ABSTRACT

This paper presents a statistical technique to determine the number of red giant branch (RGB) stars in mixed samples from the Andromeda spiral galaxy (M31) and foreground Milky Way (MW) dwarf stars. The data for this analysis were obtained from the DEIMOS instrument on the Keck II 10-m telescope. The technique involved creating empirical probability distribution functions (PDFs) for the M31 RGB and MW dwarf populations using samples of stars with reliable spectral classifications. Multiple such samples were used to represent the RGB and dwarf populations as basis functions and each one's effectiveness was evaluated. The basis functions were constructed in three two-dimensional diagnostic spaces and similar distributions were created for mixed stellar populations of undetermined RGB and dwarf compositions. The weighting of RGB and dwarf PDFs for which the mixture of the two basis functions most closbely resembles that of the populations was then determined. The method was further extended to determine the optimum mixture of more than two parent populations. Applying the method to several samples defined based on radial distance from the center of M31 yields results consistent with other modes of analysis, affirming the accuracy of this technique.

Subject headings:

1. INTRODUCTION

As our interior position relative to the MW biases our view of the Galaxy's halo, studying M31, a spiral galaxy like our own, provides excellent insight into the processes that govern the MW as well. We have a phenomenal vantage point from which to analyze our neighbor because the two galaxies are separated by a relatively small distance of 783 kiloparsecs and M31 has a disk inclination angle of 78 °, giving us a wide field of viewGilbert et al. (2006). However, despite these benefits, the proximity causes foreground dwarf star contamination (Guhathakurta et al. 2005). The relatively large area M31 spans on the sky leads to such contamination, and even after photometric cuts designed to exclude MW dwarfs from observation, dwarfs still comprise a large portion of the objects surveyed by the DEIMOS instrument, especially at large distances from M31's center. Methods to determine the spectral classification of individual stars or determine the number of RGB and dwarf stars in a given population are thus important for the analysis of

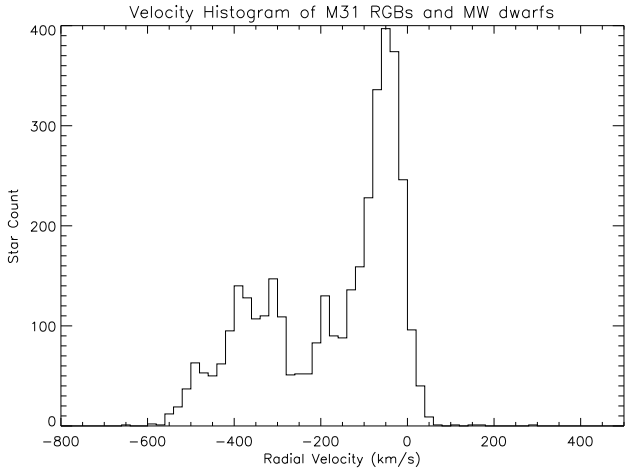

FIG. 1.— A velocity histogram of stars analyzed by DEIMOS with a clear bimodal distribution. The M31 giant distribution, corresponding to the left peak, and the MW dwarf distribution, corresponding to the right peak, clearly overlap between these two apexes.

M31. This paper focuses on the latter method: determining the spectral composition in terms of M31 RGB and MW dwarf stars of sample populations.

Though classifying objects as RGB or dwarf stars based on radial velocity is a traditional means of separating these two populations, Gilbert et al. (2006) notes that their radial velocity distributions overlap. While the bimodal distribution of radial velocities in Figure 1 strongly suggests the presence of two distinct populations, the common area between the two peaks renders this distinguishing method imperfect. A method that would determine the number of giants and dwarfs in a given sample overcomes these limitations by comparing the distribution of the sample to two standardized distributions: the RGB and dwarf PDFs. Another possible way of classifying stars is their brightness, as RGB stars are generally expected to be brighter than dwarfs. However, MW dwarf stars are far closer to the Earth than M31 RGB stars; thus, dwarf stars appear brighter relative to M31 RGB stars than they actually are. Indeed, the differences in distances between these two populations are such that the apparent brightness ranges of the two populations overlap significantly.

While Gilbert et al. (2006) presents a method to classify individual stars in a sample, the technique presented here is not applicable to individual stars. Rather, we statistically estimate the composition of the whole population, which provides a useful tool for analyzing the density of RGB stars across various regions of M31. This technique overcomes a key limitation of the individual star method: the need to deal with stars whose individual classifications are uncertain. As those stars whose categorizations are not firm cannot be reliably assigned to one population or the other, the individual star method encounters difficulties with determining the composition of a population.

This paper seeks to facilitate the study of galaxies such as M31 by providing a new method for quantifying the number stars of a certain spectral type in a given population. The basis functions used serve as estimates of where, probabilistically, an "average" star of a given spectral classification would be located on a diagnostic space. Finding the optimum combination of basis functions for an actual population yields an estimate for the number of stars of each class in the population. This technique's underlying purpose is to quantify the number of stars of different spectral types in a mixed population of stars. This quantification has utility in constructing surface brightness profiles and density functions for M31's halo (Gilbert et al. 2006).

2. METHODS

2.1. *Isolating RGB and Dwarf Star Populations*

As the basis functions for RGB and dwarf stars are intended to represent the probability distribution of solely one spectral class of star, the populations used to create them must necessarily have a high degree of purity. This section describes the process of selecting these populations and creating the various RGB and dwarf basis functions used in the study. As we have created multiple basis functions to represent each spectral class, we evaluate each PDF's accuracy in Section 3.3. Furthermore, since the RGB star distribution in the 3 diagnostic spaces changes for different star metallicities, as noted in Section 3.2, PDFs that represent the metal-rich and metal-poor RGB stars separately are created as well.

