Revealing the age of NGC 2509 with Gaia EDR3

Thinh Nguyen

November 2021

Abstract

The age of the open cluster NGC 2509 has been a confusing topic due to widely different values determined by different studies. With the recent significant improvement of *Gaia* Early Data Release 3, we aim to resolve the confusion by using the updated data to determine the cluster's age and other parameters such as parallax, metallicity, and absorption. We employ pyUPMASK, an unsupervised cluster algorithm, to calculate the membership probability and then select 254 cluster members whose probability is larger than 0.5. These stars are then used as input for the BASE-9 tool to find the best fitted stellar isochrone using the Markov chain Monte Carlo. After 30000 iterations, we obtain the estimated value of 9.2037, 0.3469, 0.4422 milli-arc-seconds, and 0.0206 for logAge, [Fe/H], parallax, and absorption A_V, respectively. Our determined age is consistent with about half of the previous studies while disagrees with the others. The results of the metallicity and absorption are greatly different from the values in the literature. However, it should be noted that earlier studies assumed solar metallicity for this cluster.

A Introduction

One of the drawbacks of astronomy is that researchers cannot do direct experiments on the objects they are interested in. This yields great difficulties when studying celestial bodies because numerous parameters come into play simultaneously whereas researchers have no easy ways to constrain those parameters. Fortunately, while we cannot change the properties of stars by ourselves, the universe provides us with many giant natural laboratories that fix certain parameters while varying others, providing us a useful tool to study stars. Those laboratories are open clusters.

As stars in an open cluster are born approximately at the same time and distance as well as composed of identical chemical compositions, mass is the determining variable that differs among the stars (von Hippel, 2005). Because more massive stars leave the main sequence earlier and open clusters have stars from a wide range of masses, the Hertzsprung-Russell (HR) diagram of open clusters usually has a distinct feature called the main-sequence turn-off point. Thus, one can fit an isochrone, a curve on the HR diagram representing a stellar population with similar age, to a star cluster to compute cluster parameters. Isochrones are calculated by using simulations to evolve an initial set of stars with the same composition to certain ages. Different theoretical isochrones are compared with the observational HR diagram to determine the best match.

Even though isochrone fitting is a good method to determine open cluster parameters, it has some drawbacks when the data is sparse or inaccurately defined. NGC 2509 is an example of such a case. NGC 2509 is an open cluster (Kharchenko et al., 2013) located approximately at 08:00:48 hour for right ascension (RA) and -19:03:06 degrees for declination (Dec). Recently, the distance to NGC 2509 is confirmed to be around 2500 parsec (Cantat-Gaudin et al., 2018) using

Gaia Data Release 2 (DR2) (Gaia Collaboration et al., 2018). The cluster had been a topic of debate as different studies resulted in very different and confusing values for its age (Carraro & Costa, 2007; Sujatha & Babu, 2003). In particular, even though they all used the same method of isochrone fitting, Sujatha & Babu (2003) reported the age to be 8 billion years old, Ahumada (2000) computed it to be 1 billion, in Carraro & Costa (2007)'s work it is 1.2 billion, and in the most recent paper, de Juan Ovelar et al. (2020) estimated it to be 861 million years old. This discrepancy might be explained by the inconsistency of data used to plot the HR diagram: Cantat-Gaudin et al. (2018) and de Juan Ovelar et al. (2020) used *Gaia* DR2 data while Sujatha & Babu (2003) and Ahumada (2000) used CCD photometry of their telescopes. Furthermore, unlike other open clusters, NGC 2509 has a quite narrow main sequence turn off (de Juan Ovelar et al., 2020) and does not contain a lot of stars in that region, which brings greater challenges in fitting the correct isochrone to the cluster's morphology in the HR diagram.

