First test of a method to detect halo objects in the Milky Way using real data.

Author: Laura Rotger Piñol*
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisor: Teresa Antoja

Abstract: UFDGs are an interesting object of study, as they are systems dominated by dark matter and may offer insight into current problems of the ΛCDM paradigm. In this work, we have focused on preparing some Python routines for the treatment of the next *Gaia* data release, *DR2*, coming up in the spring of 2018. The work consisted mostly in the implementation of a program that divides the sky in cells, and in designing and applying color cuts to the data. These codes were used to select data of several objects in the Galactic halo from the catalogs HSOY, SDSS and GPS1. The data was then used to test, for the first time with real data, a method to detect halo objects. The method is based on the Wavelet Transform (WT) and uses positions and proper motions to detect overdensities. We do a first assessment of the performance of the method on the selected halo objects. Our results show that the method works with the current available data identifying objects such as a globular cluster (*Palomar 5*) and a classical dwarf galaxy (*Draco*), although it does not succeed when trying to identify UFDGs. It is expected that with an improvement on the data precision, the method will work with *DR2*.

I. INTRODUCTION

The predicted number of dark-matter halos in a Milky Way-like galaxy is greater than the number of observed dwarf galaxies. This is one of the most important issues in the ΛCDM paradigm. This is known as the "missing satellite problem". As a possible solution to this problem stands a new population of systems, the Ultra Faint Dwarf Galaxies. These systems have very low luminosity and are dominated by dark matter. The knowledge of their properties may offer insight into numerous issues such as the process of star formation or how the Ultra Faint Dwarf Galaxies (UFDGs) could have contributed to the population of today's Galactic halo[2, 4].

These newly discovered UFDGs appear like an extension of the classical dwarf spheroidals to lower luminosities. UFDGs luminosities are similar to those of globular clusters, although the presence of dark matter in UFDGs marks a clear difference. Massive globular clusters have mass-to-light ratios of up to $(M/L_V) \sim 2$. In turn, UFDGs are clearly dark-matter dominated with $(M/L_V) > 100$ [4]. These dark-matter densities of dwarf galaxies suggest a high redshift collapse for both classical dwarf spheroidals and UFDGs, but the dwarf spheroidals' population apparently continued to evolve. However, the UFDGs are the least chemically-evolved galaxies known. Their abundance patterns indicate that their star formation period was short, and their individual stellar metallicities are as low as $[Fe/H] = -3.7$. Mechanisms that could have caused a short period of star formation include reionization, gas depletion, and supernova feedback[3].

The expected results of the *Gaia* mission, launched in 2013, give way to high expectations in regards to the discovery of new UFDGs. *Gaia* is an ESA mission that is currently gathering data for the construction of a three-dimensional map of our Galaxy, the Milky Way. Over a 5 year mission, *Gaia* will measure positions, parallaxes, and proper motions for objects with magnitude brighter than about 20.7. This is, more than 1 billion objects in our Galaxy and throughout the Local Group. Besides the positional and kinetical data, multi-color photometry will be obtained for all stars, and *Gaia* will also provide astrophysical information based on the photometry[5].

The intention behind this work was to contribute to the preparations in anticipation of the next *Gaia* data release. The main aim here was to find UFDGs using the Wavelet Transform (WT) method we describe in section *II.A*, which was developed using a mock *Gaia* catalog. The intention was to test the method with real data, using the catalogs describedin section *II.B* section. The data selection is described in section *III*. We study how the WT works on different celestial objects: *Palomar 5*, *Draco*, *Bootes*, a simulation from the *Gaia* mock catalog and the object M67 (section *IV*). Our conclusions are described in section *V*.

