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

INVESTIGATING THREE-DIMENSIONAL STRUCTURE OF INTERSTELLAR CLOUDS

Using Nearest Neighbor Distance Methods and 3D Visualizations for Ophiuchus

May 18, 2018

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 (ESA), is currently in the process of collecting and disseminating unprecedented amounts of useful data on over one billion stars within the Milky Way with the hopes of constructing a high-fidelity map of our home galaxy. The first data release (DR 1) contained limited information, but the second data release (DR 2) on April 25, 2018 brought with it the full set of data so far gathered with vastly improved breadth and accuracy. In the long-term process of using the GAIA mission to map the Milky Way Galaxy, a crucial step is determining, mapping, and understanding the structure of interstellar clouds which obscure celestial views and cause extinction in observations. In this study, the location, size, and morphology of interstellar clouds in the region of the sky contained in the zodiacal constellation Ophiuchus were explored by first employing Python to use the KD Tree algorithm as a tool for nearest neighbor methods. With this in hand, various approaches, including novel functions, were applied to represent the results in three dimensions to display the appearance of interstellar clouds as well as compare findings to two-dimensional representations. The mappings and visualizations were done for both celestial coordinates (right ascension, declination, and distance) and, to a lesser degree, the realistic picture consisting of rectangular coordinates centered around the Earth. Upon doing this study, it can be concluded that nearest neighbor methods are a useful analog to extinction for approximately determining interstellar cloud location and structure. Moreover, the two-dimensional and three-dimensional visualizations discussed offer valuable insight into studying clouds and, therefore, should prove useful for mapping the Milky Way Galaxy, although this research should be continued to better understand how accurate these representations of interstellar clouds are.

Contents

1 Introduction **2**

2 Two-dimensional Nearest Neighbor Distance Mapping **4**

3 Three-dimensional Celestial Coordinate Nearest Neighbor Mapping **8**

4 Three-dimensional Realistic Coordinate Nearest Neighbor Mapping **17**

5 Conclusion **21**

References **22**

Acknowledgments **23**

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 vital aspect of this mapping work, naturally, consists of understanding, locating, and tracing interstellar clouds which block observations into deep space but are an integral part of our galaxy. In general, the presence of interstellar clouds tends to block the ability to detect and study stars in the Milky Way and, even more so, beyond our galaxy.

The work discussed here is based around the fact that, as interstellar cloud density rises, our capacity to observe stars decreases, which has the implication that clouds tend to block any detection of stars within or past them. Logically, then, the assumption can be made that, since the galaxy tends to have uniformly distributed stars on a grand enough scale, the regions of the sky where few or no stars are seen contain clouds. Therefore, the methodology utilized relies on the fact that greater distance between stars, or lower density, implies interstellar clouds.

It is from this essential idea that the nearest neighbor approach gains traction. This method extends the idea of finding a star's distance to its closest neighboring star to greater numbers of stars; in other words, for any number k, the method computes the distance to a star's kth closest neighboring star. In this way, larger nearest neighbor distance indicates lower density of stars and therefore stronger density of interstellar clouds present. To cover the entire sky for a full picture and avoid clustering of results, the nearest neighbor method was applied to uniformly spaced pixels in the three-dimensional space occupied by the stars. In particular, the region of the sky chosen to study is the constellation Ophiuchus, which lies below the celestial equator and is a part of the thirteen zodiacal constellations, with the GAIA DR 2 data obtained through the ESA's GAIA archive online reference source[2]. Within this set of data, there are a total of 336,739 stars, where 249,711 (approximately 74 percent) stars which have parallaxes provided were utilized because this allowed for simple and accurate computation of their distances from Earth. The stars ranged in right ascension (RA) from about 245.5 to 248.5 degrees and in declination from about -25.5 to -21.5 degrees. The work done and results for Ophiuchus can be easily extended to any other region of the sky, providing a key first step towards eventually achieving the end goals of GAIA's venture to map the galaxy.

Chapter 2

Two-dimensional Nearest Neighbor Distance Mapping

Once the Ophiuchus data were obtained, some winnowing down occurred to ensure the highest possible quality of data going into the results. As mentioned, only those stars which had parallax, approximately 250,000, were utilized so that distances could be computed to them in later steps. To add on to this, an additional constraint was enforced, where only stars which had nonnegative parallax measurements were kept to avoid involved data processing and to ensure that all distance values were positive, and where the stars must not be duplicate sources. No restrictions were placed on distance or any other parameters in GAIA's release. In all, the final data set included 179,237 stars, or approximately 53 percent of the total amount in the original Ophiuchus data. This refined data set was the only collection of data used throughout the entire process of examining Ophiuchus and its interstellar clouds.

To begin with, for a full treatment, the methods of this research were first applied exclusively to two dimensions, right ascension and declination, both because these coordinates are the most accurately known and because such work gives a highly intuitive representation of looking out at the night sky. A visualization of the region of the sky, as if looking out into the night, is shown in Figure 2.1. Note the distinct areas with an absence of stars or significantly lower density and their pattern.

Throughout the process, the methodology used was to create 100 linearly spaced pixels in both dimensions between their respective minima and maxima. Following this step, both the pixels and the star data were shifted to be centered around zero and normalized to eliminate different weights derived from the varying numbers of right ascension and declination (and later distance). Throughout the process of this work, the choice was made to use k=20 (select for the 20th nearest neighboring star), as the 20th neighbor was large enough to overcome

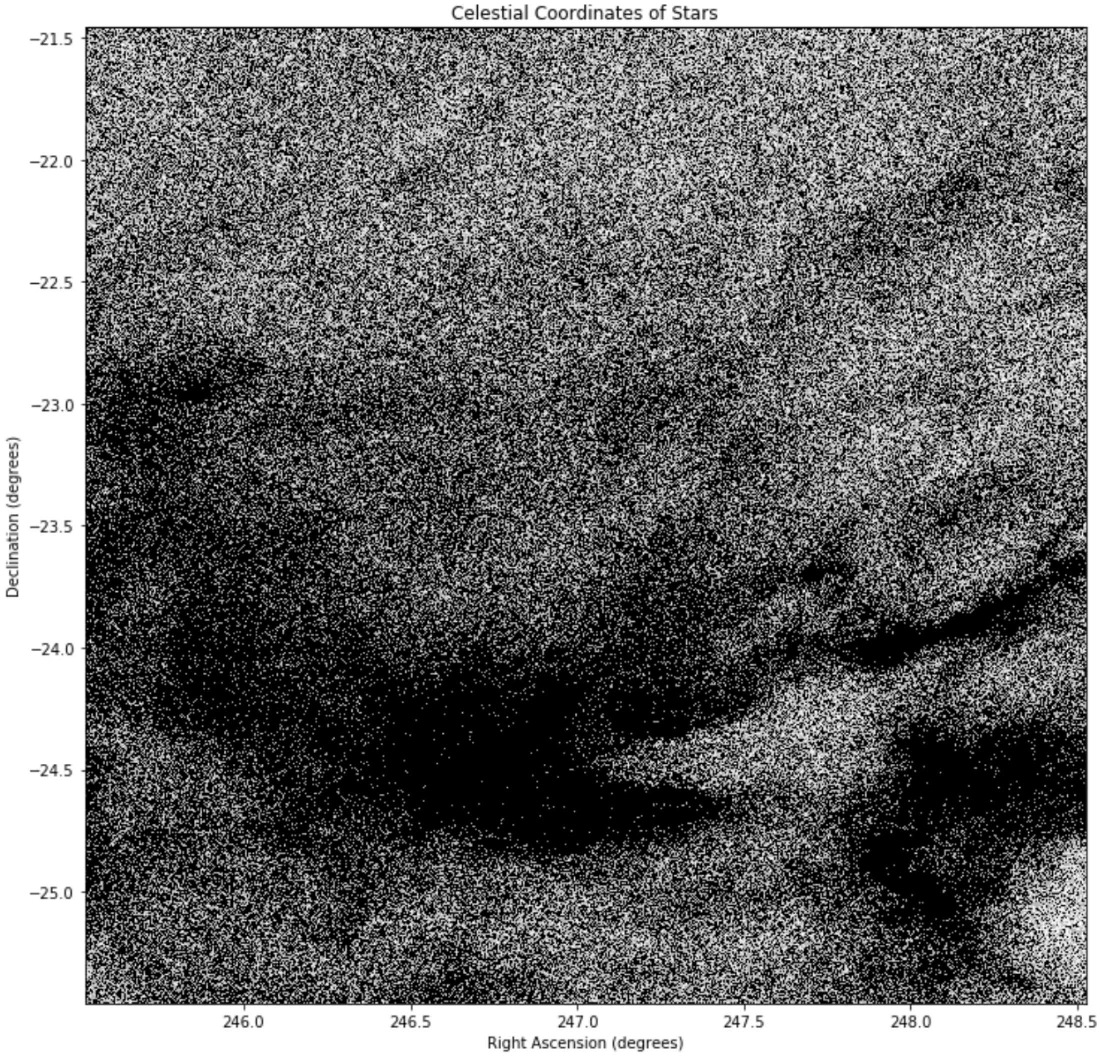

Figure 2.1: A plot of the two-dimensional locations of the stars, with stars plotted in white against a black background to give the impression of looking out into the night sky.