With the recent publication of *Gaia* Early Data Release 3 (EDR3), we propose to use the new data to verify and re-determine the age and other parameters of NGC 2509. *Gaia* (Gaia Collaboration et al., 2016) is a space mission by the European Space Agency (ESA) that surveys all astrometric, photometric, and spectroscopic measurements of our Galaxy on an all-sky scale on only one platform. The precision in *Gaia*'s astrometry is unprecedented with the accuracy down to 24 microarcseconds (Gaia Collaboration et al., 2016). Compared to *Gaia* DR2, *Gaia* EDR3 demonstrates a significant improvement in many aspects. Specifically, precision increases by 30% in parallax measurements and by 200% in proper motions (Gaia Collaboration et al., 2021). There are also better photometric measurements and a greater homogeneity across color, magnitude, and coordinate (Gaia Collaboration et al., 2021). Thus, as previous studies on the cluster were conducted with DR2 data, all of these improved elements of EDR3 are crucial to explore the confusion around NGC 2509.

The project employ robust techniques in the field of open clusters to ensure the best possible result. These include the Bayesian approach proposed by Sampedro & Alfaro (2016) to determine the cluster membership and the open-source software Bayesian Analysis for Stellar Evolution with Nine Variables (BASE-9) (von Hippel et al., 2006) to compute the cluster's parameters (age, metallicity, line-of-sight absorption, and distance modulus) by isochrone fitting. The details of these techniques will be further discussed later in the paper. Combined with improved data that minimize the interfering scattering effect, the techniques will help further refine the cluster's parameters.

The paper will be structured as follows. Section B elaborates the methods that are utilized for cluster membership calculation and isochrone fitting. The results of the computations are discussed and compared with literature in Section C. Lastly, Section D summarizes all the analysis and explains the further implication.

B Methodology

B.1 Data

The study utilizes photometry (G, BP, and RP mean magnitude) astrometry (parallax and proper motions) data from *Gaia* EDR3 public archive [1]. The query was conducted with a 14 arcmin radius search around the NGC 2509's coordinate (08:00:48, -19:03:06). Sánchez et al. (2020) calculate the radius of NGC 2509 to be 8.15 arcmins from *Gaia* DR2 data. Therefore, we overestimated the searched radius to be 14 arcmins to include all possible cluster members. As a result, the query returns 10260 targets.

[1] https://gea.esac.esa.int/archive/

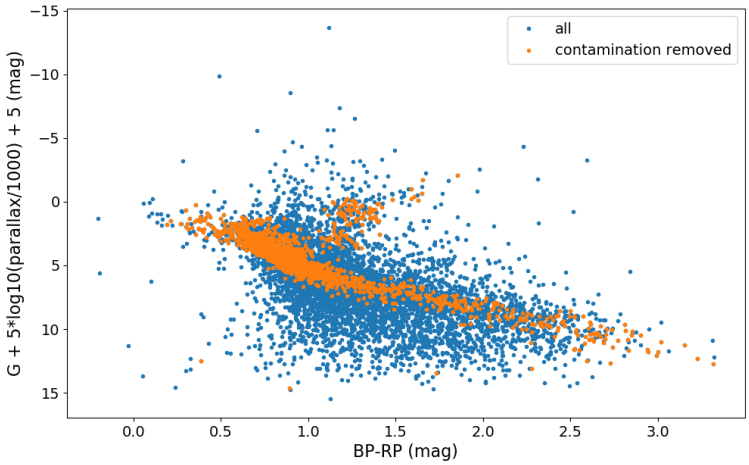

Figure 1: The color-absolute magnitude diagram of NGC 2509 from the *Gaia*'s query and after the contamination removal procedure described in Equation 1.

