UNIVERSITY OF GRONINGEN

BACHELOR THESIS

ASTRONOMY

Creating realistic mock galaxy samples for the APERTIF HI surveys

Author:
Bas ROELENGA

Supervisors:
Thijs VAN DER HULST
Manolis PAPASTERGIS

29 July, 2016

Abstract

In this project I create realistic samples of mock galaxies for two surveys to be carried out by APERTIF in the 21cm line of atomic hydrogen (HI): the shallow and medium-deep survey. The samples of mock galaxies are created according to the technical specifications for the APERTIF system, and the current design of each survey. Galaxies are given realistic HI properties (such as HI mass, profile width and HI disk radius) by using the distribution of galactic HI masses and HI profile widths (Mass-Width function) determined from previous observations, as well as other measured galactic scaling relations. Furthermore, the number density of mock galaxies follows the actual large structure observed in the nearby universe. The mock APERTIF samples created in this work can be a valuable tool for aiding the final survey design, as well as evaluating the merits of synergies with other surveys and study the effects of cosmic variance as a function of survey parameters.

Contents

1 Introduction

Currently the Westerbork Synthesis Radio Telescope(WSRT) consists of 14 radio telescope which all have a single feed system. This single feed system causes the radio telescopes to have a limited field of view which make it difficult and very time consuming to observe large areas of sky with high sensitivity. The idea is to replace 12 of these single feed systems with arrays of detectors, also called a Focal Plane Array (FPA). This technology makes use of this array of detectors which can form multiple beams on the sky. The system that is going to replace the single feed systems in the WSRT is called APERTIF (APERture Tile In Focus). This system makes use of this phased array technology and has been developed by ASTRON and is implemented on the radio telescope. The purpose of APERTIF is to perform HI and continuum surveys of the northern sky and to efficiently find pulsars and transients. When all of the 12 dishes are equipped with the FPA's its goal is to have 37 beams on the sky simultaneously which will generate a field of view of 8 square degrees. The APERTIF survey will consist out of two surveys, the shallow and medium-deep survey. The shallow survey will focus more on detecting galaxies in the nearby universe and the medium-deep survey, which has a higher integration time per pointing, will determine more detailed properties of galaxies.

In this thesis I will create a sample of mock galaxies which are predicted to be observed by APERTIF for the two HI surveys. We use the existing measurements of the distribution of HI properties of galaxies to determine the number of galaxies on the sky. Since this measurement has an overdensity of zero we will include the effects of large-scale structure in the survey volume using existing surveys. These samples of mock galaxies can than later be used to help finalizing the final survey design. Every survey that will be carried out by APERTIF will consist of its own patch of the sky. Therefore a mask will be created to filter out the galaxies that will be observed for that specific survey. The sample of galaxies itself will be created by combining the Mass-Width Function, from which numbers of galaxies will be obtained based on the properties of the APERTIF system, with the large scale structure obtained from two catalogs, a catalog of the Local Volume and the 2MRS catalog. This will create a realistic distribution of galaxies on the sky. Finally the galaxies will be given some general properties such as HI mass, profile width and the radius of the HI disk. Combining all this will leave us with a realistic sample of mock galaxies which are predicted to be observed by APERTIF. This sample of mock galaxies will contain the amount of galaxies and dwarf galaxies that are predicted to be observed in each of the surveys. But the mock sample will also contain the distribution of galaxies in mass, the distribution of galaxies in distance and the profile width of galaxies as a function of distance. The mock sample of galaxies can therefore be used to use the different surveys of APERTIF for specific purposes (if a study wants to observe dwarf galaxies, APERTIF can carry out the survey, which contains the most dwarf galaxies as predicted by the mock sample of galaxies).

2 The APERTIF HI surveys

To now what can be detected with the newly installed system, I will first have a look at what APERTIF actually is, specifically what its technical specifications are and what areas of the sky it will observe.

2.1 Technical specifications of APERTIF

The FPA system that has been developed for APERTIF will have a frequency range from 1 to 1.75 GHz. The total bandwidth will be 300 MHz. Other important technical aspects of the APERTIF system are listed below in table 1. These values are relevant for the creation of the sample of mock galaxies, because these values determine the sensitivity limit for the APERTIF surveys.[1]

System Temperature	T_{sys}	70 K
Antenna Effective Area	A_e	392.7 m^2
Number of polarizations	N_{pol}	2
Number of antennas	N_a	12
Channel bandwidth	$\Delta\nu$	12.2 kHz

Figure 1: Technical specifications for the APERTIF FPA.

