RESEARCH PROJECT: MAPPING THE MILKY WAY GALAXY USING THE ESA'S GAIA SPACE MISSION

INVESTIGATING THE IMPACT OF EXTINCTION ON SED TEMPLATES IN GAIA DATA

Creation of Template Matches for Catalog Stars and Determination of Ideal Extinction Level

December 19, 2017

Ryan Skoletsky
Student ID: 114715111
The Honors College at the University of Maryland, College Park
Department of Astronomy

Abstract

Launched in late 2013, the GAIA space mission, pioneered by the European Space Agency, is currently in the process of collecting and disseminating large data releases, the first of which has already been made public, with the hopes of constructing a high-fidelity map of our home galaxy. A vital aspect of understanding the data from GAIA and thus achieving its long-term goals is comprehending the impact of interstellar extinction on observations. In this study, the effect of extinction on the catalog constructed from cross-matching the first GAIA release and earlier open catalogs is explored. Using a set methodology, the ideal level of extinction is determined by finely modifying the reference V filter extinction at a distance of 200 parsecs from Earth and finding where the lowest average value of χ^2 for SED template-matching each star is achieved. Doing so, it can be concluded that the ideal correction for the data is zero extinction due to the inherent selection effect causing bluer stars to disproportionately appear at greater distances. This finding is the first step in the eventual Milky Way Galaxy three-dimensional map of star and interstellar dust, and this work should be extended to give a better picture of how extinction impacts GAIA observations at large.

Contents

Chapter 1

Introduction

The European Space Agency's (ESA's) Global Astrometric Interferometer for Astrophysics (GAIA), hereafter mentioned only as GAIA, is a space mission designed with the goal of creating an accurate, intricate three-dimensional map of the Milky Way Galaxy, including locations of stars, interstellar dust, and other celestial bodies[1]. In the end, the program hopes to measure locations and radial velocities of approximately 1 billion stars, comprising roughly 1 percent of the galaxy's stellar population[1]. A key piece of the puzzle of interstellar and especially galactic positioning is having a comprehensive understanding of extinction and the way it affects observations. In general, interstellar extinction causes celestial bodies to appear dimmer and redder than they are in reality, with this effect becoming more pronounced for greater distances.

In this study, the impact of extinction on fitting the data from a refined catalog of GAIA's first major release to SED templates is investigated. Within this catalog, the subset of the mission's data that are utilized are main sequence stars which have high-quality data and which are no more than 200 parsecs (pc) away from Earth. I sought out to determine what the ideal level of extinction is to maximize the goodness of fit between the observations and templates, as measured by minimizing the average of each star's reduced χ^2 value. In particular, the "level of extinction" will be used interchangeably throughout this report with what it mathematically represents, which is the reference extinction magnitude in the V filter at a distance of 200 pc using a linear extinction-distance model. The catalog and methodology will be walked through, and a solid answer that conforms with expectations was obtained. Such a study of how extinction works in conjunction with empirical space mission data is a first step in eventually achieving the end goals of the ESA's venture.

Chapter 2

Construction of the GAIA Data Catalog

The data utilized in this study came from a mixture of the first major release from the GAIA mission and open catalogs obtained through the Internet. Specifically, cross-matches were made with the *Tycho*-2, Wright 2003, and 2MASS catalogs, which are openly available online through the CDS Portal[2], [3]. The cross-match feature was employed also through the assistance of the CDS Portal[3]. In obtaining the data, a filtering selection was imposed to restrict the stars included to lie within 200 pc, corresponding to a maximum parallax angle of 5 milliarcseconds (mas).

Following the initial compilation of the catalog using the raw data release, online sources, and cross-matching, additional constraints were applied. To begin with, only main sequence stars were selected from all the astronomical sources found. This was done by using a portion of Python code designed to convert the catalog's luminosity class strings into numbers and filter out non-class V bodies. Moreover, extensive improvements in the catalog were wrought about by time-consuming investigation and work with removing bad data sources. Some of the steps included in this process were taking out duplicate sources resulting from finite cross-match radii, eliminating flagged data points, and not using (masking) those stars whose quality was not of the highest category (denoted by AAA in the J, H, and K filters, respectively).

Each step of this process improved the catalog noticeably, and by the end, the GAIA data catalog which was utilized and which will be referred to from here on out consisted solely of main sequence stars closer than 200 pc whose data were singular and of the highest quality attainable. Though it is not directly relevant to the work regarding extinction, an H-R diagram analog, a color-magnitude diagram, of the final catalog used is shown below (Figure 2.1).

