133523072	--	--	--
245464806968150400	--	--	--
245504251951140864	0.7664739345162207	1.5329478690324414	2.0499203
571201678666075776	1.1960250949787594	2.392050189957519	--
245806617648715008	--	--	--
245823861938360064	0.6355839213621689	1.2711678427243378	1.2478467
...
5917407173655038336	--	--	--
5917418134421212288	0.0297131031004485	0.059426206200897	3.9412684
5917421329878084224	1.8271662258919763	3.6543324517839526	3.8620486
5917421845281955584	0.3841134488740654	0.7682268977481308	0.6402818
5917430366489806080	0.33341474049111364	0.6668294809822273	--
5917433080917061376	--	--	--
5917434730176001152	0.5874098817723185	1.174819763544637	1.3096962
5917438024388028928	--	--	--
5917442220595980416	--	--	--
5917450741791830528	--	--	--
Length = 271779 rows			

Figure 44: Comparison with Catalog's G_absorption

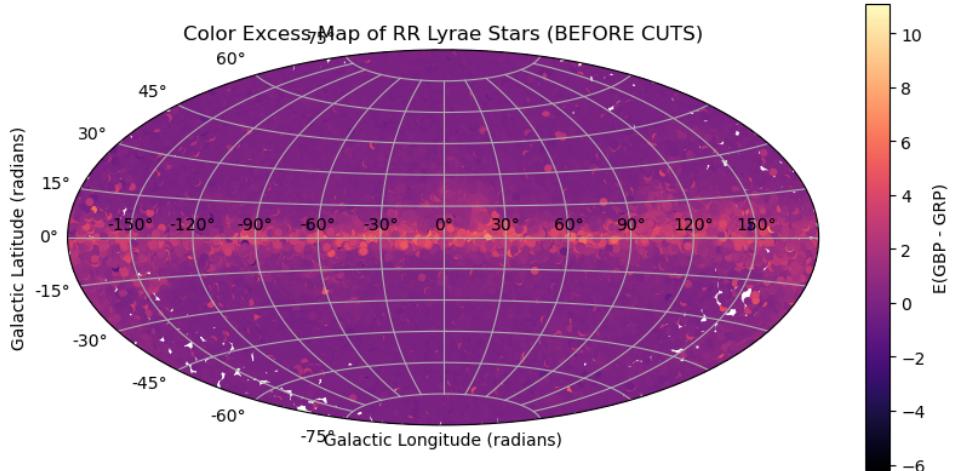


Figure 45: Plotting 2-D Map of Color Excess before creating cuts for extinction

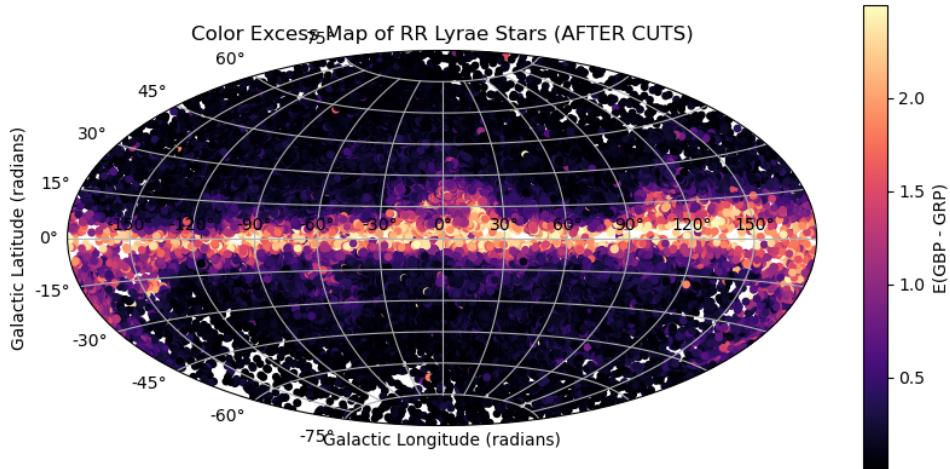


Figure 46: Cleaning Reddening Map of Milky Way with post extinction cuts

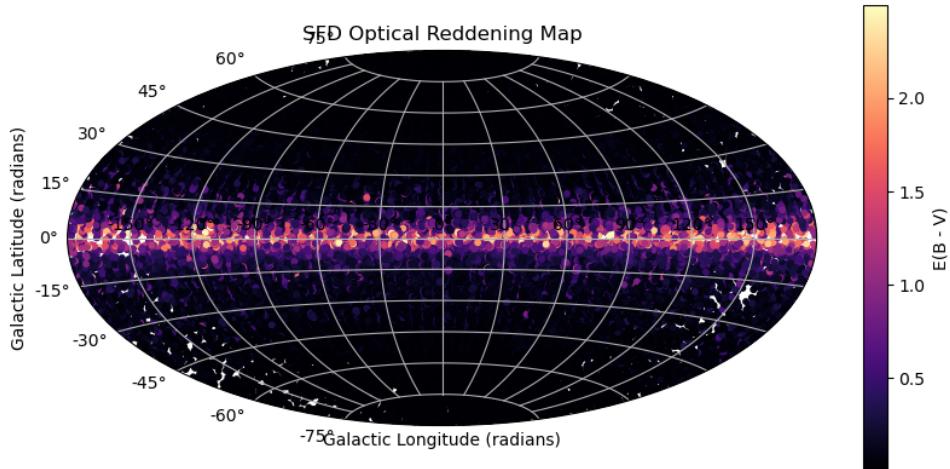


Figure 47: Comparison with notable SFD Dust Map

4 Discussion

4.1 Period Comparison

The periods for my data seem to be ranging from .3 to .6 (bottom table) whereas the ones from the RR Lyrae catalog range from .3 to .6. (top table), overall not much of a difference between both. This would demonstrate how accurate my results were with the data form the rrlyrae catalog

4.2 Functional Form Prediction

Suppose we want to predict how bright an RR Lyrae star will be at some time in the future. To do this, we need to know a functional form, $mG(t)$, which we can evaluate at some future time t . Stellar atmospheres are complicated, so predicting a closed-form expression for $mG(t)$ from first principles is hard. However, since we expect the fluctuations to be periodic, we can use some results from Fourier analysis to get a good estimate of $mG(t)$. Any periodic, smooth function $f(t)$ can be described by a sum of sines and cosines:

$$f(t) = A_0 + \sum_{k=1}^K [a_k \sin(k\omega t) + b_k \cos(k\omega t)],$$

where A_0 , a_k , and b_k are to-be-determined constants (a_k and b_k are arrays that are K terms long), and $\omega = 2\pi/P$ is the angular frequency. This equation is exact in the limit of $K \rightarrow \infty$. The

reason it is useful for us is that for typical periodic functions, it is reasonably accurate even for small K (say, $K = 5$ or 10). In the most general case, determining all the $2K + 2$ free parameters needed to represent $f(t)$ is a complex nonlinear optimization problem. However, if w is already known (say, from a periodogram), the problem becomes much simpler. If w is known, then the problem of determining the remaining $2K + 1$ free parameters can be re-cast as a linear algebra problem; i.e., $y = X^* \text{Beta}$. Here y is an array of the measured fluxes, X is a matrix that can be constructed from known quantities, and Beta is an array of unknowns. Here are X and Beta :

$$X = [1, \sin(w), \cos(w), \sin(w * 2), \cos(w * 2), \dots, \sin(w * k), \cos(w * k)] [1, \sin(w * 2), \cos(w * 2), \sin(2 * w * 2), \cos(2 * w * 2), \dots, \sin(k * w * 2), \cos(k * w * 2)] [\dots, \dots, \dots, \dots, \dots, \dots, \dots]$$

$$\text{Beta} = [A_0] [a_1] [b_1] [a_2] [b_2] [\dots] [a_k] [b_k]$$

4.3 Figure 8 and 9: Optimal Number of Terms

From looking at the table above, K would be 10 since the best MSE happens before any error happened or the point of decreasing or increasing which we see its true on these graphs. Then from the plots we could as well verify $K = 10$ is nearing 10 validating itself as the most terms that should be enough!