II. METHODOLOGY

A. Wavelet Transform method

To begin with, we will briefly explain the method we use for detecting UFDGs using both galactic coordinates and proper motions. This method uses the WT, which can be understood as a Fourier Transform, to detect over-densities in both coordinate systems. The method was tested using a model of the sky which includes the *Gaia Universe Model Snapshot* [2] as background, and

---

*Electronic address: `laura.rotger.23@gmail.com`

a simulation of 30000 UFDGs. It is important that the algorithm is able to detect over-densities of different sizes, since the apparent sizes of the UFDGs may vary. We have used a set of three different frequencies with our data for the detection of UFDGs using galactic coordinates, and three different ones for the proper motions. In the model used for the development, a set of four different frequencies is used.

Once the WT has been performed, the algorithm computes the Wavelet Probability (WP), which gives the probability that the detected over-densities are due to Poisson noise. The algorithm has three significance levels that allow us to see how likely an over-density is to be a real detection and not due to Poisson noise. The detections with a low WP are discarded.

B. Catalogs

The data we used for the development of this project came from different catalogs. We started working using the HSOY catalog, but we switched to using the GPS1 and the SDSS catalogs for the better quality of their proper motions and photometry, respectively.

The purpose of HSOY [1] is to provide the scientific community with proper motion data for objects for which Gaia's first release (Gaia DR1) only provided positions. This is achieved for 583 million objects, while Gaia provided positions for 1143 million objects. HSOY was constructed using a method that involves cross-matching different datasets (PPMXL and *Gaia DR1*) and a weighted least-square fit to derive positions and proper motions. PPMXL contains about 900 million objects. The reason why HSOY is much smaller than both its sources is that a given object had to be contained in both catalogs in order to be included. Possible causes are the low angular resolution of PPMXL compared to *Gaia*, and the incompleteness of *Gaia* in some regions of the sky in the first release due to its inhomogeneous coverage of the sky. The other catalogs we used for the last parts of our study were the ninth release (DR9) of SDSS (Sloan Digital Sky Survey), which we use for its photometry; and GPS1, which combines Gaia DR1, PS1, SDSS and 2MASS astrometry to obtain proper motions for 350 million sources.

III. DATA SELECTION

The program we developed read all the data in the catalog (583 million objects) and classified them into packets or cells that cover the whole sky. The program classifies every object into one cell, associating its spatial coordinates to a given 'central coordinate'. These 'centers' are read from a previously created file, which contains the coordinates of the centers of square cells that cover the whole sky. For the purposes of this work, we settled with selecting only some fields of interest, which are mentioned later in this section.

The WT method works using galactic coordinates, and each of the catalogs contains the spatial coordinates and proper motions in the equatorial system. Before classifying the data into its corresponding sky cell, the program changes the coordinates using the *pygaia* Python module. The module contains functions that change both the spatial coordinates and the proper motions into galactic coordinates.

The last step on the preparation of the data consisted on applying a color cut. The aim of this color cut is to leave out the population of stars that we are not interested in, in particular the dwarf stars belonging to the disk, and to include most of the visible population of UFDGs. The implementation consists mainly on checking if an object's color indexes fall into a desired range. The stars belonging to UFDGs that we are able to detect are the brightest ones of their population, which would be the giants. This allows us to isolate the population of UFDGs in a color-color diagram and implement a cut that sets them apart from a big part of the foreground. This cut can be thought of as a polygon in a color-color diagram.

The initial intention was to apply these color cuts using the data from HSOY. We wanted to apply them to the whole catalog in order to reduce the amount of data we had to work with in about 3 orders of magnitude. However, we found that the only Johnson colors in the HSOY catalog were B and V, and they also had very large uncertainties. The conclusion we drew from this was that it would be difficult to find any new objects applying color cuts to this catalog.