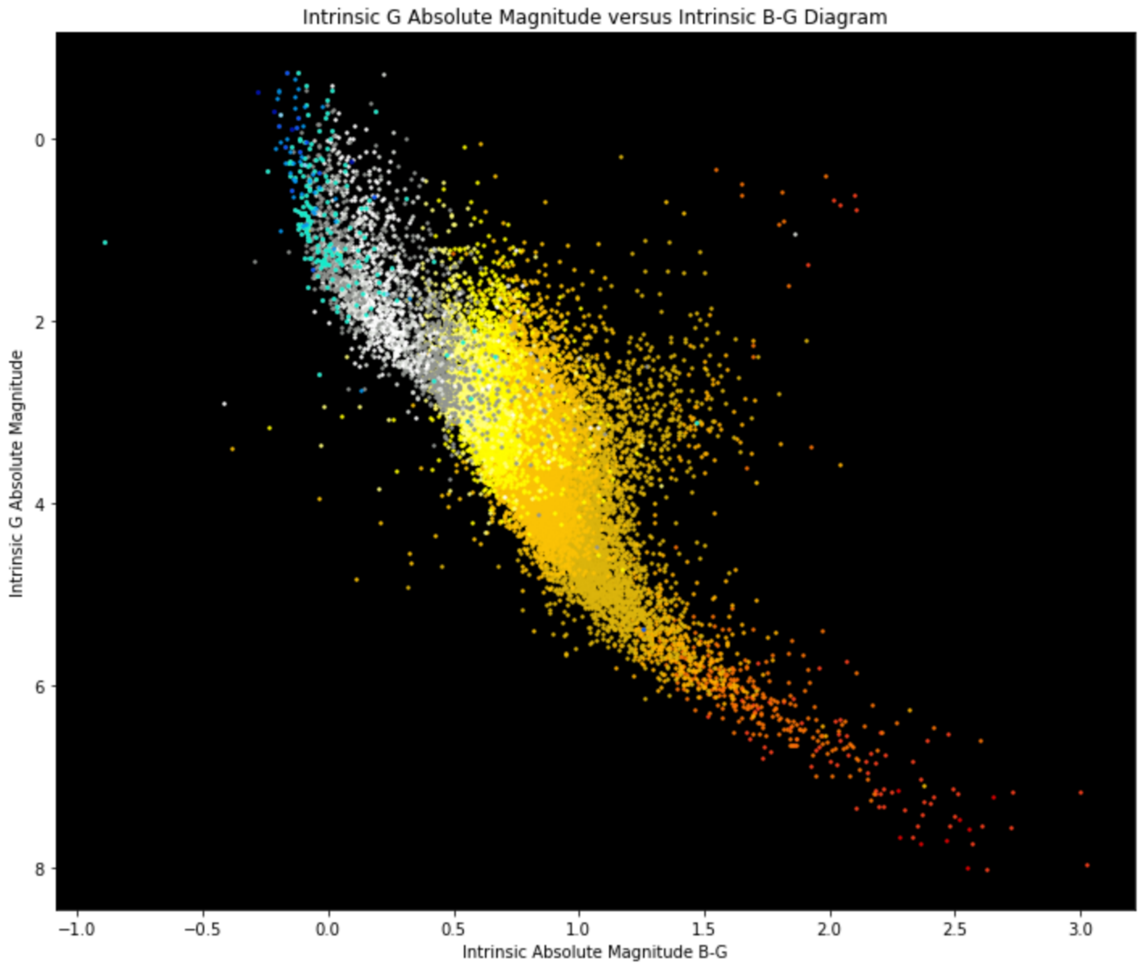


Figure 2.1: Color-magnitude diagram for the GAIA data catalog utilized in the extinction study.

Chapter 3

Matching Stellar Data to SED Templates

A spectral energy distribution (SED) is a plot often used in astronomy which displays brightness, flux, or intensity against wavelength or frequency. SEDs help to characterize a celestial object based on the relative strength at different wavelengths. In the case of the GAIA data catalog, all of the sources studied are main sequence stars, but the relative SEDs vary considerably based on photospheric surface temperature. Relative SEDs can be constructed in multiple ways, including using effective temperatures and color indices (differences in magnitudes through distinct filters). Through the work of Kevin Hall of the GAIA research team, it was found that the best way to build and match to the SED templates was by using colors, measured by differences in magnitudes[4]. A relative SED plot using colors relative to the absolute magnitude in G is shown below (Figure 3.1), where the colors represent temperature corresponding to each template: blue is hottest, while red is coolest. In constructing the relative SED and throughout this study, six wavelengths, matching up with the effective wavelengths of the major filters investigated, were used in detail: B (436.1 nm), V (527.2 nm), G (600 nm), J (1,254 nm), H (1,646 nm), and K (2,149 nm)[5]. It is these six filter wavelengths which are plotted in the figure, where the points for G are all by definition zero (the absolute magnitudes are subtracted by themselves).