random variations in smaller-scale star patterns while still low enough to calculate in a reasonable amount of time. The KD Tree algorithm, a nearest neighbor method specialized for efficiently dealing with large datasets, was exploited in running computations in Python to speed up the otherwise laborious process. Then, the resulting nearest neighbor distances were plotted as a color mapping on top of a right ascension and declination mesh, as shown in Figure 2.2. In this plot, notice the striking similarity to the basic star locations plot. The areas with no stars appear brightest with the highest 20th nearest neighbor distance, and, generally, the spots with lower density show up clearly on the nearest neighbor mapping. This initial two-dimensional success validates the efficacy of using this method, and three-dimensional findings do so even more.

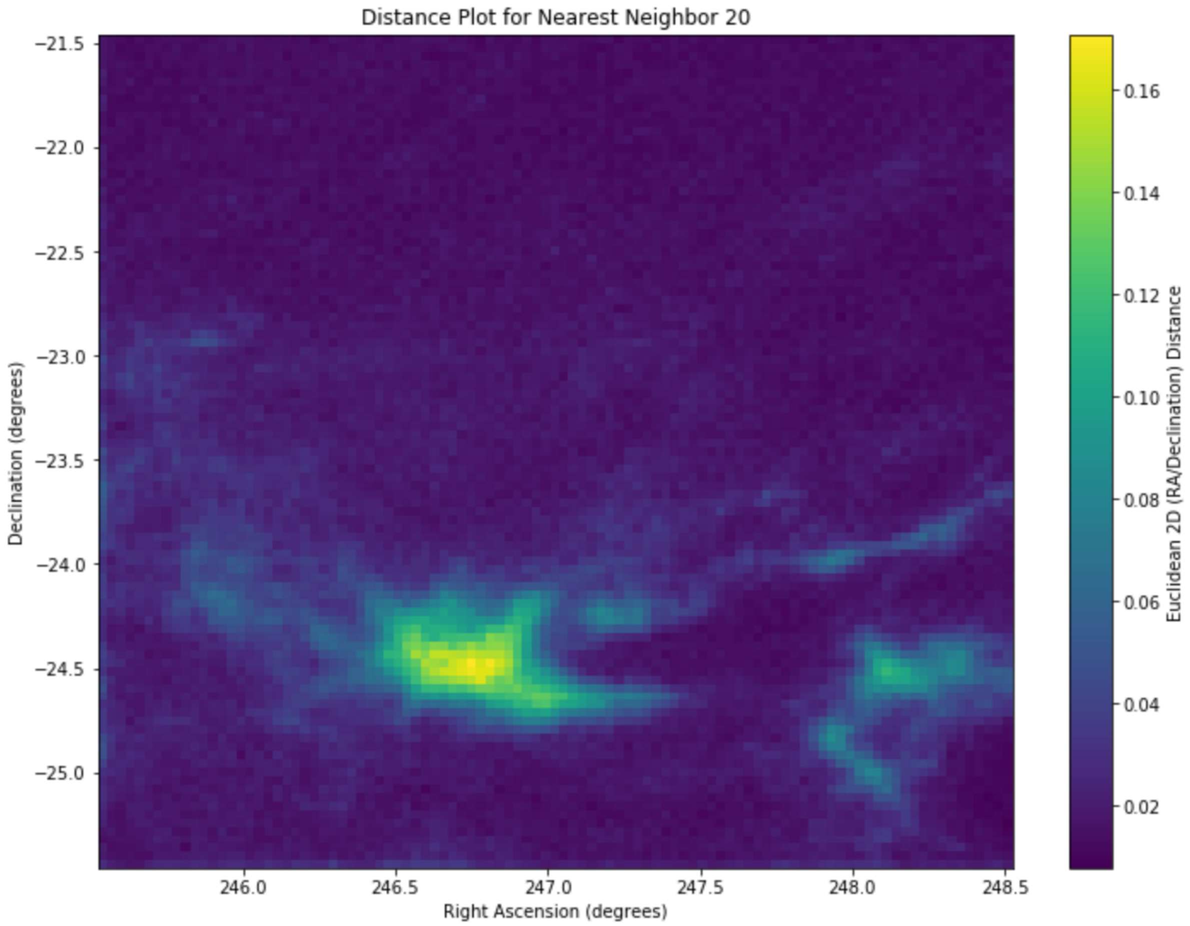

Figure 2.2: Two-dimensional 20th nearest neighbor distance mapping, where the color bar indicates distance according to the scale shown.

Following the construction of the two-dimensional distance map, another check on the results was to create a contour plot, which is shown below in Figure 2.3. Like a topographic elevation map, this plot displays the greatest nearest neighbor distance as the brightest and highest level on the color bar and also shows the gradient of change through the proximity of the contour lines. Once again, this graph shows an excellent match to the star locations plot, confirming the validity of the nearest neighbor method for displaying regions of lower star density and therefore interstellar clouds. In the next section, these methods and visualizations will be extended to three dimensions.

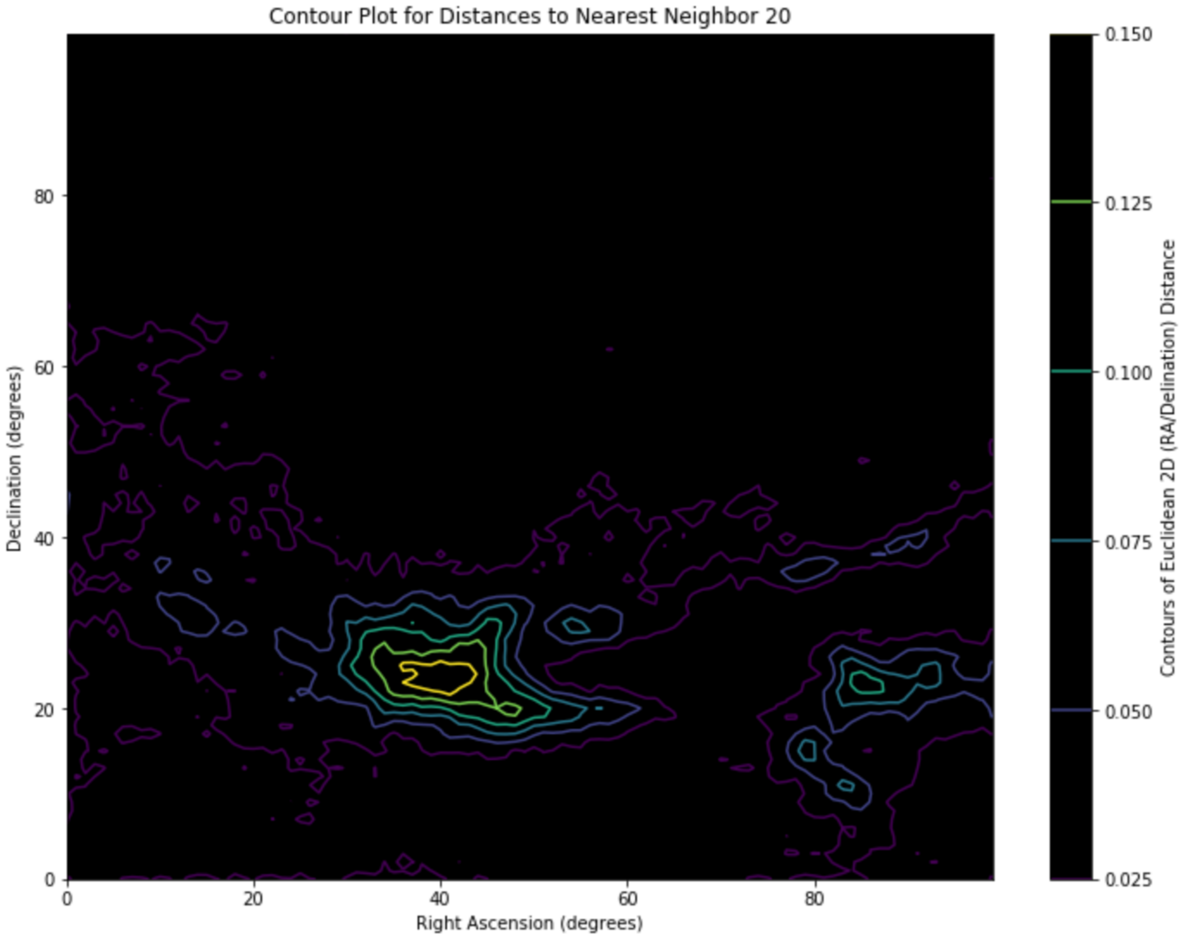

Figure 2.3: Two-dimensional 20th nearest neighbor distance contour plot, with the contour levels shown on the color bar.