2.1.1. *Dwarf Satellite Galaxies*

One of the RGB basis functions was created based on stars in the various dwarf spheroidal (dSph) satellite galaxies orbiting M31. These dwarf galaxies' members, primarily RGB stars, are similar to those in M31 and can thus be used as a basis function representing an M31 giant. A key distinction, however, lies in dwarf galaxy members' radial velocities: the satellite galaxies have velocity distributions quite different from those of typical M31 giants. Indeed, some of these galaxies have distributions that have minimal overlap with the MW dwarf velocity distribution, as evidenced by Figure 2, a bimodal distribution which includes two clearly separated peaks. The peak at the lower radial velocity is an easily isolated source of RGB stars, as it is very separate from the dwarf peak in its field.

By constructing velocity histogram for stars in dwarf satellite fields, we developed velocity ranges for each dwarf galaxy field that allowed us to definitively select RGB stars to construct a basis function. Table

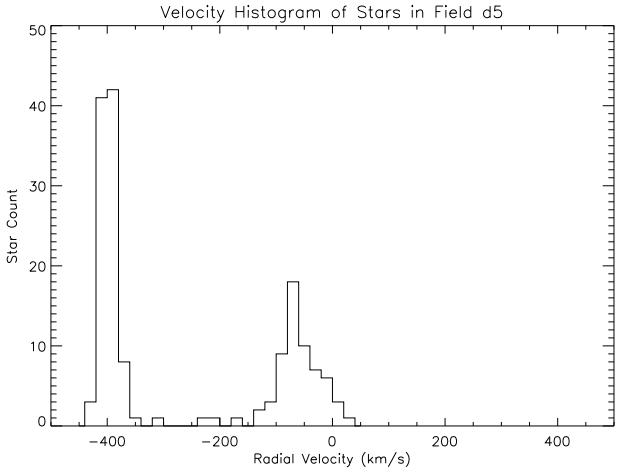

FIG. 2.— A velocity histogram of stars on belonging to field d5. Though the right peak is clear evidence of dwarf contamination, the two distributions do not overlap.

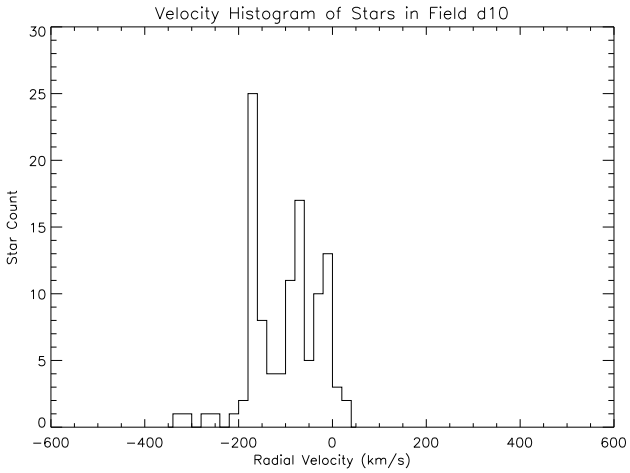

FIG. 3.— A velocity histogram of stars on belonging to field d10. There is no distinct boundary between the left red giant distribution and the right dwarf distribution; thus, this field was not included in the RGB PDF based on dwarf satellite galaxies.

1 summarizes these ranges and notes the fields without significant overlap in the two distributions. Those dwarf satellite galaxy fields in which the RGB and dwarf peak overlapped so their boundaries were not readily distinguishable, as in Figure 3 were not included in the basis function population. The stars satisfying the velocity ranges in such marked fields were incorporated into the RGB training set population. Note that though some included fields' ranges may appear to overlap with the radial velocity ranges of MW stars suggested by Figure 1, the selected RGB stars' distributions displayed no overlap in their own fields. The 575 stars satisfying the conditions in Table 1 constitute the training set used to define the satellite galaxy RGB PDF. Though this set can be used as the complete RGB PDF, analysis in Section 3.3 indicates that the population lacks metal-rich RGB stars. The set is thus more apt to represent solely the metal-poor RGB stars.

Field Name	RGB Radial Velocity Range (km/s)	Stars Included
d1[a]	$-420 < v < -330$	95
d2	$-220 < v < -160$	111
d3	$-400 < v < -300$	68
d5	$-430 < v < -360$	94
d7	$-350 < v < 270$	168
d10[b]	$-190 < v < -120$	0
d13[b]	$-200 < v < -120$	0
A170	$-500 < v < -450$	39

[a]An additional cut in photometric [Fe/H] metallicity was made for d1 to account for the effects of another population in the background. Stars in field d1 were required to satisfy both the radial velocity cut and the restriction $-3 < [Fe/H]_{phot} < -1$

[b]Stars from these fields were excluded from the basis function because the RGB and dwarf star velocity distributions overlapped in that field. The velocity ranges indicated are approximations of the boundaries.

TABLE 1
DATA REGARDING THE SATELLITE GALAXY FIELDS USED TO DEFINE AN RGB PDF.

Field Name	Average Radial Distance from M31's Center (kiloparsecs)	Stars Included
d7	220.884	40
A170	160.243	44
m11	158.992	117
A080	165.222	127
A305	169.852	148

TABLE 2
DATA REGARDING FAR FIELDS FROM M31 USED TO DEFINE A DWARF PDF

2.1.2. *Far Fields from M31*

As fields very far from the center of Andromeda are expected to have a very low number of M31 giants, we chose to use stars from these fields for one of the dwarf basis functions. Stars from various fields were included in these sets based on their radial distance from M31's center. Stars from dwarf galaxies were deliberately excluded as they are generally RGB stars despite their large radial distance from M31's center. Table 2 lists these fields, their average distance from M31's center, and the number of stars contained in each. Section 2.2 covers the method of how to calculate an object's radial distance from M31 using its right ascension and declination. The total of 476 stars in the fields listed in Table 2 composes a dwarf basis function.