B.2 Contamination removal

After obtaining the query result, we first remove targets without data on either photometry, proper motion, or parallax. Next, because calculating the cluster membership probability relies heavily on the astrometric information, to receive the best possible cluster members, we limit contaminated and unreliable sources by applying some restrictions on the *Gaia*'s astrometric parameters, specifically

$$\begin{aligned} &\texttt{ruwe} \leq 1.4 \\ &\texttt{astrometric\_excess\_noise\_sig} \leq 2 \\ &\texttt{parallax\_over\_error} \geq 5 \text{ mas}. \end{aligned} \tag{1}$$

In the equations, `ruwe` stands for the renormalized unit weight error. Sources that a have good fit to a single-star model are expected to have the `ruwe` value approximate to 1. A large `ruwe` indicates that the examined sources are not single stars or there are problems with their astrometric solution (Lindegren et al., 2021; Lindegren, 2018). We take the recommended value by *Gaia* of 1.4 (Lindegren, 2018) for `ruwe`. Regarding the second part of the restriction, the `astrometric_excess_noise` measures the inconsistency between observations and the best fitting standard astrometric models. The value is considered significant if the `astrometric_excess_noise_sig` is larger than 2. Therefore, we set the limit at 2 to obtain astrometrically reliable sources (Lindegren et al., 2021). Finally, we require the parallax measurement to be at least five times larger than its uncertainty. As a result, from 10260 targets in the query, we narrow down our sample to 1683 targets with reliable astrometric solutions. Figure 1 displays the color - absolute magnitude diagram of our sample before and after the contamination removal procedure.

B.3 Membership probability

To determine the cluster membership probability, we utilize pyUPMASK, an unsupervised cluster algorithm (Pera et al., 2021). pyUPMASK is an enhanced and faster version of UPMASK (Krone-Martins & Moitinho, 2014) and also converts the UPMASK's original code in R language to Python language. Both UPMASK and pyUPMASK share a similar 2-loop procedure to assign cluster

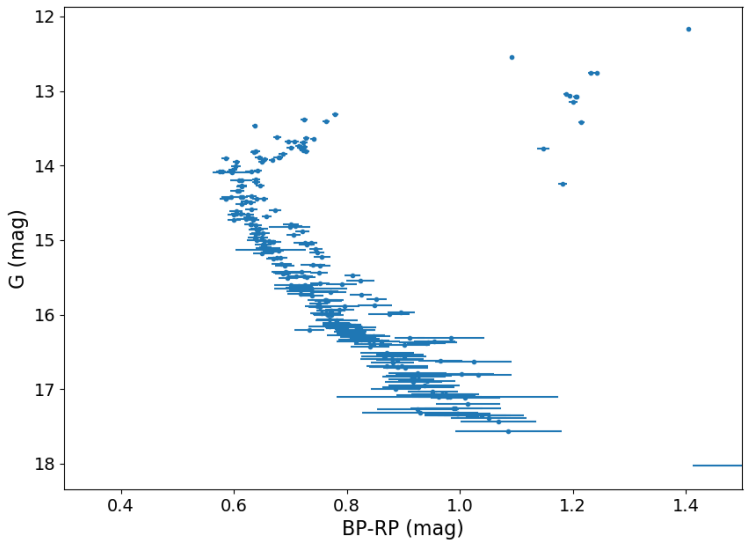

Figure 2: The color-apparent magnitude diagram of 254 NGC 2509's member stars, which are identified by pyUPMASK with the membership probability larger than 0.5

membership probability. In the inner loop, using the non-positional features (in our case, we use proper motions), the clustering method is responsible for dividing the whole cluster data into smaller clusters with N members per cluster. For each sub-cluster, the random field rejection method is applied to test for clumps in positional space. The method will compare the kernel density estimation of the coordinate space to that of a two-dimensional uniform distribution in the same range. If the two kernels are similar, all stars in that sub-cluster are rejected as field stars; otherwise, they will be retained to another iteration of the inner loop. The inner loop ends when no more sub-clusters are dismissed. Next, the outer loop will divide the original cluster again to feed the inner loop, and the whole cycle keeps repeating by the number of outer loops specified by the user. The membership probability for each star will be calculated by averaging the number of times that star is classified as a cluster member in each outer loop. For our use, we follow the pyUPMASK's default setting: mini-batch k-means for the clustering method (Sculley, 2010), 25 outer loops, 25 maximum inner loops, and 25 members per sub-cluster.