Chapter 3

Three-dimensional Celestial Coordinate Nearest Neighbor Mapping

In bringing these ideas into three dimensions, distance away from Earth is the new component, where right ascension and declination are both coordinates which concern where on our view of the sky an object is. It is important to note a major assumption which is made throughout the rest of this work: that there is negligible extinction and no interstellar clouds within 200 pc of Earth. Based on a wealth of findings, this is already a very good assumption; additionally, when conducting nearest neighbor distance tests, it was repeatedly discovered that the distance was vastly skewed to have the greatest nearest neighbor distances at small distances. This, of course, makes sense, since there is naturally a much lower density of stars close to Earth than far away, following the fact that the volume enclosed by a sphere around Earth increases more rapidly than the surface area of such a sphere with increasing radius. Therefore, including distances less than 200 pc served no purpose other than to bias color bars and results excessively. For this reason, these distances are ignored, both for good reason and based on sound science. In this study, the maximum distance examined was 1,000 pc, as the work sought to determine cloud location and structure relatively near Earth.

Having said that, the same exact methodology was followed as for two dimensions except that the distance component (with pixels uniformly spaced between 200 and 1,000 pc) was included and the KD Tree algorithm was applied to three dimensions rather than two. The 20th nearest neighbor distance then gives the distance in celestial coordinate space instead of on the sky. Once the nearest neighbor distances were calculated, they were shown in three dimensions as a color plot, shown below in Figure 3.1. On this plot, and others after it, the distance ranges from 0 to 1,000 pc to give perspective, but the data plotted begin at only 200 pc. In the plot, there appears to be a fringe effect of greater nearest neighbor distance at 200 pc, which indicates the severity to which near distances had skewed the results; however, this is artificial and only indicates the lower density of stars closer than 200 pc rather than actual cloud structure. Where there are distinct regions of greater nearest neighbor distance, it appears that they show up as columns. This, too, makes sense, as the presence of a cloud obscuring our view of stars would certainly block observations of stars even farther away; thus, once a cloud has intervened, it will show up at all greater distances.

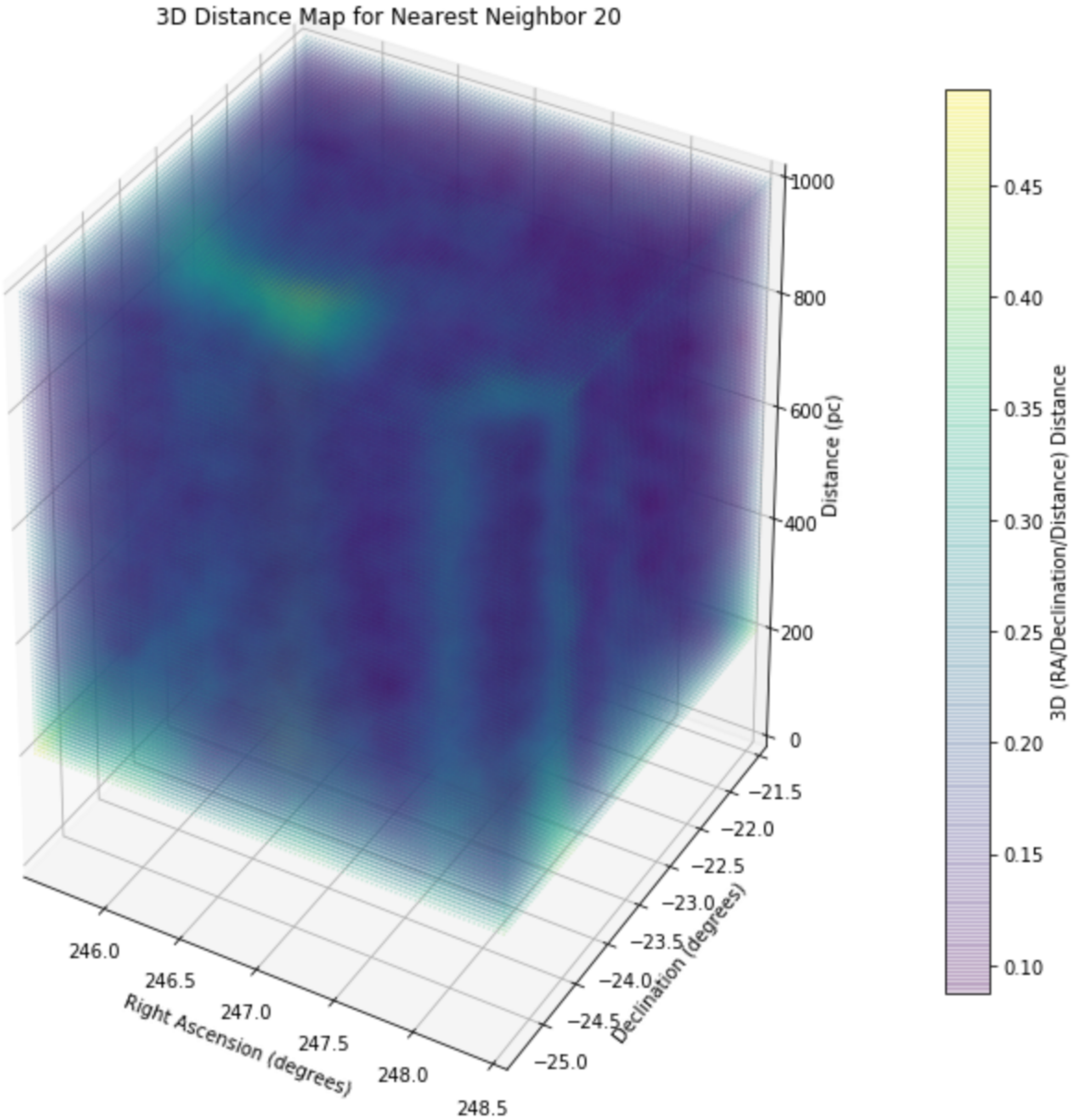

Figure 3.1: Three-dimensional 20th nearest neighbor distance mapping, where the nearest neighbor distance is represented by the colors according to the color bar.

Most importantly, the plot shows that, if viewed from above, the areas with distinctly higher nearest neighbor distances (and their related vertical columns) math up exceptionally with the two-dimensional results! This result directly affirms the reliability of the translation of the nearest neighbor method into three dimensions and allows for a view of the structure of interstellar clouds. Shown additionally below in Figure 3.2 is the same exact plot, but with the locations of stars included in green for an interesting comparison.