2.1.3. *Diagnostic-based Populations*

We also created basis functions by using a classification scheme developed by Gilbert et al. (2006) to determine the likelihood of any star being a giant or a dwarf. The classification scheme is based on 5 criteria and assigns an integer value between -3 and 3 to each star, with -3 corresponding to a very likely dwarf and 3 corresponding to a very likely giant. The classification of stars with more intermediate values

is less certain, and stars with a value of 0 are on the border between the two spectral classes. We define an RGB basis function with the population of all stars with a classification of 1 or greater and define a dwarf basis function with the population of all stars with a classification of -1 or less. These sets, which do not include stars with the intermediate classification of 0, are thus comprised of stars whose classifications are well-determined. The populations include a total of 1501 stars for the RGB basis function and 2210 for the dwarf basis function. We further split the RGB population into two separate populations, one to represent the metal-rich RGB stars and another to represent the metal-poor RGB stars. The metal-rich PDF is composed of the 750 stars with [Fe/H]$_{phot}$ ([Fe/H] determined by photometric means) greater than the median [Fe/H]$_{phot}$ of the RGB basis population, -1.2890236. The metal-poor PDF is composed of the 750 stars with [Fe/H]$_{phot}$ less than this value.

2.2. *Calculating Radial Distance from M31's center*

In order to separate the entire M31 star population into different radial fields, each star's radial distance from M31's center was calculated from our data, which included right ascension (ra) and declination (dec) values for each star. We treated M31 as a section of a sphere centered on the Earth, with a radius of 783 kiloparsecs, the distance between our planet and M31's center. All stars in M31 were approximated as lying on the sphere and thus equidistant from Earth. Based on this assumption, the ra and dec values were treated as spherical coordinates with equal radii and converted into Cartesian coordinates based on the following formulae, where ra and dec are given in radians:

$$x = \cos dec \times \cos ra \tag{1}$$

$$y = \cos dec \times \sin ra \tag{2}$$

$$z = \sin dec \tag{3}$$

These formulae are used to calculate the Cartesian unit vector (x, y, z) associated with M31's center and each star in the database. For a given star, the angle θ between the v$_{center}$, the vector pointing to M31's center, and v$_{star}$, the vector pointing to its own location, is calculated using the relation

$$\theta = \frac{v_{center} \cdot v_{star}}{|v_{center}| \times |v_{star}|} \tag{4}$$

After obtaining the value for θ, the radial distance from M31's center is approximated as the arc length between the 2 points on the sphere. This rough assumption is valid because the distance from the Earth

to M31's center, 783 kiloparsecs, generally dwarfs the radial distances of stars from the center. Moreover, these radii are only used for grouping stars into clusters; thus, they only need to be accurate relative to each other. The radial distance of a star is thus $d = r \times \theta$, where r is 783 kiloparsecs. An important point to note about this radial distance is that it makes no sense for MW dwarfs contaminating the samples chosen for analysis. Rather, for dwarfs, the distance represents what an RGB star's radial distance would be if it had the same ra and dec as the dwarf. However, the radial distance calculation still effectively categorizes M31 giants into different fields, ensuring the validity of analyzing dwarf contamination as a function of radius.

2.3. *Diagnostic Spaces*

The method was carried out in three specific two-dimensional diagnostic spaces, though it is extensible to others. The specific diagnostic spaces used were plots of the strength of the NA_I 8190 Å absorption line doublet versus $(V\text{-}I)_0$ color, [Fe/H] metallicity as determined by spectroscopy versus that determined by photometry, and $(M\text{-}D)_0$ color versus $(M\text{-}T)_0$ color. The first space plots the strength of the sodium doublet versus a measure of the color of the star. Schiavon et al. (1996) explores the sodium doublet as a possible distinguishing feature between dwarfs and giants. We chose to plot this value against the color of the star to further separate the overlapping distributions of RGB and dwarf stars when considering the strength of the sodium doublet alone. The second space plots two measures of the [Fe/H] metallicity of stars against each other. $[Fe/H]_{phot}$ is a theoretical estimate of a star's metallicity based on its position in a color-magnitude diagram, determined based on comparison to groups of RGB stars known as isochrones. $[Fe/H]_{spec}$ is based on the strength star's Ca_{II} absorption line triplet and the information it gives about the star's metallicity. These measures are calibrated based on RGB stars and are thus expected to be similar for those stars; dwarf stars, however, are expected to be below the $[Fe/H]_{phot} = [Fe/H]_{spec}$, as their Ca_{II} absorption line triplets are generally weak in comparison to that of RGB stars (Guhathakurta et al. 2005). The third plot, also known as the two color diagram, is a plot of two differently defined color values for a given star. Majewski et al. (2000) analyzes this plot as a discriminator between RGB and dwarf stars, which are located in two overlapping yet distinct distributions. Most plots in this paper use the two color diagram, as these distributions are fairly distinguishable. Though all these diagnostic spaces were created to distinguish RGB stars from dwarf stars, there is still significant overlap in each plot, making individual classification difficult for those stars in the overlapping range. This paper's method deals with these overlaps by using PDFs to represent the two spectral classes of stars.

2.4. *Gaussian Smoothing*

In order to convert a set of discrete stars on a two-dimensional plot into a probability function, we apply Gaussian smoothing to each star on the plot. This process entails dividing the plot into small rectangular pixels and distributing each individual star across the surrounding pixels. Stars are thus no longer discrete points on the plot but likelihood functions spread over a certain area. Applying this process to a whole population of stars yields a two-dimensional array of values representing the probability of a star in that population landing in a certain pixel. Moreover, if a given population of stars is expected to be purely of one spectral classification, applying the Gaussian smoothing process to the entire population and normalizing the sum of the elements in the two-dimensional array to 1 results in a distribution that represents the "average" star of that spectral type. The Gaussian function used to distribute a discrete star across the surrounding pixels has the following formula:

$$W(i,j) = e^{\frac{-i^2}{2\sigma_x^2}} e^{\frac{-j^2}{2\sigma_y^2}} \tag{5}$$

Here i and j refer to the distance a given pixel is from the pixel in which the given star belongs in the x and y directions, respectively. This formula is used to create a Gaussian kernel for the smoothing process, which is a 2-dimensional array containing floating point values used in the smoothing, with each element corresponding to a pixel. This kernel, which is normalized so that its values sum to 1, represents the weights given to the pixels surrounding a center pixel in which a star is located. As Formula 5 suggests, the largest weight is given to the star's own pixel and pixels surrounding the center receive progressively smaller weights with increasing distance (corresponding to increasing i and j). The kernel spans 3 standard deviations from the center in both the x and y dimensions; thus, its size varies based on the x and y dimensions of a pixel and the x and y standard deviations (σ_x and σ_y). This 3 standard deviation range is imposed on the kernel because the values assigned to pixels farther away than this range become vanishingly small. As an example, for a case in which a pixel's x dimension equals σ_x and its y dimension equals σ_y, the kernel would be 7 pixels by 7 pixels, encompassing 3 standard deviations in both directions for both the x and y dimensions. Such a kernel is represented by the following matrix:

$$\begin{pmatrix} 2.0\times10^{-5} & 2.4\times10^{-4} & 1.1\times10^{-3} & 1.8\times10^{-3} & 1.1\times10^{-3} & 2.4\times10^{-4} & 2.0\times10^{-5} \\ 2.4\times10^{-4} & 3.0\times10^{-3} & 1.3\times10^{-2} & 2.2\times10^{-2} & 1.3\times10^{-2} & 3.0\times10^{-3} & 2.4\times10^{-4} \\ 1.1\times10^{-3} & 1.3\times10^{-2} & 5.9\times10^{-2} & 9.7\times10^{-2} & 5.9\times10^{-2} & 1.3\times10^{-2} & 1.1\times10^{-3} \\ 1.8\times10^{-3} & 2.2\times10^{-2} & 9.7\times10^{-2} & 1.6\times10^{-1} & 9.7\times10^{-2} & 2.2\times10^{-2} & 1.8\times10^{-3} \\ 1.1\times10^{-3} & 1.3\times10^{-2} & 5.9\times10^{-2} & 9.7\times10^{-2} & 5.9\times10^{-2} & 1.3\times10^{-2} & 1.1\times10^{-3} \\ 2.4\times10^{-4} & 3.0\times10^{-3} & 1.3\times10^{-2} & 2.2\times10^{-2} & 1.3\times10^{-2} & 3.0\times10^{-3} & 2.4\times10^{-4} \\ 2.0\times10^{-5} & 2.4\times10^{-4} & 1.1\times10^{-3} & 1.8\times10^{-3} & 1.1\times10^{-3} & 2.4\times10^{-4} & 2.0\times10^{-5} \end{pmatrix}$$

In order to use this kernel to convert the population of individual stars into a PDF for one 2-dimensional diagnostic space, we first select boundary ranges for the two variables of interest to confine the smoothing process to a rectangular region of space containing all the stars in the population. This region is then divided into small rectangular pixels and is represented by a two-dimensional array of values. We then iterate over the population of stars and use the Gaussian kernel to add each star's contribution to the array. To do so, we first calculate the pixel in which the star is located and treat it as the center of the Gaussian kernel. We then superimpose the kernel on the array, aligning the kernel's center with the pixel to which the star belongs. Then each value in the area around the star's pixel is incremented by its corresponding value in the kernel. For example, for the 7 by 7 kernel above, the pixel to which a star belongs would be incremented by 1.6×10^{-1} and the 4 pixels directly adjacent to the star's pixel would be incremented by 9.7×10^{-2}. If a star falls too close to the border and part of the superimposed kernel is out of the rectangular region, the kernel's contributions in that section are neglected.

The smoothing process effectively functions as a discrete convolution of the Gaussian kernel with the two-dimensional array prior to smoothing, which is simply a two-dimensional array whose values are the number of stars in each pixel. If the smoothed population is intended as a basis function for a stellar class, the two-dimensional array is normalized by dividing such that its values sum to 1. This division makes the array a representation of a single star of that stellar class, spread into a probability distribution to represent its "average" location.

2.5. *Minimizing Residuals*

In order to quantify the number of red giants in a given set of stars, the smoothing process must be applied to three separate populations: a dwarf basis population, a giant basis population, and the mixed population of

interest. The three two-dimensional arrays resulting from this process are then used to estimate the number of red giants in the mixed population. Since the mixed population consists of both dwarfs and giants, its PDF is expected to be a combination of the dwarf and giant PDFs. An estimate for the number of red giants can thus be produced by combining the dwarf and giant PDFs with different weightings and selecting the combined distribution that most closely mirrors that of the population of interest. This task is accomplished by iterating over all possible combinations of the two spectral types. Thus, for a population with n stars, we test the combinations starting with 0 giants and n dwarfs, then 1 giant and $n-1$ dwarfs, iterating up to n giants and 0 dwarfs. The theoretical combined distribution for k giants is given by $\frac{k}{n} \times d_{giant} + \frac{n-k}{n} \times d_{dwarf}$, where d_{giant} is the smoothed two-dimensional array of the giant PDF while d_{dwarf} is that of the dwarf PDF. Since each array from the dwarf and giant basis functions sums to 1, the combined array will have a total sum of n, the same as the sum of the smoothed population of interest. Thus the arrays differ primarily in how their values are distributed.

For each combined array for k from 0 to n, we define a statistic quantifying how similar the combined array is to the smoothed population array. For $x \times y$ arrays a_{pop} and $a_{combined}$ where $a[i,j]$ represents the element in the ith row and jth column, the statistic s is defined as follows:

$$s = \sum_{i=0}^{x} \sum_{j=0}^{y} (a_{pop}[i,j] - a_{combined}[i,j])^2 \tag{6}$$

This statistic is effectively the sum of of the squares of elements in the residual array obtained by subtracting the combined array from the population array. As we perform the iteration from $k = 0$ to n, we keep track of the lowest value of s so far and the value of k for which s takes on that value. When the iteration is complete, the stored value of k is the number of red giants in the population such that the combined array most closely resmbles the population array.