As mentioned, the SED templates were defined by their color relative to G. The process for matching each star to its SED template involved conducting a χ^2 analysis between the extinction-corrected theoretical absolute magnitudes (G subtracted) and the observed absolute magnitudes (G subtracted) in the B, V, J, H, and K filters. Mathematically, this χ^2 was computed using the equation

$$\chi^2 = \sum_{i=1}^{5} \frac{[(M_i - G) - (M_i - G)_{th}]^2}{\sigma^2_{M_i - G}}$$

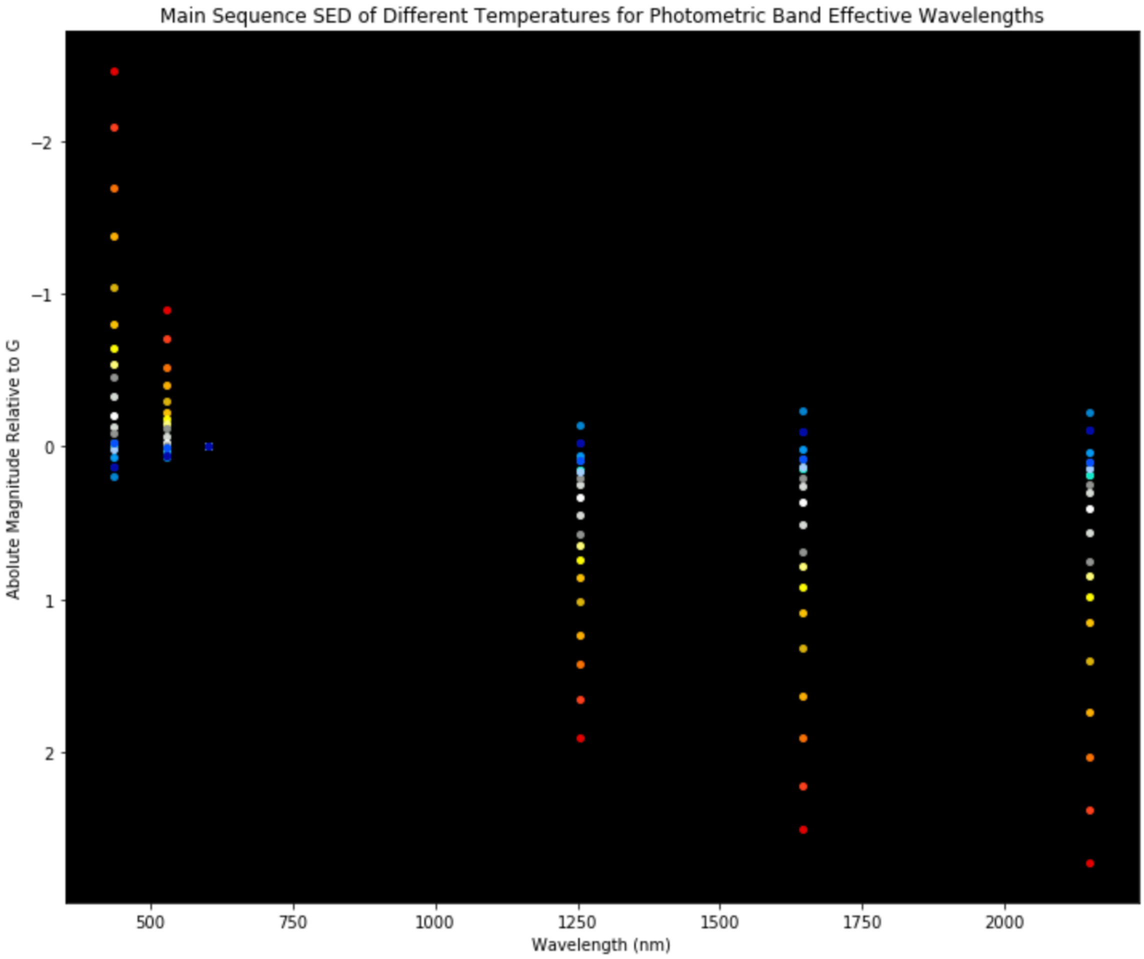


Figure 3.1: Relative SED for color templates, with color defined as absolute magnitude subtracted by absolute G magnitude, plotted against wavelength. Colors indicate relative surface temperature of template stars, with blue hottest and red coolest.

where i represents the filter considered (B, V, J, H, or K), M_i represents the absolute magnitude in that filter, G is the absolute magnitude in G, $(M_i - G)_{th}$ is the extinction-corrected theoretical absolute G-subtracted magnitude in that filter, and σ_{M_i-G} is the standard deviation of the observed absolute G-subtracted magnitude in that filter, used as the uncertainty.