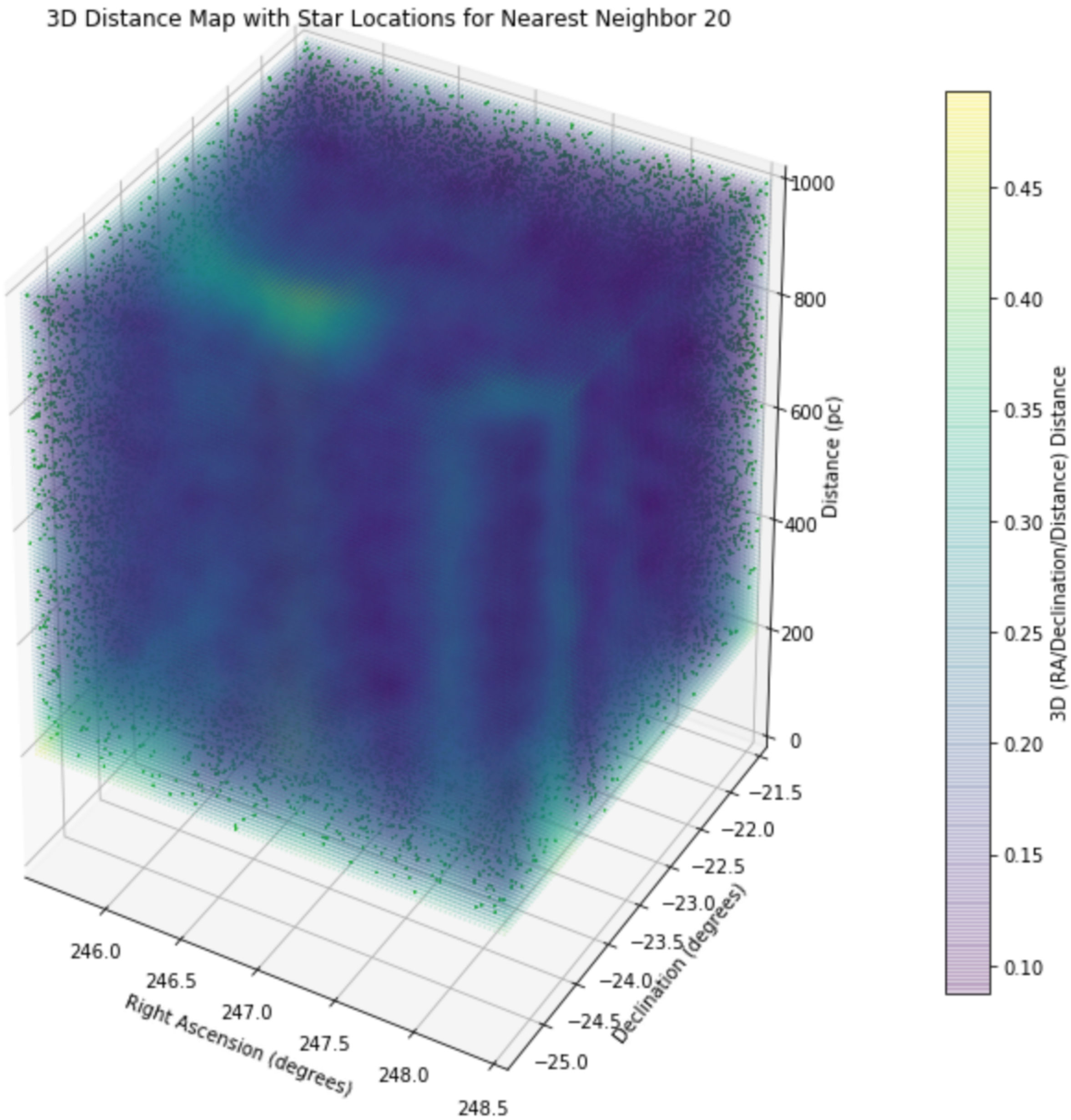

Figure 3.2: Three-dimensional 20th nearest neighbor distance mapping, where the nearest neighbor distance is represented by the colors according to the color bar, with star locations in green.

Moreover, a custom three-dimensional contour plotting function was used to display structure of clouds. In order to do this, a new vector of colors was constructed to represent the standard "viridis" color mapping in Python, which was used in the previous plots. This consisted of choosing colors that corresponded to that color mapping and placing them in an appropriate order. In total, 25 colors were selected to be used with 25 contour levels in plots, and the order of the colors is shown below in Figure 3.3, where the plot is meaningless and is simply designed to show the colors on the line y=x.

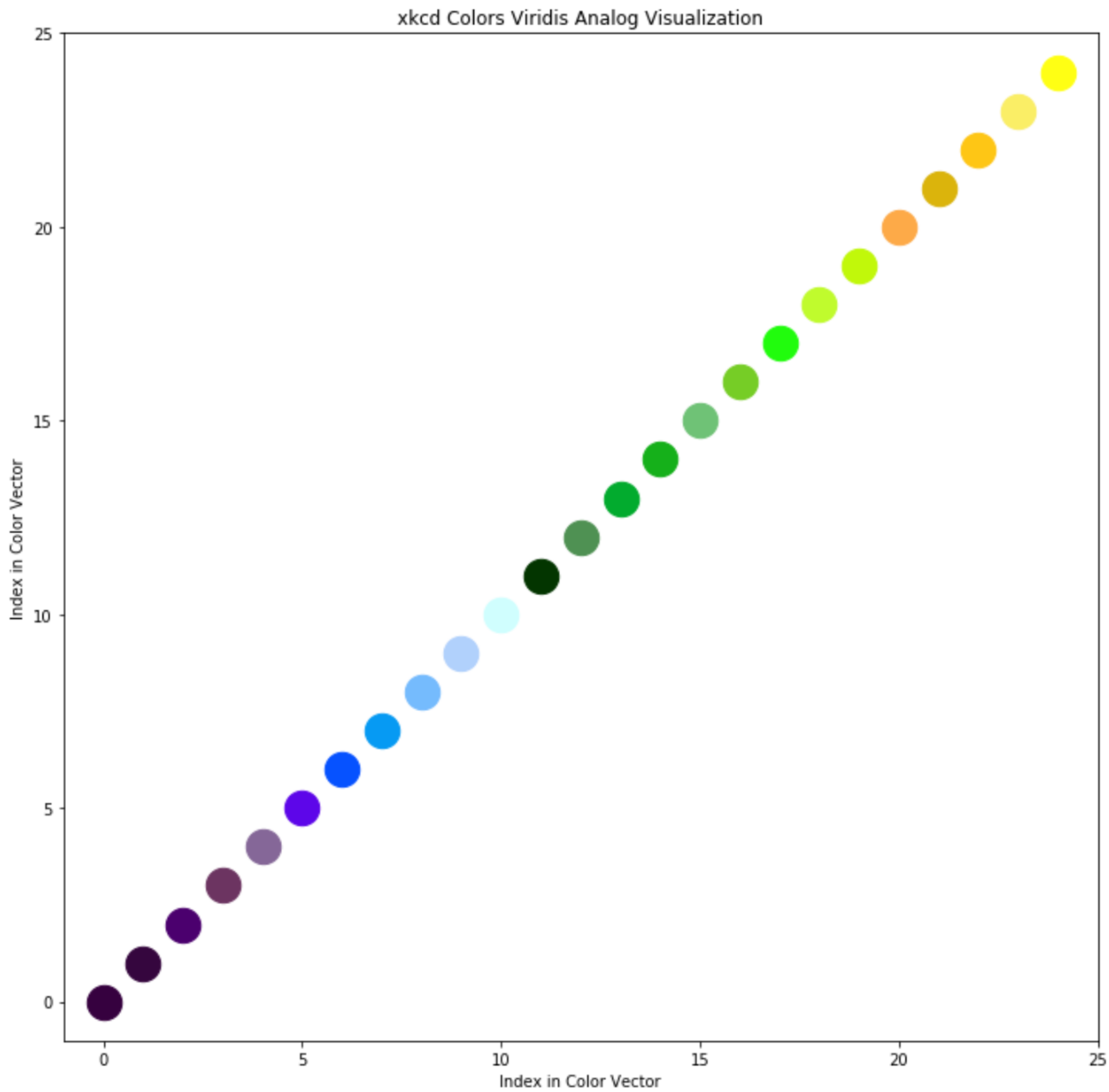

Figure 3.3: Display of the colors and order of the color vector utilized to represent Python's viridis color mapping.

The function works quite simply by dividing the nearest neighbor distances into equally spaced contour levels and parsing the pixel data to select those whose nearest neighbor distances lie between the particular contours. Then, each level of contours gets specified a color in the color vector, and those pixel data are plotted. In this way, the function creates a three-dimensional, layered contour graph that shows the structure and gradients of the clouds. As examples, in the two graphs below, Figure 3.4 and Figure 3.5, contours 1-2 and then contours 1-5 are plotted, showing increased nearest neighbor distance and volume.