2.6. *Additional Basis Functions*

This method can be extended to cover more than one basis function; that is, it can be used to estimate the composition of a population of stars containing more than 2 distinct subpopulations. This extension can be implemented by iterating over all possible combinations of the distinct populations. Consider a case with a population of n stars with 3 distinct subpopulations and let k, l, and m denote the number of stars in each subpopulation. Using a doubly-nested for loop with an outer loop iterating k from 0 to n and an inner loop iterating l from 0 to $n-k$ will exhaust all possible combinations of the multiple subpopulations. This

method, however, has drastic increases in runtime as the number of subpopulations increases. A refinement is to first perform a coarse search, perhaps incrementing k and l by a number greater than 1, noting the optimum values of k, l, and m found by this search, and performing a fine search in only that region. While pruning the search space makes this method much faster, there is the additional risk of landing in a local minimum and not finding the optimum results. Moreover, these subpopulations must all be sufficiently large so that our assumption that the subpopulations reflect an "average" mix of stars of a given stellar type holds; otherwise, our estimates might be inaccurate. While this problem applies to the 2 subpopulation scenario as well, it is of particular concern with many subpopulations. These additional complications demonstrate that the method works best when applied to fewer subpopulations.

3. RESULTS AND DISCUSSION

This section covers the results of the technique when applied to various star populations within M31. The various PDFs were first evaluated against one another and the functions that best represented the giant and dwarf populations were selected. The evaluation process is detailed in Section 3.3. The technique was then applied to populations separated based on radial distance from M31, where radius was determined by the method detailed in Section 2.2. The method was further applied different fields within M31, which consist of groups of stars belonging to the same aperture mask. These data may be converted into surface brightness profiles and density models for M31 (Gilbert et al. 2006), which possible extension of the method in the future.

3.1. *Radial Distribution of RGB Stars and Dwarf Contaminants in M31*

The M31 population was separated into 5 different subpopulations based on radius r: $r < 20$ kpc, 20 kpc $< r < 40$ kpc, 40 kpc $< r < 60$ kpc, 60 kpc $< r < 100$ kpc, and 100 kpc $< r$. As indicated in Section 3.3, the PDFs based on the classification scheme developed by Gilbert et al. (2006) were the most accurate; the data in the Table 3 were obtained using these PDFs.

Table 3 summarizes the results obtained by applying the method to radial bins in all three diagnostic spaces. Note that the table has different numbers of stars in a given radial bin for each diagnostic space; these differences are attributable to faulty data for various stars. For example, the *(M-D)*$_0$ color and *(M-T)*$_0$ color values were incorrect for all stars in the radial range from 0 to 20 kpc. None of these stars were included in the population because of their erroneous data. Such failures occur inconsistently among the

Radial Bin	Number of RGB Stars[a]	Total Stars	Proportion of RGB Stars
0-20 kpc	88/125/0[b]	147/144/0	0.598/0.868/NA
20-40 kpc	269/337/146	500/499/288	0.538/0.675/0.507
40-60 kpc	144/156/112	566/569/564	0.254/0.274/0.199
60-100 kpc	135/94/116	1223/1230/1114	0.110/0.076/0.104
100+ kpc	73/91/44	624/627/624	0.117/0.145/0.071

[a]The entries in this column and the subsequent columns are formatted to give the information for all three diagnostic spaces. The first of the three numbers is for the NA_I Absorption Line Strength versus $(V\text{-}I)_0$ Color plot, the second is for the $[Fe/H]_{spec}$ versus $[Fe/H]_{phot}$ plot, and the third is for the two color diagram.
[b]All the two color diagram values for stars from 0 to 20 kpc failed.

TABLE 3
RESULTS OF ESTIMATING THE NUMBER OF RGB STARS IN RADIAL BINS USING THE M31CLASS-BASED RGB PDF

different diagnostic spaces, leading to the differences in total values and perhaps influencing the predicted number of RGB stars. Indeed, the failures can cause discrepancies in the predicted proportion of RGB stars in different diagnostic spaces; these errors are most evident in the 59.8% and 86.8% RGB proportions estimated by different diagnostic spaces in the 0-20 kpc range. Increasing the data set and seeking to minimize faulty measurements would combat this source of error.

Despite such inconsistencies, there is a clear increasing trend in dwarf contamination of the populations with increasing radius, as evidenced by the relatively consistent decreases in the proportion of RGB stars. This proportion is monotonically decreasing as we consider more distant radial bins until the 100+ kpc range, which is marginally greater than the predicted proportion for the 60-100 kpc range. This unexpected increase perhaps reflects the tendency of the method to lose accuracy as the size of one subpopulation dwindles. In the 100+ kpc range, the population of RGB stars is quite small and may not accurately reflect the average distribution of RGB stars in the diagnostic space. Moreover, with two exceptions, the estimated proportions of RGB stars are all quite close to one another (within approximately 10% for different diagnostic spaces). In fact, the only two exceptions occur close to one extreme, where the dwarf population is small and the accuracy of the method is expected to decrease. The fact that completely different plots produce generally similar results ultimately speaks to this method's great accuracy. Both the increasing trend of contamination and the consistency between diagnostic spaces conform with the expected outcomes, attesting to this method's general validity. The technique's validity can be further tested by constructing a surface brightness profile and comparing it to that noted by Guhathakurta et al. (2005).