Once having all stars' membership probability, we decided to set a cut-off value of 0.5 for a cluster member. As a result, we have 254 cluster member stars and 1429 field stars in our sample. Detailed membership probability will be provided in a table along with the online version of this paper. In Figure 2, the color-apparent magnitude diagram of our cluster shows a clear main-sequence with its turnoff point. Red giants are also noticeable in the plot, and we do not have any white dwarfs or blue stragglers in our sample.

B.4 Fitting isochrone

Having the NGC 2509's list of members, we adopt the open-source software Bayesian Analysis for Stellar Evolution with Nine Variables (BASE-9) (von Hippel et al., 2006) to determine the cluster's parameters. This method uses the Bayesian approach to fit a set of theoretical isochrones with the observational distribution of cluster members. BASE-9 adjusts its set of parameters for each

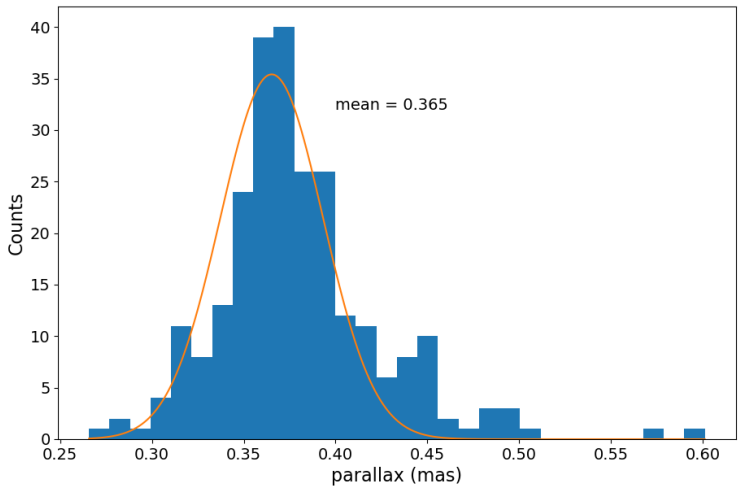

Figure 3: The parallax distribution of 254 member stars. The mean of the fitted Gaussian distribution is used as the prior mean for parallax in the BASE-9 MCMC run.

iteration by using the Markov chain Monte Carlo (MCMC) technique. MCMC involves random sampling from the probability distribution to estimate a quantity, in which one sample step is dependent on the previous step.

Among the list of parameters available in BASE-9, we aim to estimate the cluster's logAge, metallicity ([Fe/H]), the extinction in V band (A_V), and observed modulus, which accounts for both distance and extinction. We leave helium mass fraction and carbonicity for white dwarfs fixed at the BASE-9's default values of 0.29 and 0.38, respectively. We refer to literature as much as possible for our prior. We set our prior mean for the logAge and absorption A_V, to be 9.2 and 0.322, which come from Kharchenko et al. (2013). As shown in Figure 3, our parallax prior comes from fitting the Gaussian distribution to the *Gaia* parallax distribution of the members themselves. According to the fit, the prior mean for parallax is set to be 0.000365 arc-second. Since we do not have any information regarding the metallicity, we leave the prior mean as the default value of 0.01 and set its prior sigma to be 0.1. This value is a reasonable assumption as the metallicity of open clusters often falls in the range of -0.3 to 0.3 (Netopil et al., 2016). Regarding the prior sigma, we left it as 0.1 for all parameters prior. We also want to note that we chose the sigma for logAge to be uniform, as we want the age to be free to vary throughout the parameter space. The number of iteration used in our run is 30000. For the step sizes, we chose 0.01, 0.05, 0.1, 0.02 for the logAge, [Fe/H], parallax, and A_V, respectively. The summary of all the prior parameters can be found in Table 1.