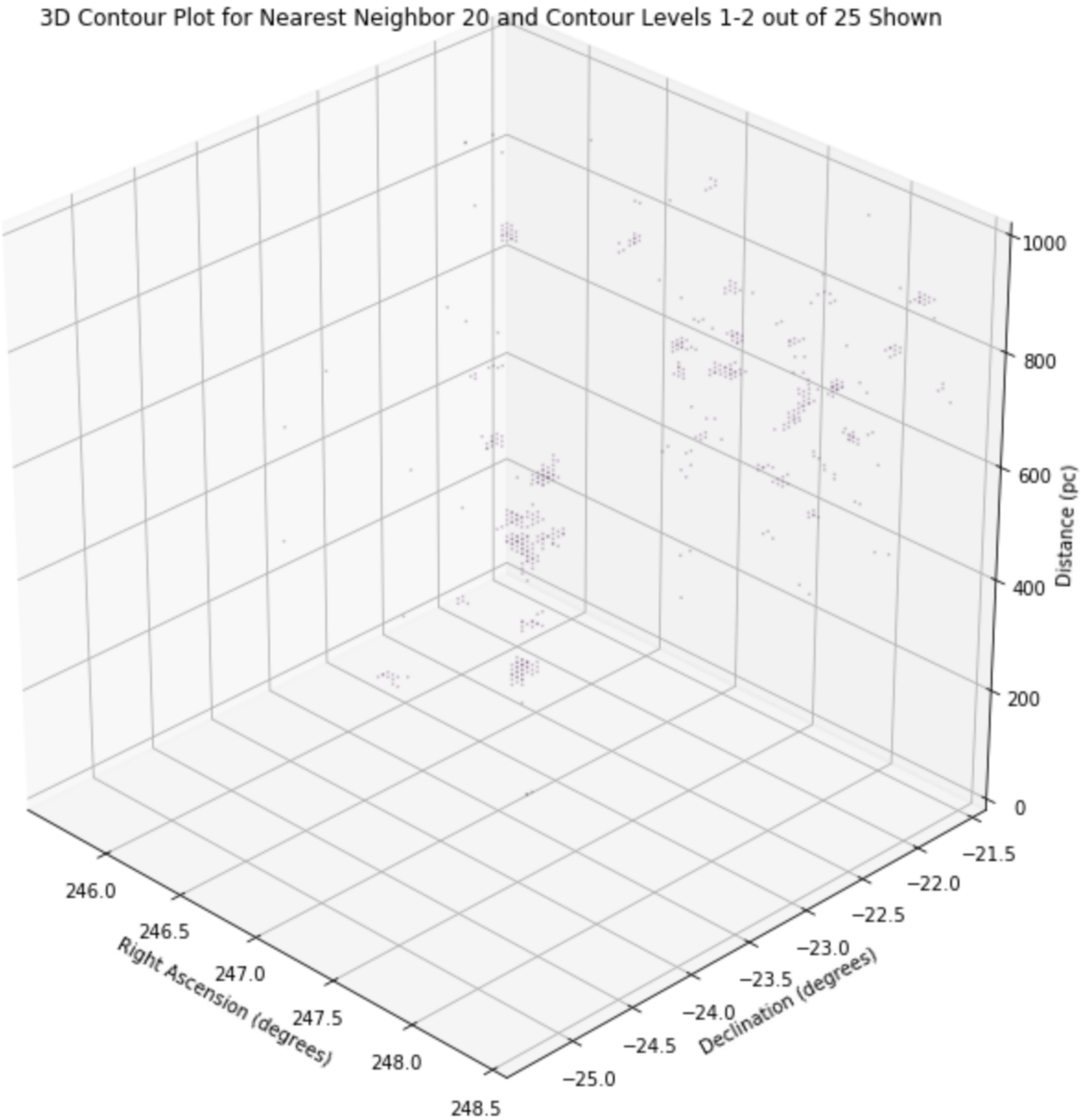

Figure 3.4: Three-dimensional contour results for contours 1-2 out of 25 of the 20th nearest neighbor distance mapping.

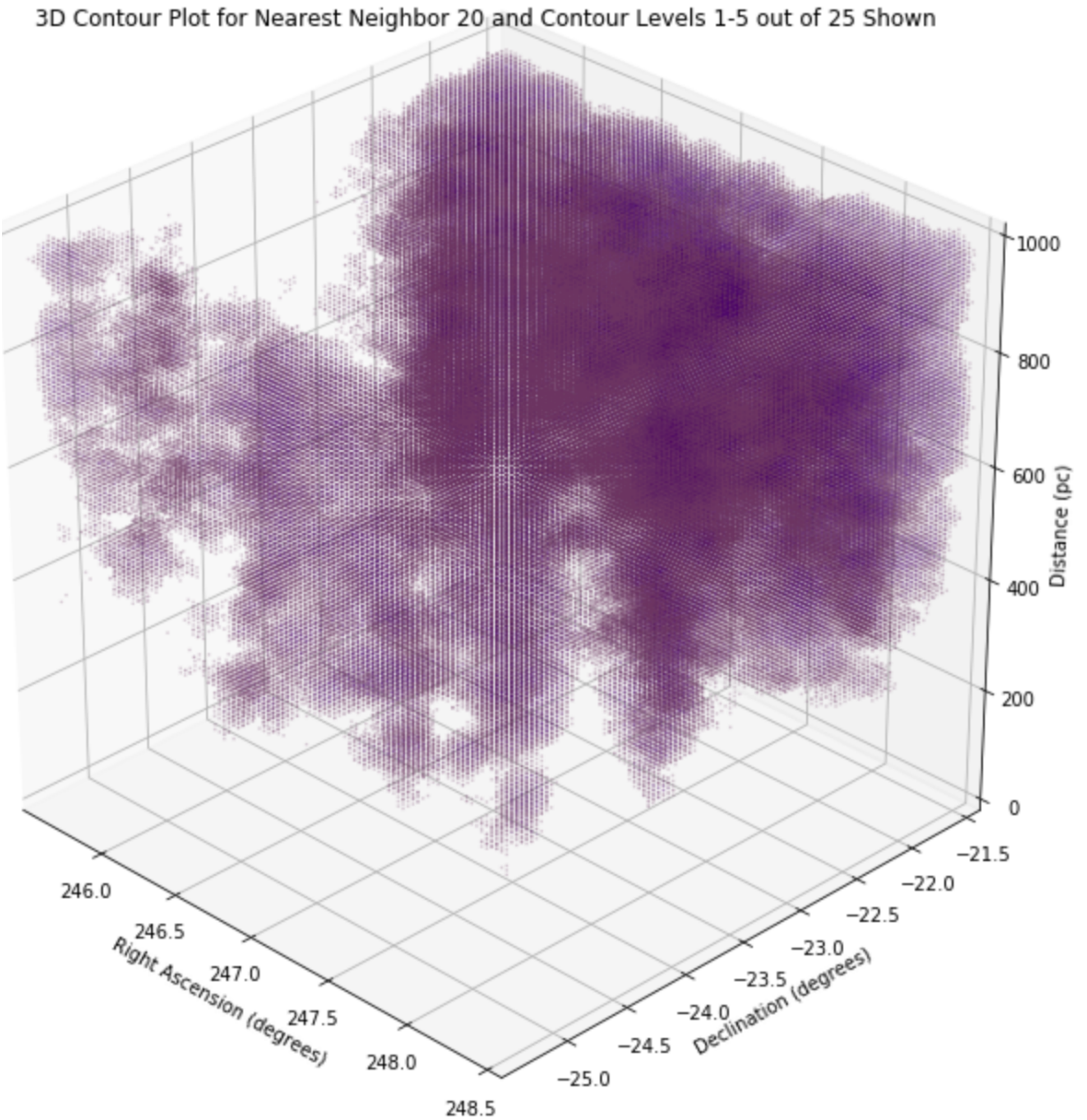

Figure 3.5: Three-dimensional contour results for contours 1-5 out of 25 of the 20th nearest neighbor distance mapping.

Now, looking at the low numbers of contour levels is not very instructive, as this simply displays regions of low stellar density and therefore absence of clouds. However, this function provides great insight into cloud location and structure for the higher contour numbers. Below, in Figure 3.6, contour levels 8-25 (out of 25) are plotted, and the results are fascinating. Once again, there is clear fringing at the bottom of the plot, an artifact of the low stellar density close to Earth and not a cloud-related effect. However, there are distinct cloud regions shaped like columns that emerge to the maximum distance. What's more, it can be observed that, if viewed from the top of the plot, these cloud columns would coincide with the patterns of the 3D and 2D distance mappings. This once again offers great confirmation that this methodology is effective for finding regions of low stellar density and determining the structure of clouds.