In order to continue with our study of UFDGs, we switched from using HSOY to using the other catalogs mentioned earlier, SDSS and GPS1. We compare how our program works implementing new color cuts on some known objects, and then apply the WT method to these. Using *topCAT*, we cross match the data for these known objects from both catalogs and keep the positions and kinematics from GPS1 and the photometry from SDSS. The color cuts applied to the new data were taken from [6, 7] and slightly modified using the classical dwarf *Draco*, the globular cluster *Palomar 5* and the UFDG *Leo II* as reference. We can observe a well-defined sequence in their color-magnitude diagram. Our implementation of the modified cut checks if a given object's colors $G-R$ and $R-I$ fall inside the polygon with the following vertexes in the $G-R, R-I$ space:

$$(0.4, 0.1)\ ,\ (1.2, 0.37)\ ,\ (1.2, 0.54)\ ,\ (0.4, 0.27),$$

while the original cuts proposed in [6, 7] were:

$$(0.4, 0.1)\ ,\ (0.9, 0.28)\ ,\ (0.9, 0.4)\ ,\ (0.4, 0.22).$$

In order to better visualize how the cut would affect the data, *FIG. 1* depicts the area of restriction superposed to the data from *Draco*.

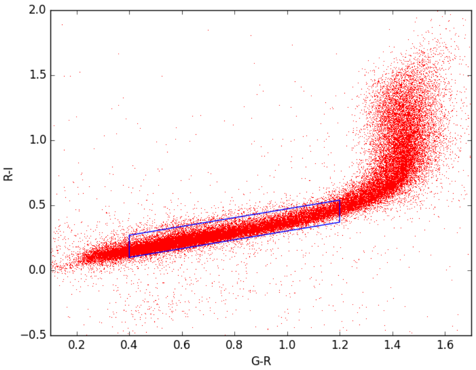

FIG. 1: Color-color diagram of *Draco*'s data (red) and the implemented color cut (blue).

IV. RESULTS

We have tested the efficiency of the color cut we implemented on the following objects: *Palomar 5*, a globular cluster; *Draco*, a classical dwarf spheroidal; and *Bootes* and *Leo II*, which are Ultra Faint Dwarf Galaxies. To begin with, we apply a manual selection of the stars belonging to the object we are studying in a spatial coordinates diagram. Afterwards, we trim the selection observing its shape in a color-magnitude diagram. We can see in the middle diagram of *FIG. 2* that there is a region with an overdensity that curves right in this space. This is the section of the color-magnitude diagram that we take to limit our data when applying the color cut. In the diagram to the right of *FIG. 2* (proper motions), we can appreciate how the selected sequence stays an overdensity.

Object	$N_O/N_{B\ \mathrm{i}}$	$N_O/N_{B\ \mathrm{f}}$	f/i
Palomar 5	$2.44 \cdot 10^{-3}$	$3.98 \cdot 10^{-3}$	1.6
Draco	$7.93 \cdot 10^{-3}$	$1.44 \cdot 10^{-2}$	1.8
Leo II	$6.14 \cdot 10^{-3}$	$1.48 \cdot 10^{-2}$	2.4

TABLE I: Comparison of the ratio of the number of stars that belong to the object of study, N_O, to the total number of stars, N_B, before (i) and after (f) applying the color cut.

TABLE I shows that the ratio of the number of stars in the manual selection, N_O, to the total number of stars, N_B, increases after applying the color cut. This implies that the cut does indeed dispose of many of the foreground stars. The larger the ratio f/i, the better the effect of the color cut. This ratio is greater than 1 for every case studied. We do not show the results from the remaining object of study, the UFDG *Bootes*, because we could not perform a manual selection on it. When plotting the right ascension to declination diagram, we could not see any hint of an overdensity, nor could we detect a sequence similar to that of the other cases in the color-magnitude diagram. The fact that the color cuts works on all of these except *Bootes* could lead us to believe that perhaps the cut does not work for UFDGs. However, if we observe the f/i ratio for *Leo II*, which is another UFDG, we find that the cut works well. The reason why we can not find *Bootes* is that it is too faint.