4.4 Figure 11 and 12: Mean Magnitude Estimation with Fourier Model

From the plot above we're able to see how the fourier models seem to be much more accurate to the gaia catalog estimates in comparison with those from part 3. For a closer analysis, the residuals between fourier and part 3 were plotted and we clearly see how the fourier has residuals nearer to 0 whereas for part 3 it's far from there.

4.5 Figure 14 to 19: Comparing RRab and RRc Light Curve Shapes

RRab stars have asymmetric, sawtooth shaped light curves with longer periods and larger amplitude variations. In contrast, RRc stars exhibit more sinusoidal curves with shorter periods and smaller amplitude variations and some intrinsic scattering.

The analysis presented in Netzel et al. 2018 suggests that a significant fraction of RRc stars, about 5.6 percent in their sample, exhibit the Blazhko effect, which is a quasi-periodic modulation of pulsation amplitude and/or phase. While the light curves of these RR Lyrae stars are well-described by a single period on average, the presence of the Blazhko effect introduces intrinsic scatter and deviations from simple periodicity. This is evidenced by the detection of multiplets and sidepeaks in the frequency spectrum, indicating modulation in pulsation amplitude and phase. Therefore, while a single period can generally describe the light curve shapes of RR Lyrae stars, there is evidence of intrinsic scatter due to phenomena like the Blazhko effect.

4.6 RR Lyrae Period-Luminosity Relation with Gaia Distances

To accurately determine the absolute magnitude of RR Lyrae stars with Gaia distances, it is essential to have a clear line of sight with minimal dust content. Dust can obscure and dim the starlight, resulting in incorrect distance estimations and, consequently, inaccurate calculations of the stars' intrinsic brightness.

4.7 Distance Estimates with Galactic Structure Prior

The geometric distances in the Bailer-Jones catalog are calculated using a Bayesian approach that incorporates a prior based on the expected distribution of stars in the Galaxy. This method provides a more reliable estimate of distance, especially for stars with large relative parallax errors. Unlike the simple inversion of parallax, which can lead to biased and uncertain distance estimates, the Bayesian approach accounts for the asymmetric nature of distance uncertainties and the non-linear relationship between parallax and distance.

4.8 Figure 23 and 24: Galactic Distribution of Targets

The trend observed in the plot indicates that for nearby stars, the naive and Bailer-Jones distance estimates are similar, but for more distant stars, the naive estimates tend to underestimate the distance compared to the Bailer-Jones estimates. This discrepancy arises because the naive method does not consider the non-linearity of the parallax-distance relationship and the asymmetry of the distance probability distribution, whereas the Bailer-Jones method employs a probabilistic approach that accounts for these factors, leading to more accurate distance estimates for distant stars.

The comparison between DR3 and Bailer-Jones parallaxes shows a negligible average difference of -0.01 mas, indicating high consistency between the two sets of measurements. The majority of data points align with the line of equality (red dashed line), with overlapping error bars suggesting that the parallaxes agree within their uncertainties. This consistency is likely due to the Bailer-Jones model's correction for biases in parallax inversion, particularly for distant stars with significant parallax errors.

4.9 Figure 26: Scatter Reduction with Quality Cuts

After applying the 2 cuts of Lindegren et al., the scattering has been eliminated for the most part !

4.10 Figure 33: Fitting Period-Luminosity Relation with M-H MCMC (i) 5

The spread between samples as a function of period is consistent with the data, as evidenced by the absolute magnitudes averaging between 0.7 and 0.9.