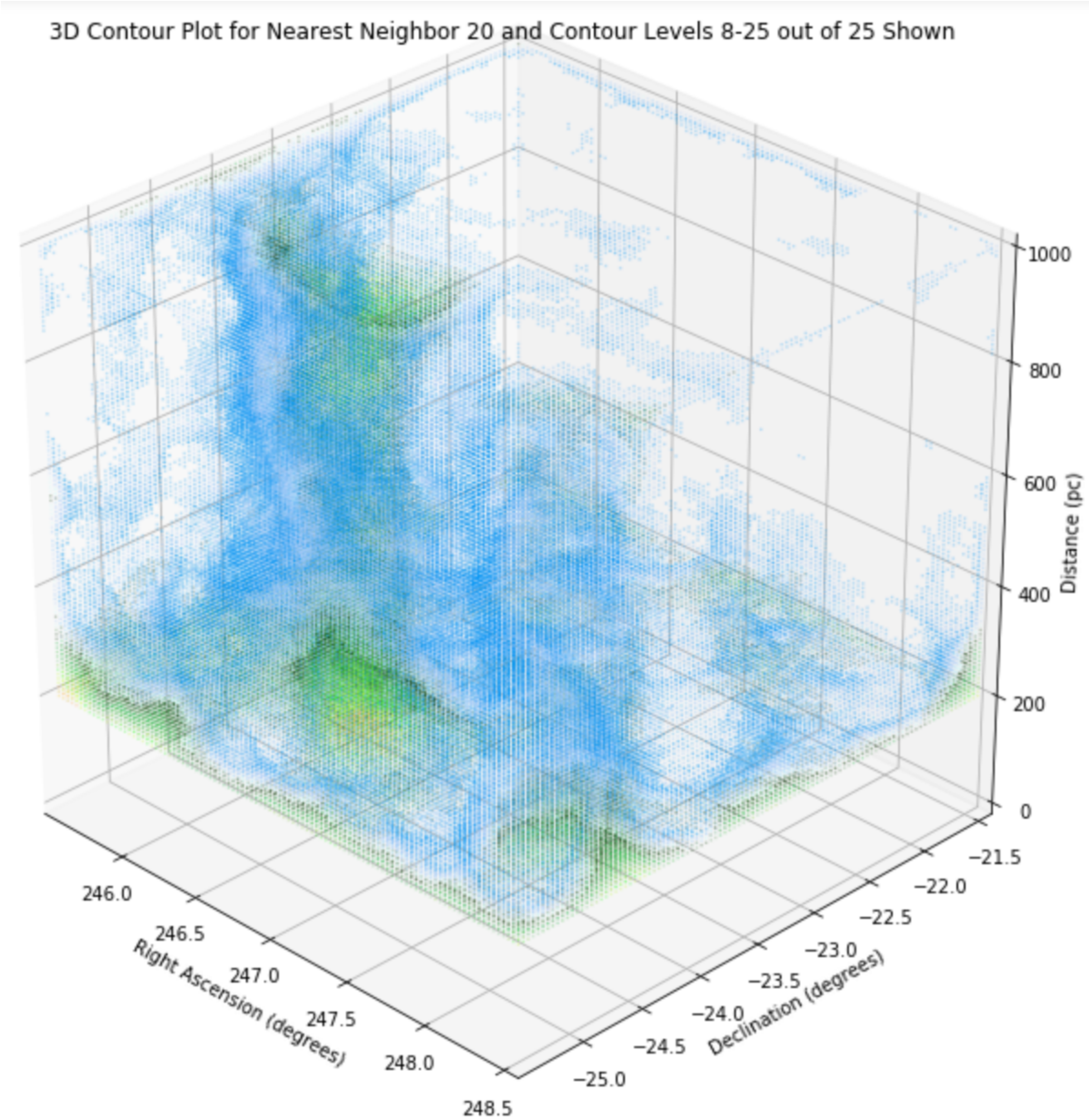

Figure 3.6: Three-dimensional contour results for contours 8-25 out of 25 of the 20th nearest neighbor distance mapping.

To provide further convincing evidence of the correspondence between the three-dimensional contour picture and the two-dimensional mapping, in Figure 3.7, the same result, the contour plot for contour levels 8-25, is shown, but the perspective is altered to be a top view.

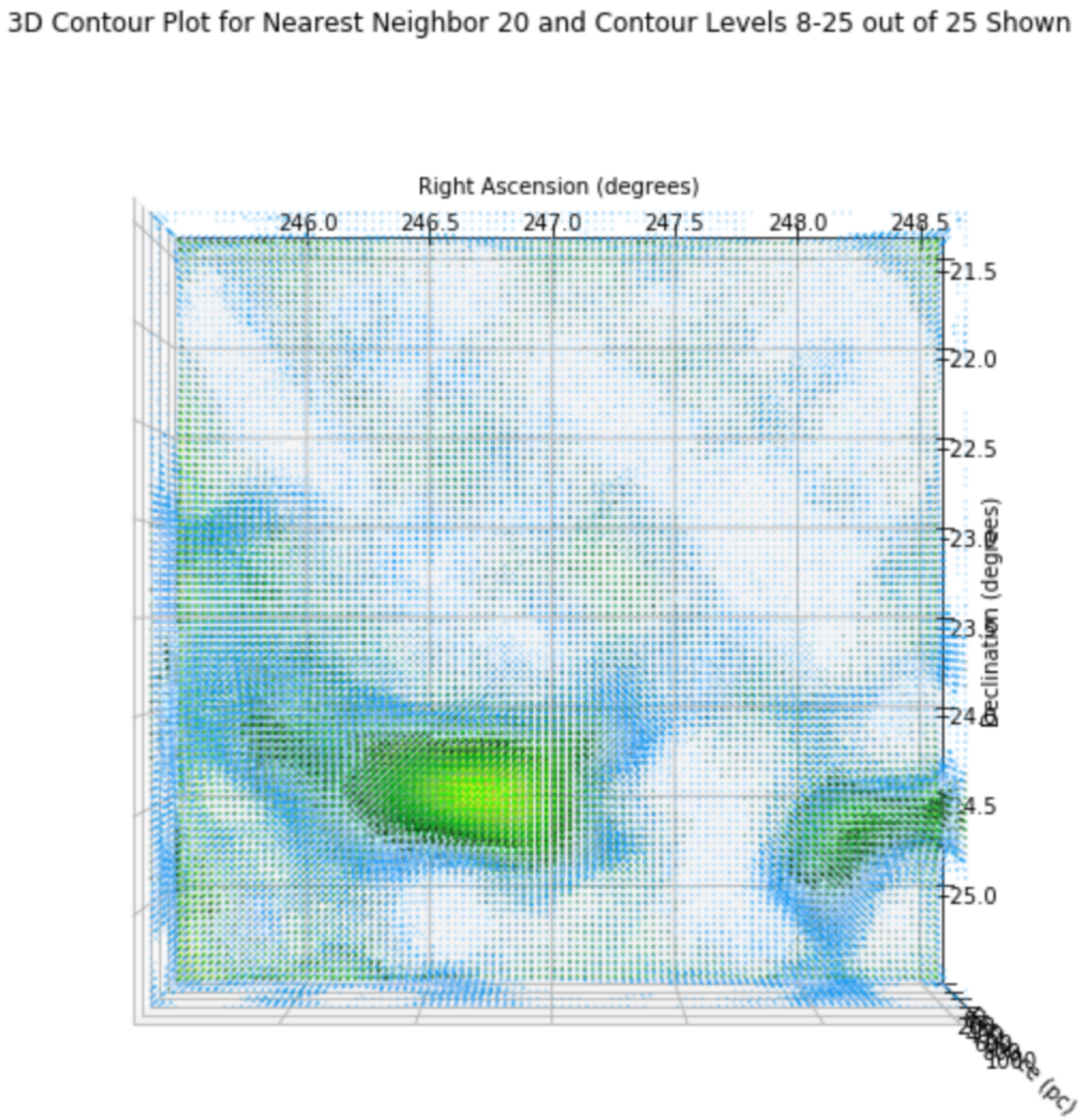

Figure 3.7: Three-dimensional contour results for contours 8-25 out of 25 of the 20th nearest neighbor distance mapping, viewed from the top to match the perspective of the two-dimensional mapping.

From this angle, the right ascension and declination coordinate axes are oriented precisely as they are in the two-dimensional plots. The various translucent blue, green, and yellowish splotches spread throughout the plot should be ignored, as these are solely the fringing effects at small distances (near Earth). Taking this into consideration, it appears that there is strong evidence that this contour plot in three dimension resembles the results in two-dimensions. This top view validates the idea that the interstellar clouds should appear as columns when looking at increasing distance based on the obscuration of stars from detection. Essentially, these three-dimensional visualizations support all of the results assembled and give very useful determinations of the locations and shapes of the interstellar clouds in Ophiuchus.