Below we present an example of the manual selection from *topCAT* in positions, color-magnitude and proper motions diagrams (*FIG. 2*). The selected population is represented in blue and the rest of the stars are represented in red. Although we may have selected some stars that do not belong to the object of interest, we won't pay much attention to this fact. Choosing any stars that do not fit in the polygon with which we have implemented the color cut would only make our results worse. The ratio pre-cut (N_O/N_Bi) would increase and the ratio post cut ((N_O/N_Bf) would stay the same, so f/i would only decrease. Our results would only improve with a more thorough selection of the stars that actually belong to the object of study. This case corresponds to the classical dwarf spheroidal galaxy *Draco*. We have chosen this example to illustrate the selection process because the sequence in the color-magnitude diagram is clear, although there is a 'hole' in the spatial coordinates diagram. This means that in either one of the catalogs we used to obtain these images (SDSS and GPS1) there was a missing section of the sky. However, the lack of this portion does not affect our study. Since the missing section is far from our region of interest, he effect this 'hole' would have on the results presented in *TABLE I* would only be positive, as it would only reduce the ratio N_O/N_B pre-cut.

Once we have checked that the color cut works in our favor, we take the reduced data and use it to find overdensities using the WT method.*Draco* is pictured in *FIG.3*, with well determined spatial coordinates and proper motions. In *FIG. 3* are represented the results in the spatial coordinates and proper motions planes from the same case of study, *Draco*. The darker the blue, the more dense the region.

We can see that there are crosses of different colors. Each cross marks an overdensity with a different level of significance. This WT method will only consider peaks with high WP, as we mentioned in section *II.A*. The green crosses mark peaks with $WP \geq 99.7\%$. The orange crosses mark peaks with $95.4\% \leq WP < 99.7\%$, and the red ones peaks with $68.2\% \leq WP < 95.4\%$. These percentages are similar to $> 3\sigma$, $2\sigma - 3\sigma$, $1\sigma - 2\sigma$, respectively, in the Gaussian case.

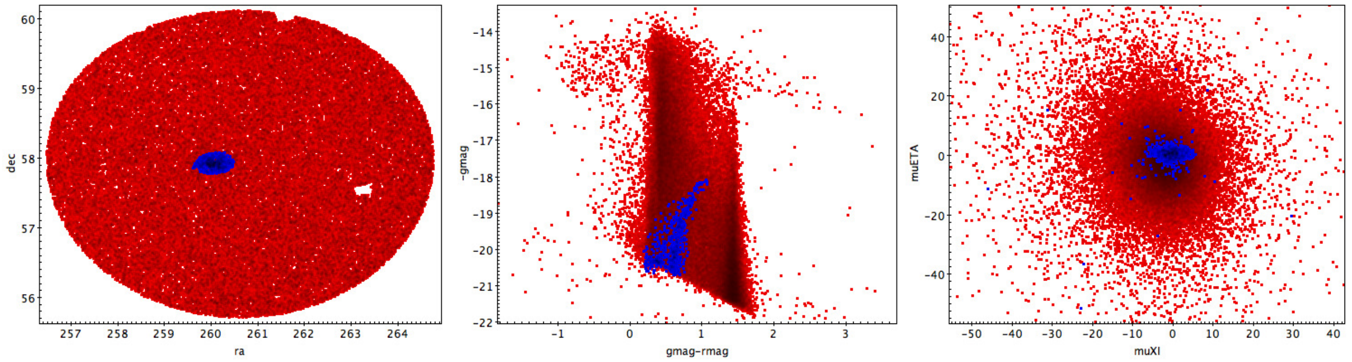

FIG. 2: Positions, color-magnitude and proper motions diagrams of Draco (blue) selected manually in contrast with the background (red).