χ^2 was computed for every star and every SED template. Whichever SED template produced the minimum value of χ^2 for that star was taken as the template for said star. Herafter, χ^2 simply refers to the minimum χ^2 for each star; that is, the χ^2 for each star when comparing it with its best SED template. This process was done for every main sequence star in the GAIA data catalog. The results of SED template-matching will not be considered here; what is of greater importance in this study is which level of extinction leads to the best goodness of fit in matching to the templates.

Chapter 4

Determination of Ideal Extinction Level for GAIA Data

4.1 Methodology for Extinction Analysis

When discussing the extinction levels applied in this investigation, models which are linear with increasing distance were used. In particular, the amount of interstellar extinction was defined by the reference V extinction at 200 pc. This quantity simply refers to the number of magnitudes dimmer a star would be observed at compared to its true magnitude if it were located exactly 200 pc away. To clarify, an example is helpful: assuming a star 200 pc away from us would have an apparent V magnitude of 10 with no interstellar extinction, we would observe it to have an apparent V magnitude of 10.2 with extinction present, assuming a reference V extinction at 200 pc of 0.2. In this way, the reference V extinction at 200 pc quantifies the level of interstellar extinction. While this reference value was arbitrarily varied, the levels of extinction in the other filters were dependent on it according to the following equations.

$$A_B = 1.32 A_V$$

$$A_G = 0.8 A_V$$

$$A_J = 0.28 A_V$$

$$A_H = 0.175 A_V$$

$$A_K = 0.111 A_V$$

In these equations, A_V is the extinction in V (linearly defined to have the reference V extinction at 200 pc for a distance of 200 pc), and the other extinctions are multiples of it. From the coefficients, it is clear that extinction is greater for shorter wavelengths (i.e. B) than it is for longer wavelengths (i.e K).

In order to determine the level of extinction which was ideal for the refined GAIA data catalog, the goodness of fit for the template-matching process was used as the criterion. Multiple equally viable methods of quantifying this exist, but the one chosen for this analysis was to find the reference V extinction at 200 pc for which the average of the main sequence stars' χ^2 values obtained a minimum. Essentially, the methodology consisted of finding the level of extinction that, when used as a correction for fitting each catalog star to a template, would produce the best fit between data and theory.

For the variation between the levels of extinction that were tested, a few constraints were applied. To begin with, the lowest amount of extinction which was incorporated into the trials was zero extinction. Although both positive and negative extinction corrections can make mathematical sense in analyzing data, negative extinction lacks physical meaning and therefore was excluded from the analysis. Secondly, the maximum reference V extinction at 200 pc was kept to 0.3 magnitudes. This choice was made to reflect the general notion that the extinction at 1 kpc is approximately 1 magnitude in all directions, which corresponds to an extinction of 0.2 at 200 pc; in addition, and more practically, as will be explained in the next section, the goodness of fit becomes increasingly worse for larger and larger levels of extinction, so it was unnecessary to consider higher values. Finally, the extinction was tested in intervals of 0.01 magnitudes of reference V extinction at 200 pc. The reason for this is twofold: extremely fine determination of the precise ideal extinction is not needed given the eventual results, and the computations require a large amount of time to run, so finer incrementation of extinction would necessitate an exorbitant amount of computational power or time. Based on these constraints, the amounts of reference V extinction at 200 pc were varied between 0 and 0.3 with intervals of 0.01 magnitudes; as such, a total of 31 extinction levels were tested.

With these considerations made, the computations could be run. The way in which the analysis worked is outlined here. First, the extinction correction was applied based on the level being tested, and each template was experimented with. The template which provided the minimum χ^2 was selected for each star. With this done, the average and standard deviation of the stars' minimum χ^2 values were found, which are the final results for analysis. Ultimately, the level of extinction which leads to the lowest average χ^2 in a statistically significant sense will be taken as the ideal amount.