The average posterior parameters are determined by calculating the weighted mean for each posterior distribution obtained from BASE-9. The weight is the negative inverse of the log of the unnormalized posterior probability, provided by BASE-9 as logPost. These posterior distributions only come from the adaptive run of the MCMC, and we do not take into account values from the two burnin stages. The weighted mean and its corresponding standard deviation of each parameter is given in Table 1.

All the mean posterior parameters are utilized with the module `makeCMD` in BASE-9 to compute the average isochrone that fits the color-apparent magnitude diagram of the cluster. Figure 4 demonstrates a good match of the found isochrone through the data points. Even though there are

Table 1: The BASE-9 prior and posterior parameters of NGC 2509's logAge, metallicity [Fe/H], modulus, extinction A_V, helium mass fraction, and carbonicity.

	Prior mean	Prior σ	Step size	Posterior mean	Posterior σ
logAge	9.2	uniform	0.01	9.2037	0.00283
Fe/H	0.01	0.1	0.05	0.3469	0.01576
parallax (mas)	0.365	0.1	0.005	0.4422	0.00249
Av	0.322	0.1	0.02	0.0206	0.00868
He mass fraction	0.29	0.0	0.0	0.29	0.0
Carbonicity	0.38	0.0	0.0	0.38	0.0

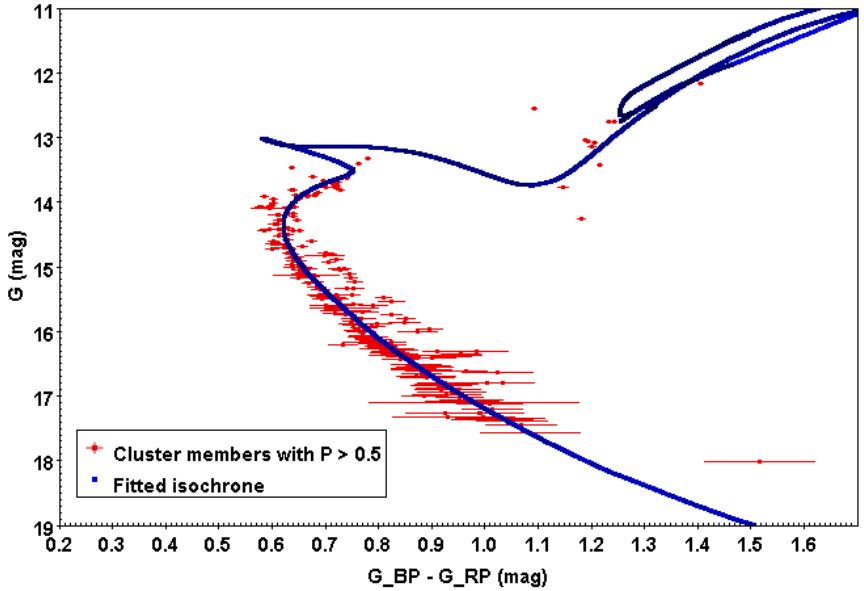

Figure 4: The fitted isochrone with parameters listed in Table 1 is over-plotted on the NGC 2509's member stars.

few red giants, the isochrone still fits well with the turnoff point and the red giant branch.

C Discussion

C.1 Membership comparisons with literature

We compare our list of members from using *Gaia* EDR3 data with that from using *Gaia* DR2 data (Cantat-Gaudin et al., 2018). Similar to us, Cantat-Gaudin et al. (2018) used the UPMASK method (Krone-Martins & Moitinho, 2014) to detect 253 NGC 2509's members whose membership probability is larger than 0.5. For us, that number is 254, as discussed in Section B.3. We perform the cross-match by comparing the EDR3 coordinates (RA and DEC) in our list with their DR2 coordinates. Given the maximum separations between 2 matched points of 1 arcsec, we obtain 203 overlapped targets (80% of our list). This high percentage emphasizes the consistency between our work and the literature and between the two *Gaia*'s data releases. The list of the cross-matched targets will be provided in the membership probability table along with the online version of this table.