3.2. *Metal Rich versus Metal Poor Stars*

As evidenced by Figure 4, the M31 RGB population appears to have two distinct subpopulations: metal-rich stars and metal-poor stars. These stars have different distributions in the two color diagram diagnostic

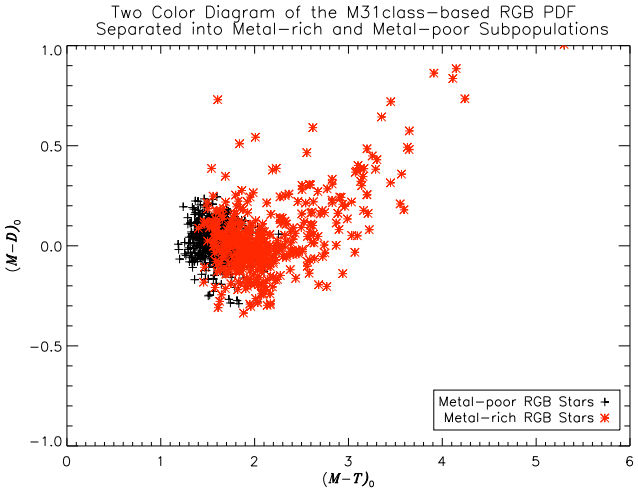

FIG. 4.— A two color diagram of the M31class-based RGB PDF. There is a distinct curve above and to the right of the main mass of points that appears to be a dwarf PDF characteristic; however, it is actually characteristic of metal-rich RGB stars. The two separate subpopulations were isolated by taking stars with [Fe/H$_{phot}$] greater than the mean value for the metal-rich stars and the ones with lower [Fe/H$_{phot}$] for the metal-poor stars (the mean value was -1.29)

.

Radial Bin	Number of Metal-rich[a] RGB Stars	Number of Metal-poor RGB Stars	Total Stars	Proportion of Metal-rich Stars (Out of all RGB Stars)
0-20 kpc	112/136/0[b]	6/12/0	147/144/0	0.949/0.918/NA
20-40 kpc	300/335/157	43/55/39	500/499/288	0.875/0.859/0.781
40-60 kpc	110/129/108	51/43/39	566/569/564	0.683/0.750/0.735
60-100 kpc	57/35/50	73/55/68	1223/1230/1114	0.438/0.389/0.424
100+ kpc	39/47/30	35/44/21	624/627/624	0.527/0.516/0.588

[a]The entries in this column and the subsequent columns are formatted to give the information for all three diagnostic spaces. The first of the three numbers is for the NA$_I$ Absorption Line Strength versus $(V\text{-}I)_0$ Color plot, the second is for the [Fe/H]$_{spec}$ versus [Fe/H]$_{phot}$ plot, and the third is for the two color diagram.
[b]All the two color diagram values for stars from 0 to 20 kpc failed.

TABLE 4
RESULTS OF ESTIMATING THE NUMBER OF RICH AND POOR RGB STARS IN RADIAL BINS USING M31CLASS-BASED PDFS

space, as evidenced by the curve of metal-rich stars extending up and to the right of the main concentration of RGB stars. This situation lends itself to an extension of the method developed in this paper to populations consisting of more than two subpopulations. The results of dividing the M31class-based RGB PDF into two distinct PDFs based on [Fe/H$_{phot}$] value are noted in Table /refrichpoortable. This table clearly demonstrates a trend of decreasing metallicity, with one exception in the 60-100 kpc range, in RGB stars with increasing radial distance from M31. Figure 5 further supports this trend and may even explain the exception, as a group of stars with radius greater than 200 kpc appears to have higher metallicities than expected. This analysis suggests a metallicity gradient in M31 RGB stars that varies with radius.

3.3. *Evaluation of PDFs*

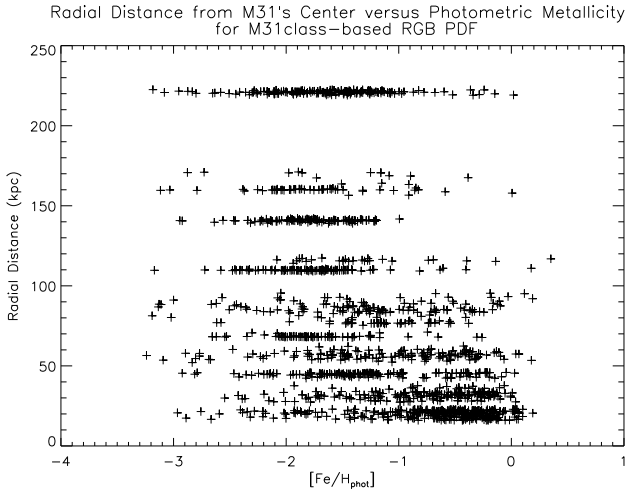

FIG. 5.— A plot of radial distance from M31's center (kpc) versus photometric metallicity for the M31class-based PDF. The plot illustrates the fact that, with increasing radius, the general metallicity of RGB stars in M31 tends to decrease. An exception is the group of stars with radius greater than 200 kpc, which seems to have a greater average metallicity than some of the groups below it.

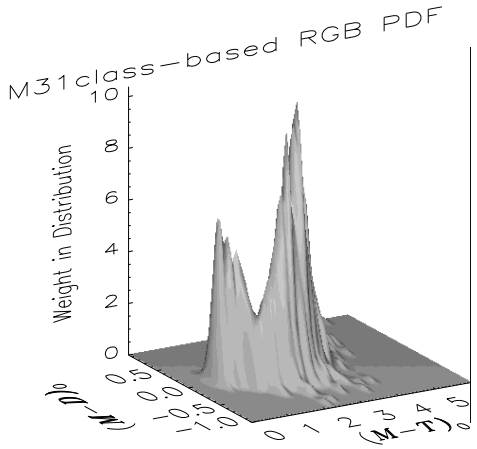

FIG. 6.— A three-dimensional surface plot representing the M31class-based RGB PDF. The PDF is derived from the basis population's distribution in the two-color diagram diagnostic space.

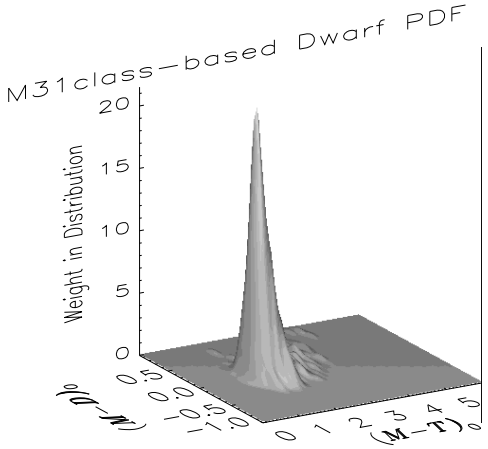

FIG. 7.— A three-dimensional surface plot representing the M31class-based dwarf PDF. The PDF is derived from the basis population's distribution in the two-color diagram diagnostic space.