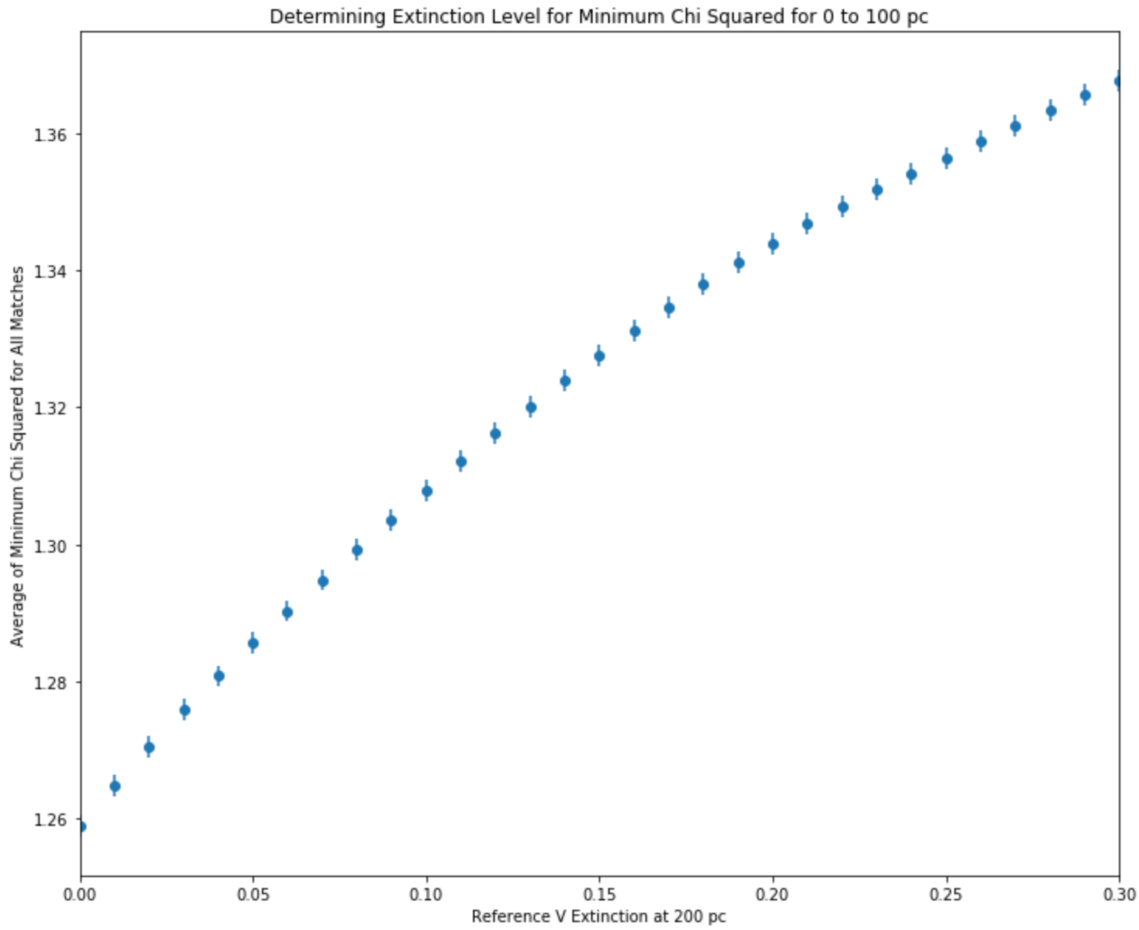
4.2 Findings for Best Template-Matching Goodness of Fit Extinction Level

Upon conducting my analysis of how well the GAIA catalog data match to the templates for varying levels of extinction, based on the reference level of V extinction at a distance of 200 pc, a strikingly clear result emerged. For each distance range (0 to 100 pc, 100 to 150 pc, and 150 to 200 pc), the level of extinction which resulted in the best fit in template matching, quantified by the minimum value of the average χ^2, was zero extinction–in other words, a reference V extinction magnitude correction of 0. Though it may be a surprising conclusion, this assertion is backed up strongly by the results.

To begin with, the same methodology, outlined previously, was used for each distance range, so the results for each case can be directly compared. As it turns out, no comparison between the ranges is necessary whatsoever, as the results provided answers that were certainly sufficient by themselves. For each distance range, the product of my analysis is broken up into two parts. First, a plot is created which graphs the average χ^2 for matching each star in that distance range to the best template; it is important to note that this quantity is not the probability of the distribution but rather is the arithmetic mean of each star's minimum χ^2. In theory, based on statistics, a perfect fit would be indicated by a low average χ^2 value. This is by all means an adequate way to make a judgment of the best extinction level, as the number of degrees of freedom have been taken into account. What's more, the process and degrees of freedom are identical across each trial reference V extinction correction; hence, the minimum average value of χ^2 directly tells which extinction level is ideal. Error bars are included to represent the standard deviation of the χ^2 values of all the matched stars. These allow one to see if the minimum occurs without any overlapping of error bars for other extinction amounts. The reference V extinction at 200 pc which corresponds to the minimum average χ^2 was computed using Python and is displayed just below each plot. For both parts of the product, the conclusion is apparent.

Examining the impact of extinction between 0 and 100 pc, the results for this distance range are shown below (Figure 4.1). It is indisputable from the plot that zero extinction leads to the minimum average value of χ^2, which represents the best fit and hence the most desirable correction. To add on to the certainty of this conclusion, the next highest average χ^2 does not even overlap within the error bars with those for zero extinction. Moreover, the sentence appearing below the graph, showing the final answer from the calculations in Python, agrees with this finding. Thus, between 0 and 100 pc, zero extinction is ideal.

Moving on, the results for the distance range 100 to 150 pc are provided below (Figure 4.2).