4.11 Figure 38: Fitting Period-Luminosity Relation with PYMC Model (iii) 2

The No-U-Turn Sampler (NUTS) is a more efficient and user-friendly extension of the Hamiltonian Monte Carlo (HMC) sampler, designed to require less manual tuning than traditional Metropolis-Hastings samplers. NUTS leverages gradient information for exploration, automatically adjusts its parameters, and often achieves better convergence with fewer samples, making it particularly effective for high-dimensional or complex posterior distributions.

4.12 G-band period-luminosity relations VS the V-band relations in studies

Comparing our G-band period-luminosity relations with the V-band relations in studies like Beaton et al. 2018 and Klein and Bloom 2014 reveals systematic discrepancies. These differences could arise from variations in the wavelength range of the bandpasses, differences in sample selection criteria, and diverse methods for correcting extinction and reddening. Furthermore, differences in the techniques used to fit the period-luminosity relations and manage uncertainties might also explain the observed discrepancies between our results and those from the literature.

4.13 Figure 44: Comparison with Catalog's $G_{absorption}$

AG and $G_{absorption}$ seem to be relatively close to their values

4.14 Figure 46: Cleaning Reddening Map

The distribution of RR Lyrae stars in the Gaia catalog is not uniform primarily due to the structure of the Milky Way, with higher concentrations in the central bulge and halo, and interstellar dust obscuring stars in the Galactic plane, leading to lower detection rates in these regions. These factors result in a non-uniform distribution across the sky, reflecting the intrinsic properties of the Galaxy and the effects of extinction

4.15 Comparison with SFD Dust Map

The overall structure of my map aligns with the SFD map, although the SFD map exhibits higher excess coloring near the center (the margin is much more thin) and less scattering, resulting in a more uniform appearance at the center compared to my map.

We do not anticipate an exact match between our color excess map and the SFD dust map, as they measure different aspects of dust distribution. The SFD map, based on far-infrared measurements, gives a broader view of total dust along the line of sight. In contrast, our map focuses on the differential reddening of RR Lyrae stars, providing a more targeted view of dust along their specific paths, which may not encompass all features seen in the SFD map.

5 Conclusion

In this study, we explored the intricate relationship between RR Lyrae stars and Galactic dust using the wealth of data provided by the Gaia mission. Through our analysis, we delved into the period-luminosity relation of RR Lyrae stars, utilizing their role as standardizable candles to estimate distances and subsequently map the distribution of interstellar dust. Our findings revealed the complex nature of dust attenuation and its impact on astronomical observations, highlighting the importance of accurate distance measurements and quality cuts to mitigate the effects of dust.

We compared our derived period-luminosity relations with those from the literature, noting systematic differences that could be attributed to variations in bandpass wavelength ranges, sample selection criteria, and methods for correcting extinction and reddening. Our study also underscored the non-uniform distribution of RR Lyrae stars in the Gaia catalog, reflecting the structure of the Milky Way and the influence of interstellar dust.

By constructing a color excess map and comparing it to the SFD dust map, we gained insights into the differential reddening of RR Lyrae stars and the overall dust distribution in the Galaxy. While our map did not perfectly match the SFD map, it provided a valuable perspective on the localized dust environment around these pulsating variables.

Overall, this exploration has deepened our understanding of the interplay between RR Lyrae stars and Galactic dust, reinforcing the significance of these stars as tools for probing the structure and evolution of our Galaxy.

References

- Netzel et al. 2018: <https://academic.oup.com/mnras/article/480/1/1229/5055626>
Bailer-Jones et al. 2020: <https://arxiv.org/abs/2012.05220>
Lindegren et al. 2018: <https://arxiv.org/abs/1804.09366>
Klein and Bloom 2014.: <https://arxiv.org/abs/1404.4870>
Schlegel, Finkbeiner, and Davis 1998. : <https://arxiv.org/abs/astro-ph/9710327>
Gaia Collaboration, Brown, A.G.A., Vallenari, A., Prusti, T., de Bruijne, J.H.J., et al. (2018).
Gaia Data Release 2: DOI: 10.1051/0004-6361/201833051