Figure 2: The Focal Plane Array used in APERTIF

The FPA itself consists out of a dual-polarised Vivaldi array which has 11 x 11 x 2 elements as shown in figure 2. The system temperature of the FPA ($T_{sys} = 70$ K) is higher than the single feed system ($T_{sys} = 30$ K) that is now installed in the WSRT. The reason for this is that the physical size of the FPA is now so large that it cannot be cooled cryogenically. On the other hand, this loss of sensitivity is partially compromised by the fact that the aperture efficiency of the FPA is higher than that of the single feed system. Another advantage of the FPA over the single feed system is that the FPA suffers much less from the effects of standing waves, making the calibration of Westerbork data easier now.[1]

2.2 The survey layout

There are two surveys planned to be carried out by APERTIF: the shallow survey and the medium-deep survey. The current plans regarding the sky coverage of surveys can be seen in figure 3. Both of the surveys are divided into two subregions, one will be in the fall sky (RA = 22h - 5h) and the other in the spring sky (RA = 10h - 17h). By dividing the footprint of the two surveys into these regions located on opposite sides of the sky APERTIF can observe around the clock.
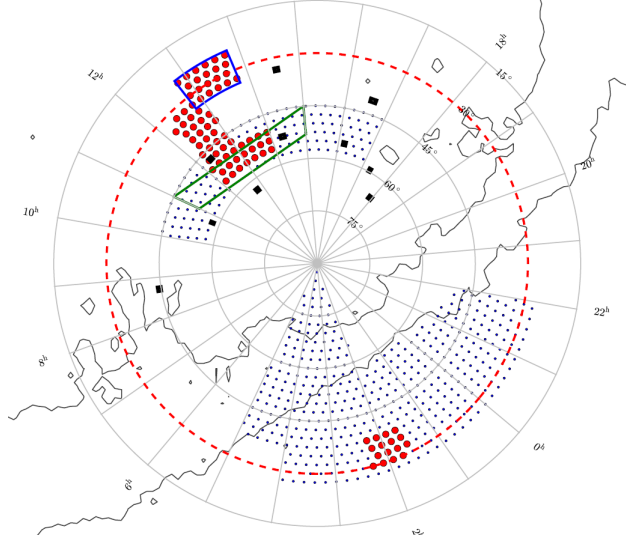
Figure 3: The planned survey coverage for the shallow and the medium-deep APERTIF surveys. Red dots represent the area to be surveyed by the medium-deep survey and the black dots indicate the area which will be surveyed by the shallow survey.[2]

2.2.1 The shallow survey

The shallow survey is optimized for the largest possible sky coverage. Since the total survey time is fixed the trade off between survey area depth survey means that the shallow survey will be less sensitive that the medium-deep survey. Because the survey is over such a large area, the shallow survey will detect more galaxies in the nearby universe. The consequence of this is that the survey sample will include more low mass galaxies and more galaxies with spatially resolved HI data. The shallow survey will provide a measurement of the gas content of galaxies over a large range of environments, which will allow for more studies (such as the gas content of the galaxies as a function of luminosity), and it will provide the width of the HI profiles, which allow for the mass of the galaxies to be determined. The characteristics of the shallow survey are listed in table 1.

2.2.2 The medium-deep survey

The medium-deep survey will have a greater sensitivity than the shallow survey. The reason for this is that the medium-deep survey will survey over a smaller area of the sky and will have a larger integration time per pointing than the shallow survey. This allows the medium-deep survey to detect low surface brightness HI emission, and determine more detailed morphologies and kinematics of HI gas in galaxies. The characteristics of the medium-deep survey are listed in table 1.

Survey	shallow	medium-deep
Integration time (h)	12	7 x 12
Sensitivity (mJy/beam)	0.65	0.25
Sky coverage (deg^2)	3050	450

Table 1: The characteristics of the shallow and the medium-deep surveys.[3]

3 Distributing mock galaxies on the plane of the sky

Both of the surveys, as described in the previous section, will only observe parts of the northern sky. To recreate the survey sky coverage we take the following steps. First we distribute galaxies uniformly over the entire northern sky. We then create a sky mask that can be used to select the mock galaxies that are located within the planned survey footprint of figure 3.

3.1 Uniform sky distribution over whole northern sky

In order to test the algorithm that the mask uses a sample of galaxies needs to be created. This sample will consist of galaxies uniformly generated across the sky. Only galaxies in the northern sky are generated since the survey will also only take place in the northern sky. The sky can be seen as a sphere, this means that the limits for only observing the northern sky are 0 to 360 degrees RA and 0 to 90 dec. The declination is defined in such a way that the Dec coordinate at the north and south equatorial pole is 90 and -90 respectively, while at the equator is 0.