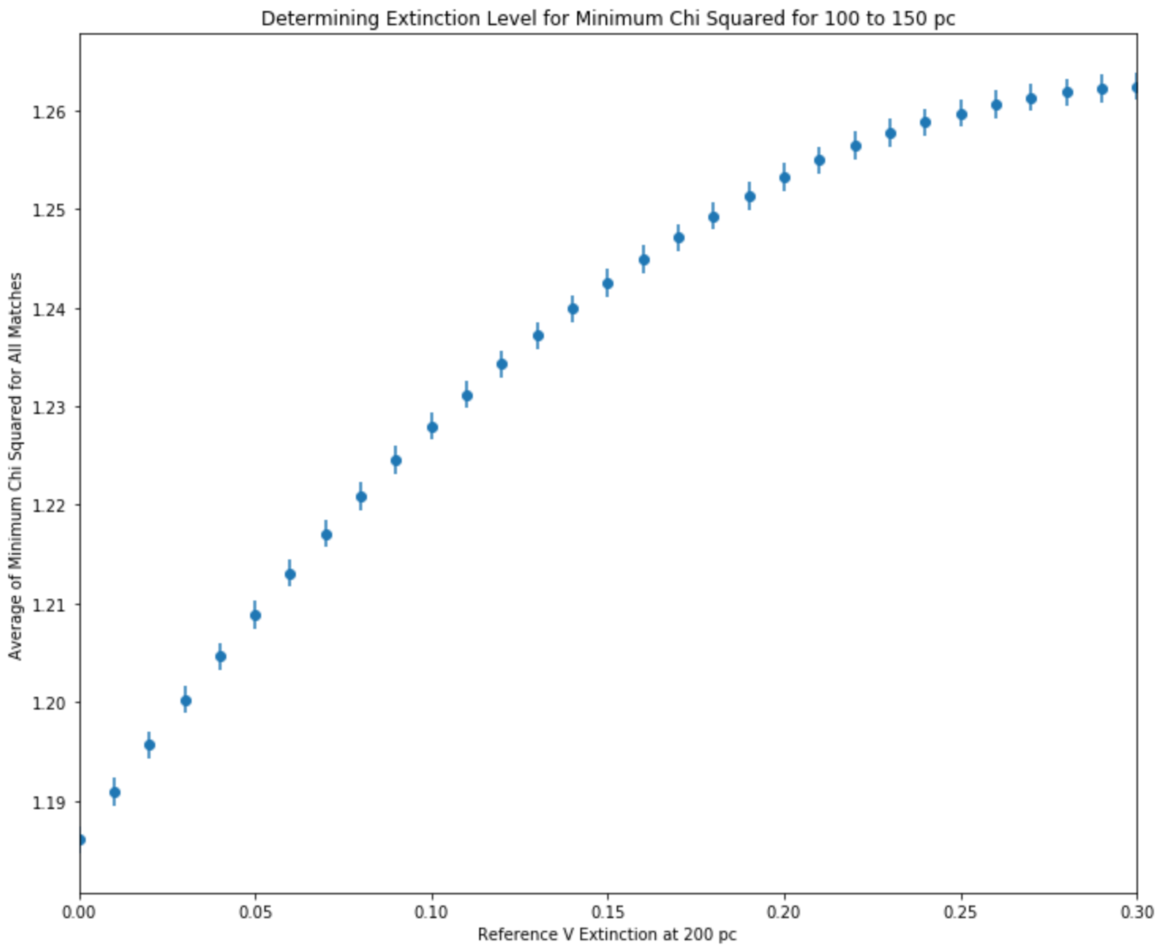

```
For 0 to 100 pc, the best reference V extinction value at 200 pc is about 0.
```

Figure 4.1: Plot of average χ^2 for template matching all GAIA catalog main sequence stars within 0 to 100 pc as a function of the reference V extinction at a distance of 200 pc. Error bars are included, showing the standard deviation in the χ^2 values. As can be seen visually, the ideal level of correction for the catalog is zero extinction, which is also corroborated by the calculations run.

As was found for the closer distance range, zero extinction leads to the minimum average value of χ^2, meaning that the best template fits occur when no extinction is present. The same findings as before also hold true for 100 to 150 pc, providing sound evidence that zero extinction is the ideal level for stars within this range.

Finally, the last subset of stars considered consisted of those between 150 and 200 pc, and the results for this range can be seen below (Figure 4.3). Once again, the best template fitting occurs when zero extinction is considered in making corrections, shown by the minimum in the average value of χ^2 at that reference V extinction. It is true like before that the conclusion to be drawn from these results is definitive and backed up strongly.