As can be seen in *FIG. 3*, the WT method detects overdensities in three different scales. A peak is best detected when the scale is similar to the apparent size of the object. *TABLE II* shows the results of the WT for all objects we studied. We see in this table that that neither of the UFDGs is well detected in the proper motion plane. This is disappointing, since the whole method is based on the detection of these systems using proper motions. However, we have to take into account that the data we used for the study of these objects is not comparable to the quality of the data that is expected from the next Gaia release. These proper motions have been derived from the positions of these stars. These stars all have high magnitudes, that is, low luminosity. This makes any measure more prone to errors than that of a brighter star. We can appreciate that there are overdensities in the proper motion plane that are detected, although it should be taken into account that the one in the largest scale may also only be representative of the centroid of the proper motions diagram.

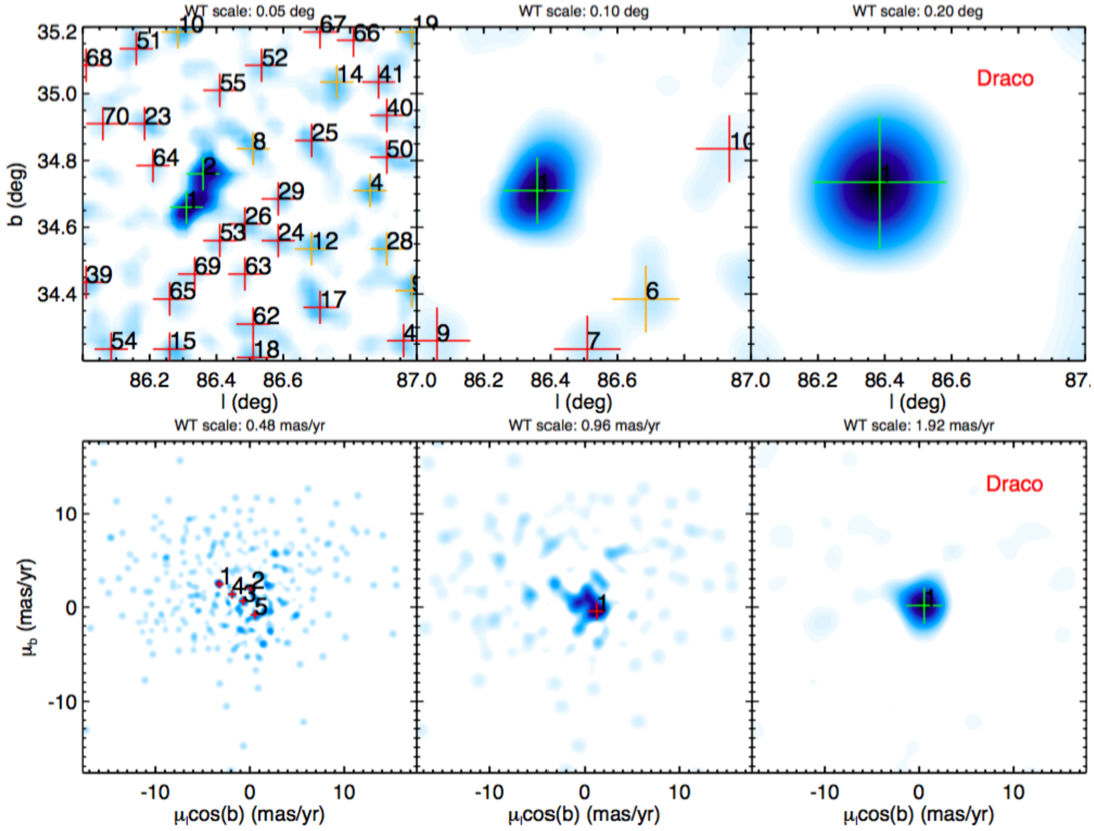

FIG. 3: Detections of the UFDG Draco in the spatial coordinates plane (above) and the proper motions space (below) in three different scales.

Object	Catalog	Type	σ_{lb}	σ_{pmlpmb}
Simulation	Mock	UFDG	> 3	> 3
M67	HSOY	OC	> 3	> 3
Palomar 5	SDSS	GC	> 3	> 3
Draco	SDSS	DS	> 3	> 3
Leo II	SDSS	UFDG	> 3	$1 < \sigma < 2$
Bootes	SDSS	UFDG	$1 < \sigma < 2$	X

TABLE II: Level of significance of the peak detected for each object in the spatial coordinate space (σ_{lb}) and the proper motion plane (σ_{pmlpmb}). An X indicates a lack of detections.