To generate the galaxies each of the galaxies needs to be given a value RA (α) and a Dec (δ). This coordinate system represents the sky as a 2D plane (it is a spherical coordinate system). For a population that is randomly distributed on the sky, the coordinates are generated in the following way: the RA coordinates can be generated using a uniform distribution between the limits. But the Dec coordinates need to be corrected by the fact that we are looking at a sphere. In other words, this means that at the equator there are more galaxies in a given RA range than at the poles. This effect is included in equation 2.

$$\alpha = uniform(0, 360) \tag{1}$$

$$\cos(\delta) = uniform(0, 1) \tag{2}$$

The uniformly distributed population of galaxies in the RA, Dec plane can be seen in the left panel of figure 4, while the projection onto the equatorial plane can be seen in the right panel of figure 4. In this figure 1000 galaxies are generated within the limits that were mentioned before.
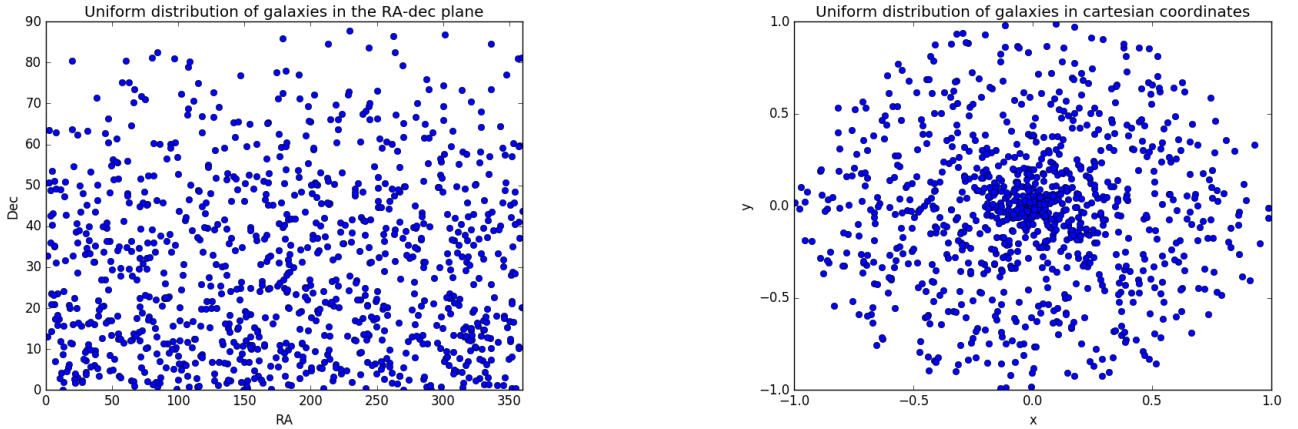
Figure 4: Left Figure: Distribution of a spatially uniform population of 1000 galaxies in the RA, Dec plane. Right Figure: Projection of the same galaxies on the equatorial plane.

3.2 Creation of a sky mask for the APERTIF surveys

3.2.1 Determining the survey coverage

Creating the mask consist of two steps. First the points of the mask need to be generated and then these points can be used to create the mask. The points that are needed to create a mask can be extracted from the APERTIF survey layout plans[2] as seen in figure 3. Using a plot digitizer program points can be selected on this map and RA-dec coordinates are generated for use. Since it needs to be checked if a galaxy is within the mask or not more points are needed to obtain a more precise result. This can be done be linearly interpolating between the points. There will be two surveys, namely the medium-deep survey and the shallow survey. The medium-deep survey consists of two disjoint areas and the shallow survey also consists of two disjoint areas. One practical complication is that one of the areas covered by the shallow survey crosses the RA = 0 line, this causes an abrupt switch of RA values from 0 to 360. For this reason, the second masks of the shallow survey has been divided into two parts. The masks for these surveys are shown in the figures (figure 5 and figure 6) below.
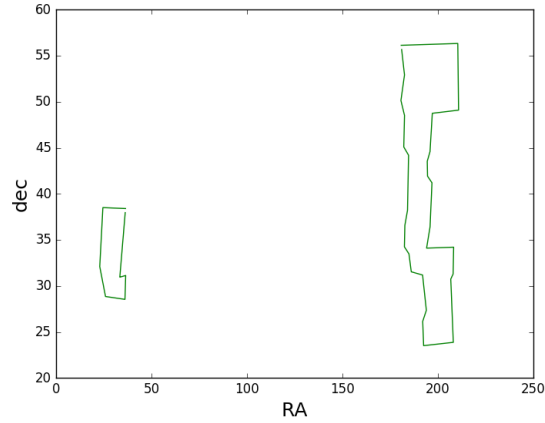
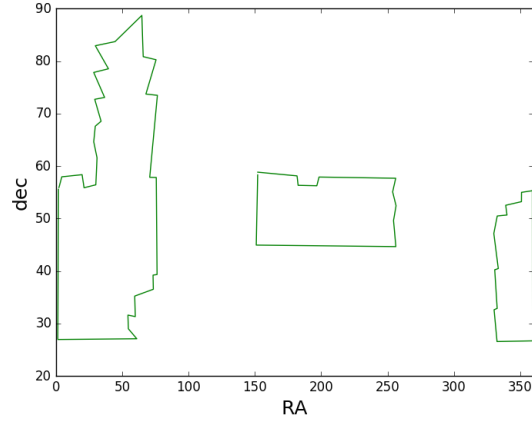**Figure 5:** Area coverage for the medium-deep survey.