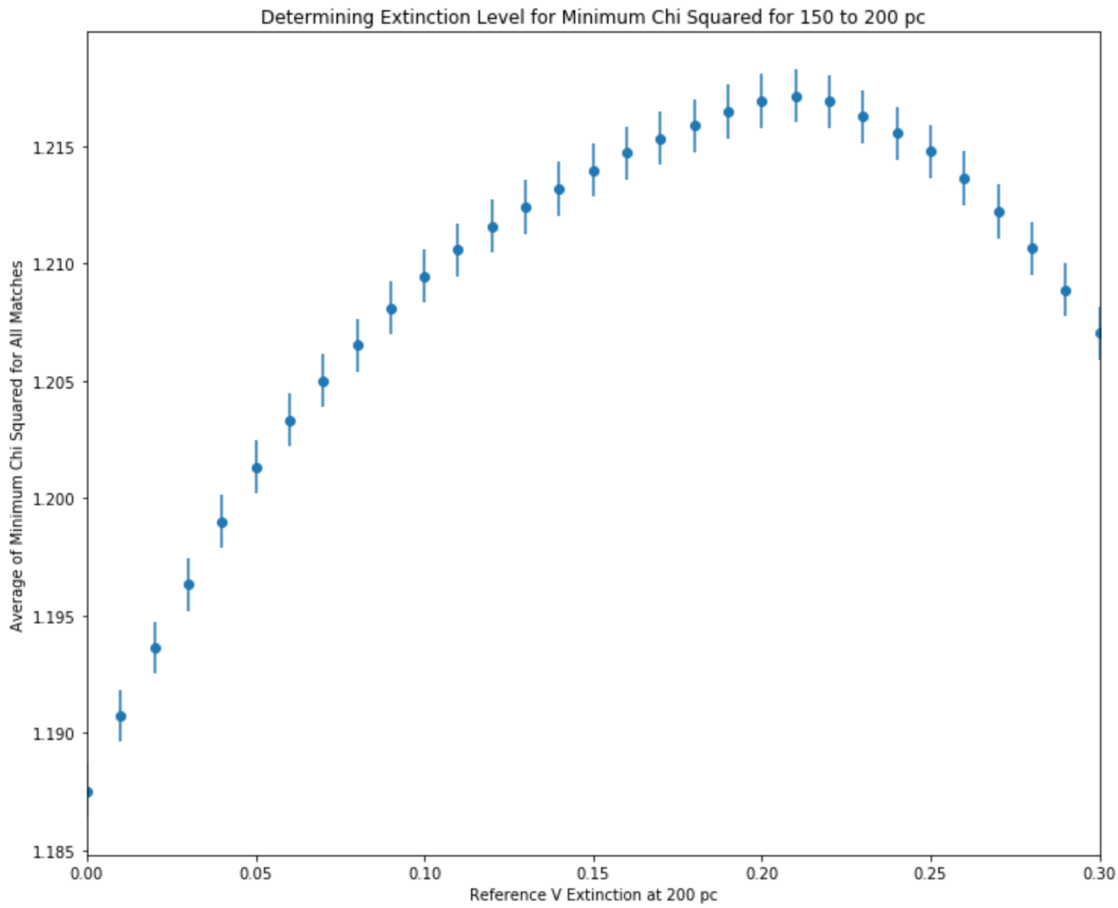
Because it was found convincingly for all three of the distance ranges examined, it is




```
For 100 to 150 pc, the best reference V extinction value at 200 pc is about 0.
```

Figure 4.2: Plot of average χ^2 for template matching all GAIA catalog main sequence stars within 100 to 150 pc as a function of the reference V extinction at a distance of 200 pc. Error bars are included, showing the standard deviation in the χ^2 values. As with the range of 0 to 100 pc, the ideal correction is zero extinction, shown by the plot and the result of calculations.

abundantly clear that zero extinction is the ideal level for working with the GAIA data catalog. Therefore, I can conclude from the results that, in order to obtain the best fit between the stellar data and SED templates, no corrections should be made in matching each star with its best template. Doing so will provide the most accurate and physical representation of the astronomy being studied using the GAIA space mission and, in the interest of good science, is of course the best choice.




```
For 150 to 200 pc, the best reference V extinction value at 200 pc is about 0.
```

Figure 4.3: Plot of average χ^2 for template matching all GAIA catalog main sequence stars within 150 to 200 pc as a function of the reference V extinction at a distance of 200 pc. Error bars are included, showing the standard deviation in the χ^2 values. The best correction is zero extinction, which is apparent from the plot and Python calculation result.

4.3 Discussion of Extinction Analysis

Although the conclusion may superficially seem surprising, the fact that the ideal extinction level for the best goodness of fit when crafting the GAIA templates is zero makes sense. In fact, upon in-depth consideration of the tendencies that exist in the catalog, this result should be the expected finding. The key fact driving this is that a selection effect exists favoring bluer stars at greater distances.