The acronims in the *Type* column in *TABLE II* correspond to Ultra Faint Dwarf Galaxies (UFDG), an Open Cluster (OC), a Globular Cluster (GC) and a classical Dwarf Spheroidal (DS). We have selected the most relevant scale when giving the significance values.

It is also worth mentioning that the data in the last column of *TABLE II* may not strictly be representing the real significance of the peaks in the proper motions plane. The WT method detects overdensities, but it would be necessary to cross-match the points detected in the position space with the ones detected in the proper motions space in order to be able to consider them relevant and true.

V. CONCLUSIONS

We have implemented a selection tool for the classification of all objects into their corresponding parcel in the sky. We have also implemented a method for the selection of the stellar objects of interest, the UFDGs, based on their positions and proper motions, and designed and applied color cuts based on [6, 7], modified for our case and more complete. As we have mentioned, the initial intention was to apply these cuts on the HSOY catalog in order reduce it and find new UFDGs using the WT method. However, we found the errors in the HSOY photometry and its proper motions data to be too large for a good application of the program. The data from *Gaia DR2* is expected to be much more accurate, and this color cut could be applied to the next catalog obtaining satisfactory results. In order to test how this cut would work on a catalog with less uncertainties in photometry, we combined the SDSS and the GPS1 catalogs and obtained satisfactory results (*TABLE I*) when studying a group of known objects. Finally, we tested the WT transform method with these objects and found that the method works well for a simulation from the Gaia mock catalog, *M67*, the globular cluster *Palomar 5* and the classical dwarf spheroidal *Draco*. However, the method is not successful in finding peaks for the two UFDGs studied, *Leo II* and *Bootes*. Although the method successfully detects peaks in the spatial coordinate plane, it is not capable of detecting the same peaks in the proper motion plane. As we mentioned earlier, the cause for the difficulties found in the proper motion plane may come from the large error associated to this variables. This error in turn is due to the high magnitudes of these stars. This problem should decrease with Gaia DR2, having much more precise proper motions that will allow the WT method to detect relevant overdensities in both spaces.

A possible solution for the problem we have experienced when trying to detect peaks in the proper motion plane would be to associate the detections in the spatial coordinates plane. That is, if a peak is detected in the spatial coordinates plane but not in the proper motions plane, the WT method could be altered in order to seek the difference in distributions between the stars that belong to the previously detected peak and the rest. In the proper motion plane the distribution of the stars detected in the position space is indeed different from the total. This aspect could be worked on in order to improve future results.

Acknowledgments

I would like to express my gratitude towards my advisor, Dr. Teresa Antoja, for the time she invested in me and for her continued attention and guidance; and towards Dr. Carme Jordi, for her advice and supervision. I would also like to thank my family and friends, for their support and encouragement during these past four years.

---

[1] Altmann, M., Roeser, S., Demleitner, M., Bastian, U., & Schilbach, E. 2017, A&A, 600, L4
[2] Antoja, T., Mateu, C., Aguilar, L., et al. 2015, MNRAS, 453, 541
[3] Brown, T. M., Tumlinson, J., Geha, M., et al. 2014, Memorie della Societa Astronomica Italiana, 85, 493
[4] Brown, T. M., Tumlinson, J., Geha, M., et al. 2012, ApJL, 753, L21
[5] Jordi, C., Gebran, M., Carrasco, J. M., et al. 2010, A&A, 523, A48
[6] Koposov, S., Belokurov, V., Evans, N. W., et al. 2008, ApJ, 686, 279
[7] Willman, B., Dalcanton, J., Ivezić, Ž., et al. 2002, AJ, 123, 848