Chapter 4

Three-dimensional Realistic Coordinate Nearest Neighbor Mapping

Using celestial coordinates and distance from Earth to study interstellar clouds can be quite useful, but, as discussed, in reality, with greater distance, the volume of space increases drastically, and stars become much farther apart in space than right ascension and declination alone suggest. As a result, for the true, comprehensive treatment, this research requires investigating clouds using realistic space coordinates. Using the mathematical transformation from spherical coordinates (where the theta coordinate is 90 degrees minus the declination) to rectangular coordinates, the data were converted into arbitrary x, y, and z space coordinates centered at the Earth. In order to remain consistent with the previous work, and also to limit the scope of the study to avoid complications (the farthest star was hundreds of millions of parsecs away), the maximum distance (rho coordinate) was restricted to 1,000 pc.

With the star data converted to realistic coordinates, pixels were created by first making them in spherical coordinates (the distance was again limited to 1,000 pc as the maximum) and then converting these to rectangular coordinates to ensure the same region of space as the star data. Because the computations for this work are more intensive and require more memory and time, the number of pixels per coordinate (in other words, the number of linearly space points created for each of the coordinates) was set to 50 instead of 100 in the previous work. Due to the "stretching" of spherical coordinates into space as distance increases, the pixels appear as solid angle squares growing in size. To display the situation as it appears in space, Figure 4.1 shows the Earth at (0,0,0), the star locations in pale blue, the celestial equator (z-plane) in red, north celestial pole going directly upward in dark blue, and the pixels in green. As shown, the stars in Ophiuchus are located below the celestial equator, since they have negative declination, and the pixels do not perfectly match the conical shape of the stars.

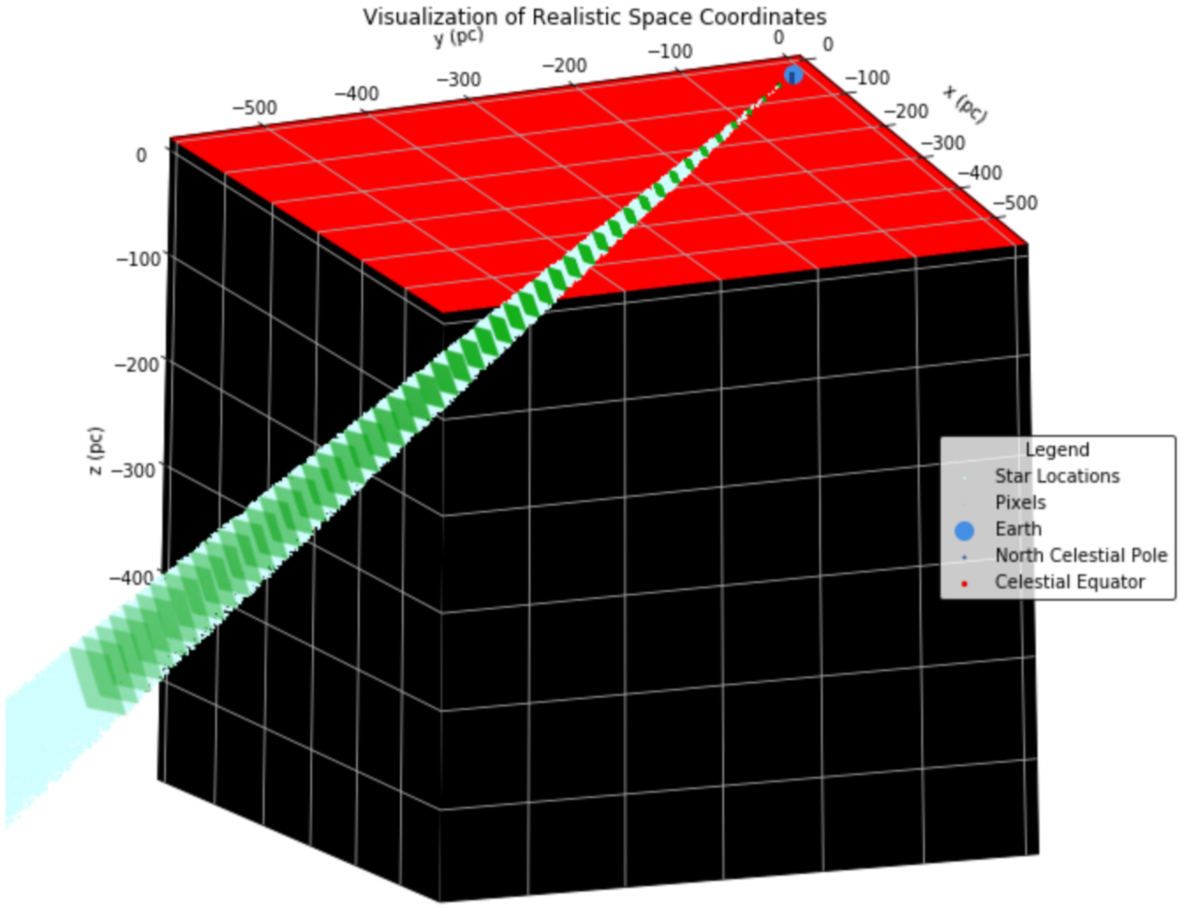

Figure 4.1: Three-dimensional picture of the Ophiuchus data and setting in realistic space coordinates. On the figure, where it is difficult to discern from the legend, the stars are pictured in pale blue, and the pixels are in green.

As before, all of the data and the pixel points were shifted to be centered around zero and normalized to eliminate discrepancies in the weights of the coordinates. With the star data converted and the pixels set up accordingly, the same process using the KD Tree algorithm was utilized to find the 20th nearest neighbor distances. Following this step, the three-dimensional distance map was created, this time in realistic space coordinates, and this is shown below in Figure 4.2.

There is great difficulty in gleaning any relevant information about interstellar cloud location or structure due to the bias of having larger nearest neighbor distances very close to the Earth because of the much lower density of stars. While not at first apparent, this can be seen as the small teal and yellow pixel areas directly adjacent to the Earth in the plot. As such, not too much can be gained from this type of three-dimensional plot. The same graph, with