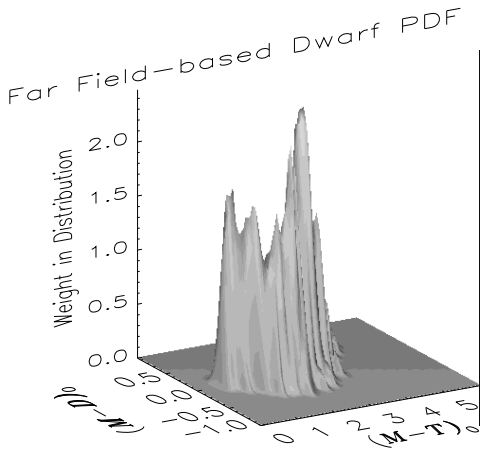

FIG. 8.— A three-dimensional surface plot representing the far field-based dwarf PDF. The PDF is derived from the distribution of stars with large radii ($r > 150$ kpc) in the two-color diagram diagnostic space.

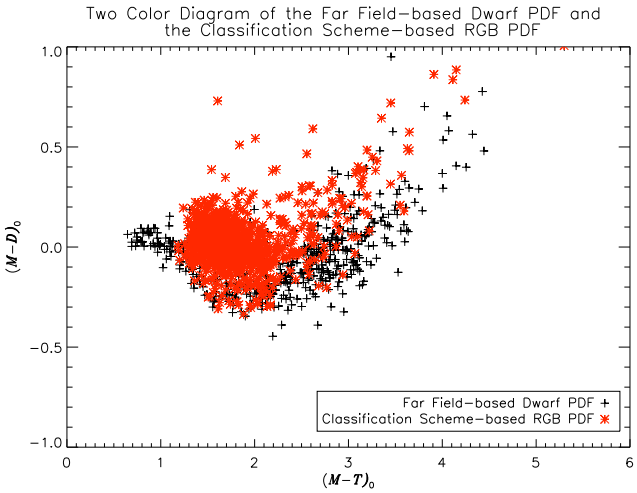

FIG. 9.— A two color diagram of the populations used to define the far field-based dwarf PDF and classification M31class-based RGB PDF. The distinct overlap between the two regions in the area where the far field PDF has an anomalous peak suggests the presence of distant RGB stars in M31.

The selection of PDFs to represent the different subpopulations within a given population is an important consideration for our method. Though we have defined multiple PDFs for both the dwarf and RGB PDFs, not all of them are equally accurate. As some of the PDFs are merely rough approximations of stars of that spectral type, their distributions may be slightly off from the average distribution and they may be contaminated by stars of other spectral types. Though these distributions may be reasonably accurate in cases where no other distributions can be created, the access to the classification scheme developed by Gilbert et al. (2006) provides a more apt approximation of the stellar classes in this case. This method assigns a value denoted M31class to each star ranging from -3 to 3, where negative values indicate dwarfs and positive values indicate giants. As such a classification scheme cannot be developed for all populations to which this method is applicable, we have employed other methods of creating dwarf and RGB PDFs,

namely isolating the populations based on far fields and dwarf galaxies to represent the dwarf and RGB PDFs, respectively. In this section we analyze the shortcomings of these proposed PDFs and suggest new methods to refine their use in the future.

Three-dimensional surface plots of the dwarf PDFs based on the classification scheme and on radially distant fields and the RGB PDF based on the scheme demonstrate that the distant fields are contaminated by RGB stars. While Figure 6 demonstrates that the two color diagram distribution of the M31class-based dwarf PDF is clearly bimodal, Figure 8 displays an anomalous peak in the distribution of the distant field-based dwarf PDF. While observing these two plots alone does not suggest that one is more accurate, Figure 7, a surface plot of the M31class-based RGB PDF, suggests that the peak of the RGB distribution coincides with the anomalous peak of the far field distribution. Figure 9 substantiates this conclusion, as it demonstrates a clear overlap between the far field dwarf and M31class-based RGB populations at the location of interest. This overlap suggests the presence of RGB stars in the far fields of M31. Moreover, the results of applying the method using the distant field-based PDF to represent dwarf stars also suggest the presence of RGB stars. As seen in Table 5, using the contaminated distribution generally underestimates the number of RGB stars in the populations, especially for populations that are more radially distant from M31. Indeed, when comparing the far field-based PDF's results to those of the M31class-based PDF, there are 8 changes of more than 10% in the proportion of RGB stars, 7 of which are decreases in the estimates of RGB stars. Such underestimation is expected for a contaminated dwarf PDF because the normalized distribution represents slightly less than 1 dwarf star and slightly more than 0 RGB stars. Furthermore, this bias is even more pronounced in radially distant populations with many dwarf stars: more dwarf stars entail more overestimation to compensate for the error. The contamination of this population could be rectified by analyzing stars from even more radially distant locations in M31; however, such data were not available for this study.