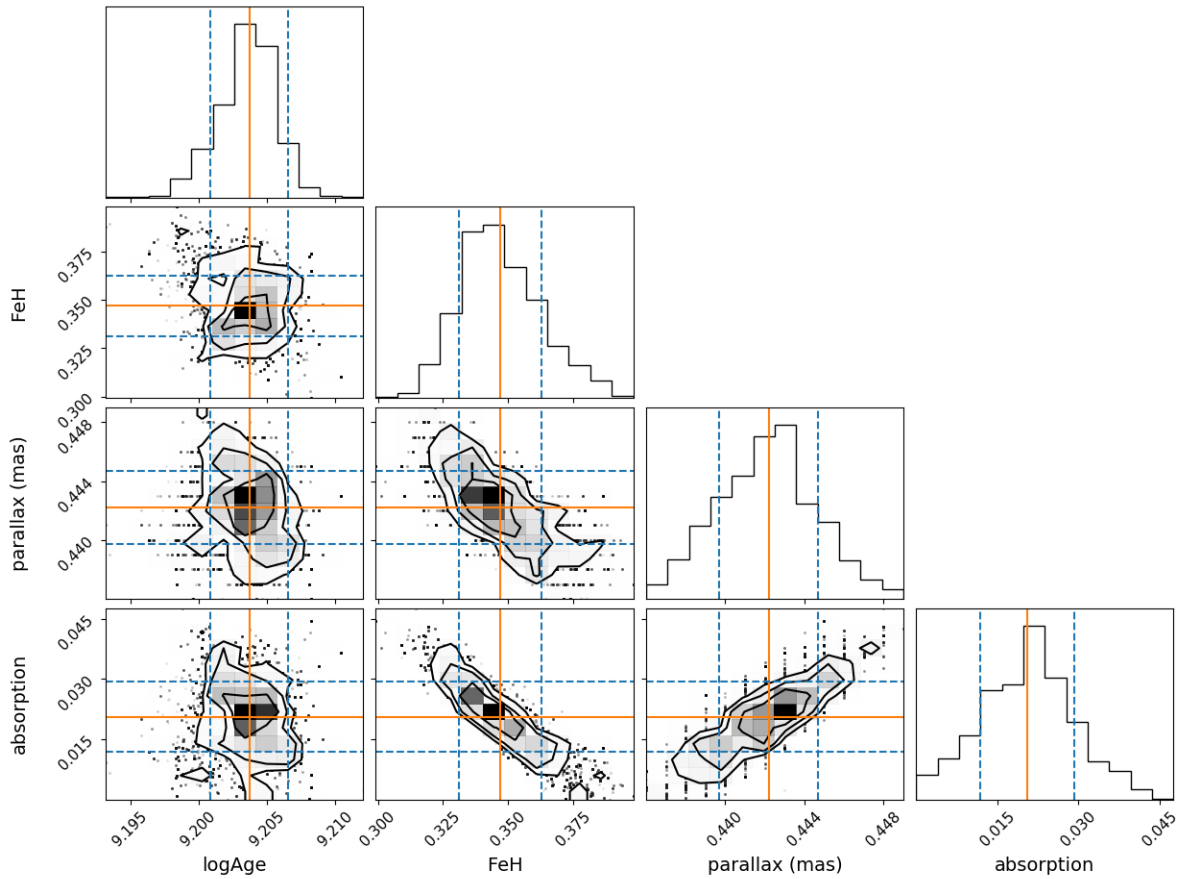

Figure 5: The corner plot showing the correlations between the posterior parameters.