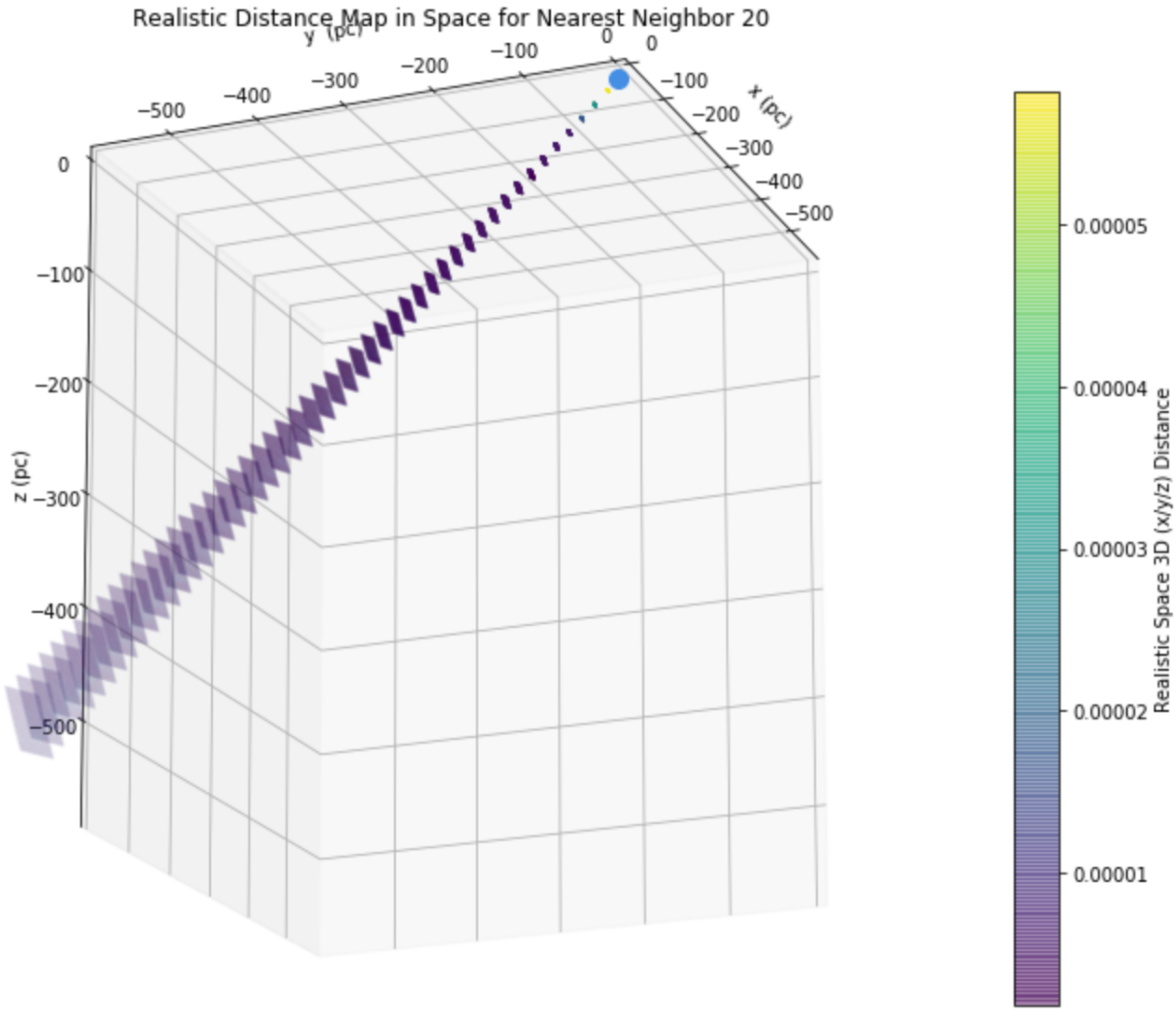

Figure 4.2: Three-dimensional 20th nearest neighbor distance map in realistic space coordinates. The nearest neighbor distances are shown according to the color bar.

the addition of the star locations in red, is displayed below in Figure 4.3. This plot backs up the previous point made: the density of stars is great far away from Earth, where the stars appear as an opaque cone, but the density is significantly lower near the Earth, which causes the drastically larger nearest neighbor distance. It is for this reason that such a realistic space coordinate analysis is at this time uninformative.

Due to time limitations on using the DR 2 data, this work has not been completed or brought to its full potential. Improvements can definitely be made to the realistic space analysis and visualizations given more time and work done on Python code. Currently, two major limitations on the scope of the usefulness of the realistic space branch of the work are the inability to use larger numbers of pixel divisions and the strong skew towards maximum

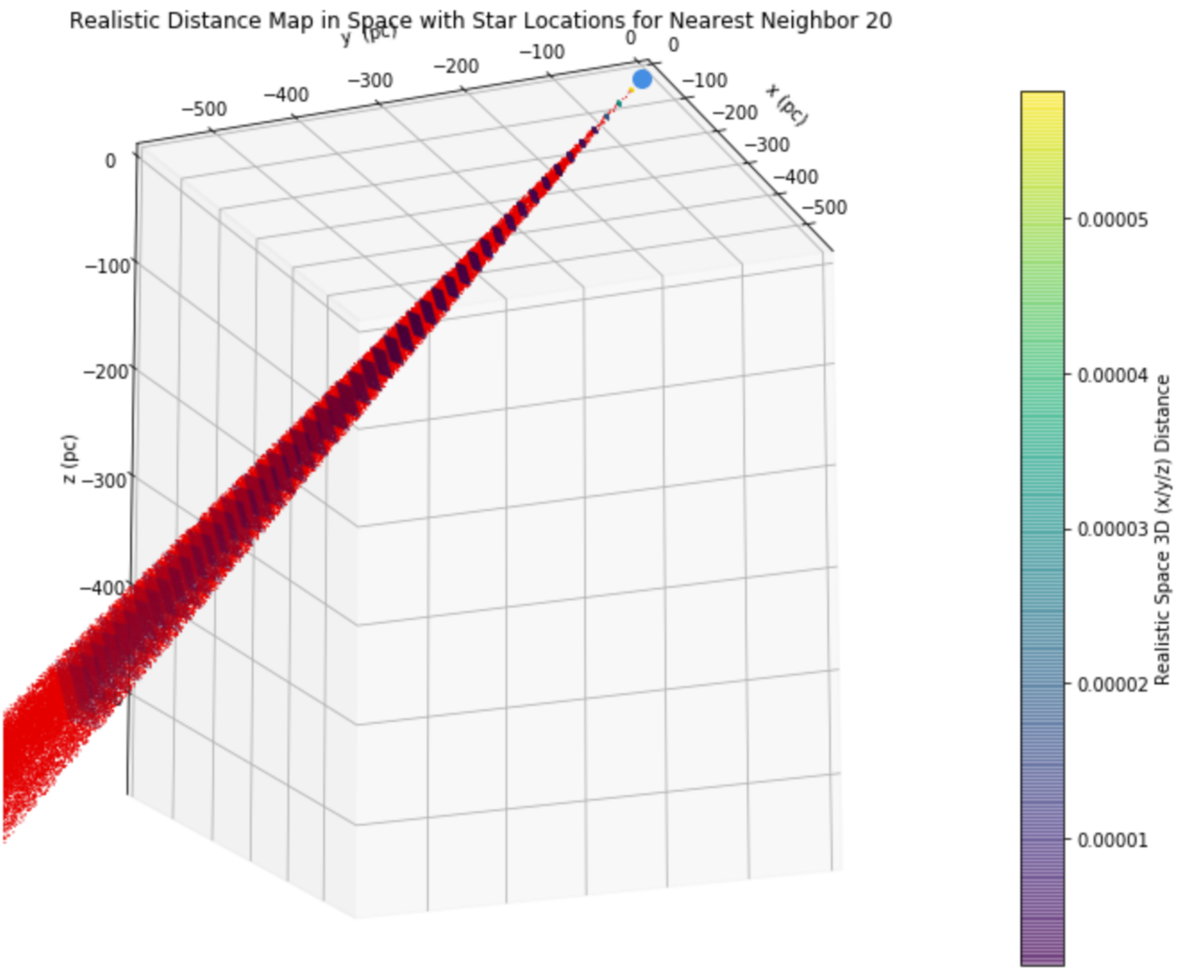

Figure 4.3: Three-dimensional 20th nearest neighbor distance map in realistic space coordinates with Ophiuchus star locations. The nearest neighbor distances are shown according to the color bar, and the stars are pictured in red.

nearest neighbor distance close to Earth. For the first issue, increasing the number of pixel divisions per coordinate to a number far greater than 50 would resolve the apparent distinctions and allow for discovery of cloud structure continuously. For the second, addressing this bias, perhaps by only beginning the analysis a few hundred parsecs from Earth or removing those skewing data, would allow for determination of where nearest neighbor distance has relative differences. Further work would allow for conclusions to be made and is highly recommended to be undertaken. Currently, however, no conclusions can be made regarding the realistic space coordinate three-dimensional analysis other than the fact that it has strong potential.

Chapter 5

Conclusion

As described, the methodology employed to appropriately set up DR 2 star data, create coordinate pixels, and put into effect the KD Tree nearest neighbor tool worked effectively on the zodiacal constellation Ophiuchus. Such methods can just as easily be applied to any other region of the sky to create a 360-degree view around Earth, allowing for interstellar cloud analysis of all of space. The nearest neighbor approach proved incredibly useful, as the results from two and three dimensions matched each other well in all facets, which indicates that the work functions as it should. Moreover, the visualizations, both in 2D and 3D are helpful to understand and picture the structure of interstellar clouds and "put a face" to the esoteric dust and gas clouds millions of miles away from Earth. The realistic space coordinate way of analyzing the data is certainly the best approach to understand clouds as they truly are, but time limitations with DR 2 made that impossible for the scope of this study, although massive potential exists in this regard.

Further research is recommended making improvements to the realistic space coordinate analysis and three-dimensional visualizations. In doing so, the major improvements that can be made are to increase the number of pixel divisions of each coordinate and making changes to remove the bias of low stellar density close to Earth. In addition, studies of nearest neighbor distance methods and their applications at distances beyond 1,000 pc could reveal helpful or surprising answers and are worth delving into fully. Of course, once this work is improved and its breadth risen, this same research can and should be applied to many other intriguing regions of the sky, such as Sagittarius, which looks directly into the center of the galaxy. Then, hopefully, understanding the locations, patterns, and structures of interstellar clouds in the sky around Earth will be a key step in the path towards using the GAIA mission to map the Milky Way Galaxy in unprecedented detail.

References

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

[2] A. G. A. Brown *et al.*, "GAIA Data Release 2," European Space Agency, April 25, 2018.

Acknowledgments

In concluding this investigation of interstellar clouds with nearest neighbor methods using 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, made me enthusiastic about the subject, and led me to my work in the field.