The two color diagram also revealed shortcomings in the dwarf galaxy-based RGB PDF compared to the M31class-based RGB PDF. Initially, an analysis of Figure 4, a two color diagram of the M31class-based RGB PDF, reveals a curve to the right of the main mass of stars that seems to be a dwarf PDF feature, as indicated by Figure 11. A two color diagram of the dwarf galaxy-based RGB PDF seems to confirm this interpretation, as it does not have this telltale feature. However, a closer inspection of two plots reveals that this curve is characteristic of not only dwarf populations but also metal-rich, interior RGB stars; the dwarf galaxies of M31 are generally positioned far from the galaxy's center and have mostly

Radial Bin	Number of RGB Stars[a]	Total Stars	Proportion of RGB Stars	Percent Change Compared to the M31class-based RGB PDF[b]
0-20 kpc	89/126/0[c]	147/144/0	0.605/0.875/NA	+0.012/+0.008/NA
20-40 kpc	269/324/146	500/499/288	0.538/0.649/0.507	+0/-0.039/+0
40-60 kpc	157/122/101	566/569/564	0.277/0.214/0.179	+0.091/-0.219/-0.101
60-100 kpc	171/3/93	1223/1230/1114	0.139/0.002/0.083	+0.264/-0.974/-0.253
100+ kpc	18/16/19	624/627/624	0.029/0.026/0.030	-0.752/-0.821/-0.577

[a]The entries in this column and the subsequent columns are formatted to give the information for all three diagnostic spaces. The first of the three numbers is for the NA_I Absorption Line Strength versus $(V\text{-}I)_0$ Color plot, the second is for the $[Fe/H]_{spec}$ versus $[Fe/H]_{phot}$ plot, and the third is for the two color diagram.
[b]Defined as $\frac{p_{\rm far}-p_{\rm M31class}}{p_{\rm M31class}}$, where $p_{\rm far}$ is the proportion of RGB stars in a given bin obtained using the far field-based RGB PDF and $p_{\rm M31class}$ is the proportion obtained with the M31class-based RGB PDF.
[c]All the two color diagram values for stars from 0 to 20 kpc failed.

TABLE 5
RESULTS OF ESTIMATING THE NUMBER OF RGB STARS IN RADIAL BINS USING THE FAR FIELD-BASED RGB PDF

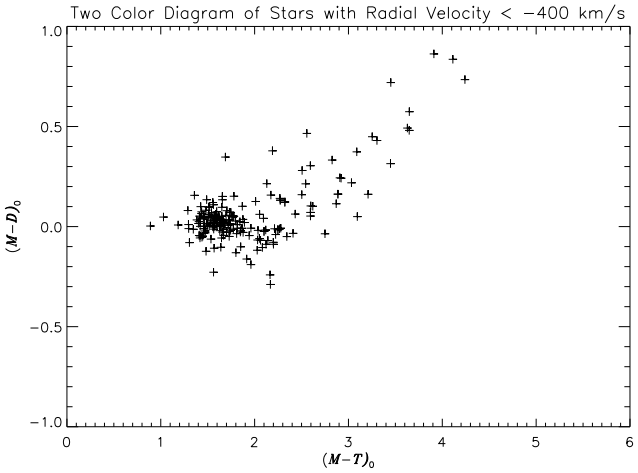

FIG. 10.— A two color diagram of stars with velocities less than -400 km/s. As stars with such low velocities are virtually guaranteed to be RGB stars, this plot demonstrates that the curve above and to the right of the main mass of points is not a sign of dwarf contamination.

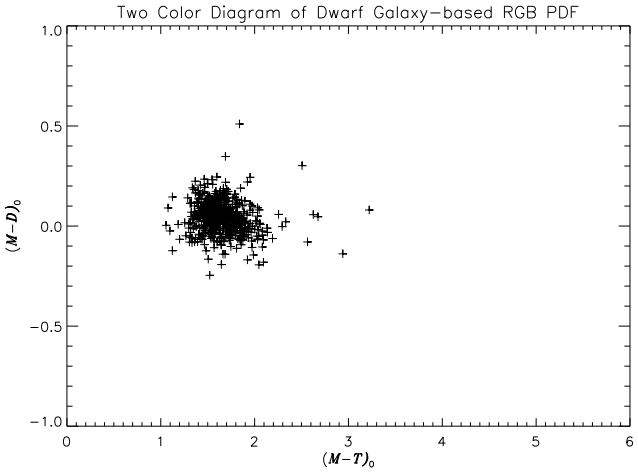

FIG. 11.— A two color diagram of the dwarf galaxy-based RGB PDF. The lack of the curved feature characteristic of high-metallicity stars indicates that it does not accurately represent the RGB population.

metal-poor stars. Indeed, Figure 10 shows that the population of all stars in the database with radial velocity less than -400 km/s exhibits this characteristic curve. At such an extreme velocity, the vast majority of stars are M31 RGB stars; in fact, the classification scheme indicates that only 2 of the 206 stars on the plot are dwarf stars. Splitting the M31class-based population into two subpopulations based on [Fe/H_{phot}] metallicity demonstrates that the curve is characteristic of dwarf stars with high metallicity, a class that the dwarf galaxy-based PDF does not contain. Constructing a separate PDF to represent metal-rich stars and using more than 2 basis functions to represent subpopulations is an improvement to this method and can be done by using stars from the very center of the galaxy. This study did not have sufficient data to construct such a PDF, but Section 3.2 analyzes the implications of representing the RGB stars as two separate subpopulations.

REFERENCES

Gilbert, K. M., Guhathakurta, P., Kalirai, J. S., Rich, R. M., Majewski, S. R., Ostheimer, J. C., Reitzel, D. B., Cenarro, A. J., Cooper, M. C., Luine, C., & Patterson, R. J. 2006, ApJ, 652, 1188

Guhathakurta, P., Ostheimer, J. C., Gilbert, K. M., Rich, R. M., Majewski, S. R., Kalirai, J. S., Reitzel, D. B., & Patterson, R. J. 2005, ArXiv Astrophysics e-prints

Krisciunas, K., Hastings, N. C., Loomis, K., McMillan, R., Rest, A., Riess, A. G., & Stubbs, C. 1999, Uniformity of V minus Near Infrared Color Evolution of Type Ia Supernovae, and Implications for Host Galaxy Extinction Determination

Majewski, S. R., Ostheimer, J. C., Kunkel, W. E., & Patterson, R. J. 2000, ASTRON.J, 2550

Schiavon, R. P., Barbuy, B., Rossi, S. C. F., & Milone, A. 1996, The Near Infrared NaI Doublet Feature in M Stars