C.2 Corner plot

To study the correlation between the samplings in the 4-parameter space, we produce the corner plot by employing the Python module `corner.py` (Foreman-Mackey, 2016). The visualization plots the two-dimensional projection of the MCMC samplings from BASE-9 to reveal possible covariances and to test the samplings' convergence. In our corner plot displayed in Figure 5, all posterior distributions look Gaussian, and thus they suggest that MCMC has reached convergence with the iteration number of 30000. In the first column, the 2-dimensional contour plots show a concentric pattern around a peak density region, emphasizing that the sampling of logAge is random. On the other hand, the elliptical shapes of their contour plots demonstrate that there are correlations between [Fe/H], parallax, and absorption A_V samplings. Specifically, [Fe/H] and absorption have a strong negative correlation, parallax and absorption have a strong positive correlation, and [Fe/H] and parallax have a moderate negative correlation.

The corner plot also helps us evaluate the accuracy of our average posterior parameters. In the plot, the orange lines represent the weighted means of the posterior parameters (as given in Table 1). Two blue dashed lines show one standard deviation around the weighted mean. As all intersections among the orange lines are located at the highest density region ad within the 1-σ regions of the contour plots, we are more confident with our results for NGC 2509.

Table 2: The robustness analysis of the Bayesian approach of BASE-9. We examine the effect of the prior on the posterior by using 2 different logAge values and 2 different absorption values (4 combinations in total).

Prior				Posterior			
logAge	FeH	Parallax (mas)	absorption	logAge	FeH	Parallax (mas)	absorption
9.2	0.01	0.365	0.322	9.2037	0.3469	0.4422	0.0206
9.08	0.01	0.365	0.322	9.2037	0.3470	0.4421	0.0202
9.2	0.01	0.365	0.465	9.2040	0.3446	0.4427	0.0216
9.08	0.01	0.365	0.465	9.2034	0.3453	0.4425	0.0220

C.3 Robustness analysis

Another method to evaluate the convergence of the MCMC run is to do the robustness analysis with different starting priors. Beside the set of prior values described in Section B.4, we use another logAge value of 9.08 from Carraro & Costa (2007) and another absorption value of 0.456 from Dias et al. (2002), resulting in an addition of three more prior sets. Other configuration parameters such as step size, number of iteration, and prior σ are kept the same as in Section B.4. Table 2 shows the MCMC results for these four sets. According to the table, the different prior values ineligibly affect the posterior parameters. For example, at the prior absorption of 0.322, when we change our prior logAge from 9.2 to 9.08 (0.12 difference), the posterior logAge remains identical to the fourth decimal place. Similarly, at the prior logAge of 9.08, the change of 0.143 in the prior absorption value (from 0.456 to 0.322) only returns the changes of 0.0018 in the posterior value (from 0.0220 to 0.0202). Furthermore, the posterior [Fe/H] and parallax do not significantly alter among the four prior sets.

D Conclusion

In this study, we use *Gaia* EDR3 data to explore the logAge, metallicity [Fe/H], parallax, and absorption A_V of the open cluster NGC 2509. Our calculation using pyUPMASK algorithm (Pera et al., 2021; Krone-Martins & Moitinho, 2014) results in 254 stars whose membership probability is larger than 0.5, which is our criterion to declare a cluster member. The color-apparent magnitude diagram of the cluster members shows a clear main sequence with a noticeable turn-off point. The result of BASE-9 (von Hippel et al., 2006), a method to fit isochrone by using Bayesian analysis with MCMC, are 9.2037, 0.3469, 0.4422 mas, 0.0206 for logAge, [Fe/H], parallax, and A_V, respectively. We also show that the test reached convergence by examining the corner plot and conducting the robustness analysis with different sets of priors.

In response to the confusion around the age of NGC 2509, we found the age of NGC 2509 to be approximately 1.6 billion years. This value is relatively consistent with the values of Ahumada (2000); Carraro & Costa (2007); Kharchenko et al. (2013), but is different from the values of de Juan Ovelar et al. (2020); Dias et al. (2002) and especially the value of 8 billion years old by Sujatha & Babu (2003).