As explained previously, the final dataset from which the extinction was studied consisted exclusively of main sequence (luminosity class V) stars. With main sequence stars, there is an unavoidable trend that brighter stars are more massive and bluer than dimmer ones. As distance from Earth increases, dimmer (redder) stars become less and less likely to be detected due to the inverse-square dependence of light on distance. Thus, for greater distances, greater and greater proportions of the stars we can see are bluer.

Now, in reality, extinction at increasing distance tends to make celestial objects appear dimmer and redder than they truly are, represented by positive extinction. However, in the catalog utilized, increasing distance causes the celestial objects we pick up to be bluer than should be found due to the aforementioned selection effect. So, this bias in the GAIA data causes the ideal correction for the template matching, the one which has the best goodness of fit to the empirical data, to be the lowest amount of extinction, which leads to zero extinction being the answer. In fact, though it lacks physical meaning, the ideal reference V extinction at 200 pc (which minimizes average χ^2) is likely a *negative* number, which would account for the bluer-at-greater-distance selection effect.

This ultimate result is quite interesting, because it has two important implications. First, it means that, if we eventually want to observe the general population of stars that represents what we know to be true of stellar formation trends (i.e. small, red stars are much more plentiful than massive, blue stars), our telescopes need to improve significantly. This study only concerned stars up to a distance of 200 pc, which is minute even in the scheme of the Milky Way Galaxy. As a result, our technology must take leaps forward before detailed astronomy of the rest of the galaxy can be considered.

The other main implication is that, in theories extrapolated from the GAIA mission's data, models that use negative extinction corrections could be quite successful. Given the strong evidence showing that zero extinction is best for this catalog, as well as the slope of each plot at zero, it is not unreasonable to believe that theories incorporating negative "extinction" could match what we currently see very well. This fact might be a major revelation for the second large data release of the GAIA mission in the coming spring as astronomers attempt to create adequate templates and make models for studying the rest of the galaxy and should therefore be kept in mind moving forward.

Chapter 5

Conclusion

As has been illustrated beyond a reasonable doubt, the impact of extinction on the templates for the GAIA data catalog is the opposite of what it should traditionally be. The catalog consists of main sequence stars that lie within 200 pc of Earth, and the templates being matched with correspond to temperature ranges on an SED which are 500 K wide, beginning with 3,500 K and ending with 14,000 K. With such conditions, the ultimate conclusion for every distance range (0 to 100 pc, 100 to 150 pc, and 150 to 200 pc) is that zero extinction is the ideal level, as measured by the minimized average value of χ^2) in the matching process. This means that the best scenario, the one with the greatest goodness of fit to the observed data, has no extinction correction whatsoever.

Further research is recommended trying out models that incorporate negative "extinction" corrections and testing these models against the data from the GAIA space mission. In addition, studies of the impact of extinction in the catalog past 200 pc could reveal surprising answers and should consequently be undertaken. The second, major data release from GAIA is expected to be made available in the coming spring, and much work will be able to be done with this exciting collection of information. Understanding extinction's impact, or lack thereof, is a crucial step towards working with the data to study stellar locations and distribution of interstellar dust in the broad region surrounding Earth. Hopefully, it will serve as the first step in the journey towards creating a detailed and comprehensive three-dimensional map of the Milky Way Galaxy.

References

[1] P. Tanga, "Gaia," About Science and Technology, European Space Agency, December 12, 2017.

[2] D. Michalik *et al.*, "The *Tycho-Gaia* Astrometric Solution," A and A 574, Astronomy and Astrophysics, December 24, 2014.

[3] "CDS Portal," Centre de Donnees Astronomiques de Strasbourg, UDS, December 9, 2017.

[4] K. Hall, "Final Report," GAIA Research Project Team, UMD Department of Astronomy, December 2017.

[5] O'Connell, "Magnitude and Color Systems," Department of Astronomy, California Institute of Technology.

Acknowledgments

In concluding this investigation of extinction in the data from the GAIA space mission, I would like to acknowledge some of those who helped me with my research and education leading up to this project. First and foremost, I sincerely recognize and thank Dr. Lee Mundy and Dr. Isabelle Joncour of the Department of Astronomy at the University of Maryland, College Park for their mentorship, assistance each step of the way, and guidance in conducting research on the GAIA team. Their help has been instrumental throughout the whole semester and will certainly lead the way in future work on the subject. In addition, I would also like to acknowledge and thank Dr. Melissa Hayes-Gehrke, who has been extremely knowledgeable as a professor, helpful as my advisor, and guiding as a general mentor in the field, and Dr. Derek Richardson, whose instruction in astronomy illuminated many topics, provided strong background, and showed the fun side of research. These individuals have helped me get involved in astronomy research and made me enthusiastic about the prospects of working in the field as a career.