Among the four parameters, the posterior logAge and parallax do not shift much from the prior values, while the posterior metallicity [Fe/H] and the absorption A_V change significantly. Even though they are still possible, the two posterior values of [Fe/H] and A_V are surprising. We found that NGC 2509 is metal heavy ([Fe/H] = 0.3469) even though it is a young open cluster. However, previous studies did not provide metallicity or assumed solar metallicity when fitting isochrone

(Dias et al., 2002; Kharchenko et al., 2013). Thus, we have not reference to compare our result. Also, the absorption value is small compared to the great distance that the cluster is at.

E Acknowledgement

We would like to thank Dr. Andrej Prsa, Dr. Scott Engle, Dr. von Hippel, and Dr. Laurent Eyer for the valuable help and discussion while doing this project. Notably, this study relies heavily on using BASE-9, and Dr. von Hippel helps us learn to use the code much faster.

F Bibliography

Ahumada, J. A. 2000, in Astronomical Society of the Pacific Conference Series, Vol. 198, Stellar Clusters and Associations: Convection, Rotation, and Dynamos, ed. R. Pallavicini, G. Micela, & S. Sciortino, 43

Cantat-Gaudin, T., Jordi, C., Vallenari, A., et al. 2018, A&A, 618, A93, doi: 10.1051/0004-6361/201833476

Carraro, G., & Costa, E. 2007, A&A, 464, 573, doi: 10.1051/0004-6361:20066350

de Juan Ovelar, M., Gossage, S., Kamann, S., et al. 2020, MNRAS, 491, 2129, doi: 10.1093/mnras/stz3128

Dias, W. S., Alessi, B. S., Moitinho, A., & Lépine, J. R. D. 2002, A&A, 389, 871, doi: 10.1051/0004-6361:20020668

Foreman-Mackey, D. 2016, The Journal of Open Source Software, 1, 24, doi: 10.21105/joss.00024

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1, doi: 10.1051/0004-6361/201629272

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1, doi: 10.1051/0004-6361/201833051

—. 2021, A&A, 649, A1, doi: 10.1051/0004-6361/202039657

Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., & Scholz, R. D. 2013, A&A, 558, A53, doi: 10.1051/0004-6361/201322302

Krone-Martins, A., & Moitinho, A. 2014, A&A, 561, A57, doi: 10.1051/0004-6361/201321143

Lindegren, L. 2018. http://www.rssd.esa.int/doc_fetch.php?id=3757412

Lindegren, L., Klioner, S. A., Hernández, J., et al. 2021, A&A, 649, A2, doi: 10.1051/0004-6361/202039709

Netopil, M., Paunzen, E., Heiter, U., & Soubiran, C. 2016, A&A, 585, A150, doi: 10.1051/0004-6361/201526370

Pera, M. S., Perren, G. I., Moitinho, A., Navone, H. D., & Vazquez, R. A. 2021, A&A, 650, A109, doi: 10.1051/0004-6361/202040252

Sampedro, L., & Alfaro, E. J. 2016, MNRAS, 457, 3949, doi: 10.1093/mnras/stw243

Sánchez, N., Alfaro, E. J., & López-Martínez, F. 2020, MNRAS, 495, 2882, doi: 10.1093/mnras/staa1359

Sculley, D. 2010, in Proceedings of the 19th International Conference on World Wide Web, WWW '10 (New York, NY, USA: Association for Computing Machinery), 1177–1178, doi: 10.1145/1772690.1772862

Sujatha, S., & Babu, G. S. D. 2003, Bulletin of the Astronomical Society of India, 31, 9

von Hippel, T. 2005, arXiv e-prints, astro

von Hippel, T., Jefferys, W. H., Scott, J., et al. 2006, ApJ, 645, 1436, doi: 10.1086/504369