Figure 6: Area coverage for the shallow survey.

3.2.2 Constructing the mask

Once the edges of the survey coverage have been determined the construction of the mask can begin. First off, to increase the precision of the mask, points are generated between the edges using linear interpolation. The next step is to create a function to check whether a galaxy is in the mask or not. This can be done by using a odd-even rule and can be explained in a few steps(The following steps are also illustrated in figure 7).

- Step 1: The program checks if the RA and dec coordinate of the galaxy (which were created in Sec 3.1) is between the minimum and maximum value of the mask. If this is the case the program continues to step 2. If not, the galaxy is not in the mask and will be discarded.

- Step 2: The dec coordinate of the galaxy is used to find out how many times this coordinate is between 2 dec coordinates of the mask. If the dec coordinate is found to

lie between these 2 coordinates the RA coordinate belonging to these 2 dec coordinates are saved for using them in the next step.

- Step 3: Now the odd-even rule comes into play. The program checks how many times the RA coordinate of the galaxy can pass the saved RA coordinates of the mask. If it passes the only coordinate it means that the galaxy is in the mask, if it passes 2 coordinates the galaxy is outside of the mask. This rule can also be used to check if a galaxy is in a mask if the mask has a gap in the middle of it. Thus if an galaxy has passed the mask an even amount of time the galaxy is outside of the mask but when it has passed the mask a odd amount of times it is inside the mask.
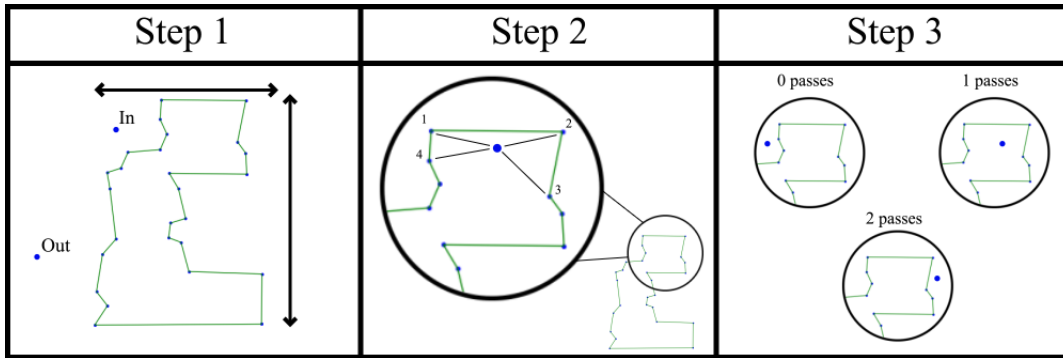

Figure 7: Illustration of the working of the mask.

With this algorithm one can determine from a sample which of the galaxies can be used and which can be discarded. Using the mask in figure 5 and 6 the following results are obtained which are visible in figure 8 and 9. The uniform distribution that was created in section 3.1 was used to obtain these results for the mask.

To use the mask in the catalog that will be generated a true-false table needs to be created. This means that at each RA and dec coordinate, at a certain interval, it will be determined if that coordinate is inside or outside the mask. This way a program creating the catalog only needs to check if the coordinates of the galaxies are at a true or false location to decide whether to keep them.

Figure 8: 1000 galaxies uniformly distributed across the sky, green galaxies are in the mask and the blue galaxies are outside the mask. The masks are from the medium-deep survey.
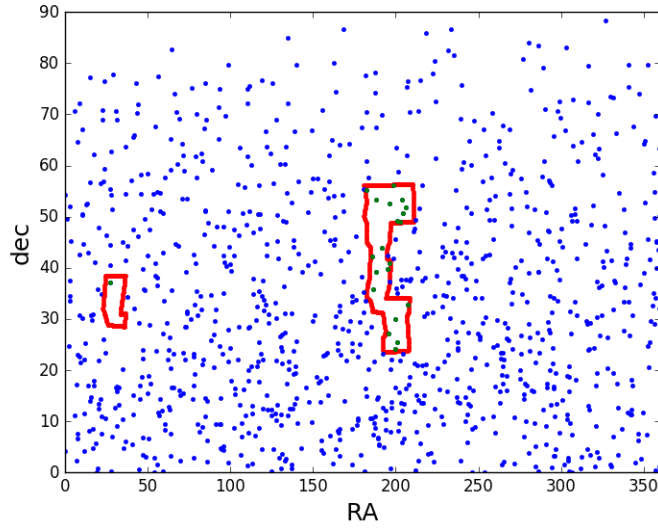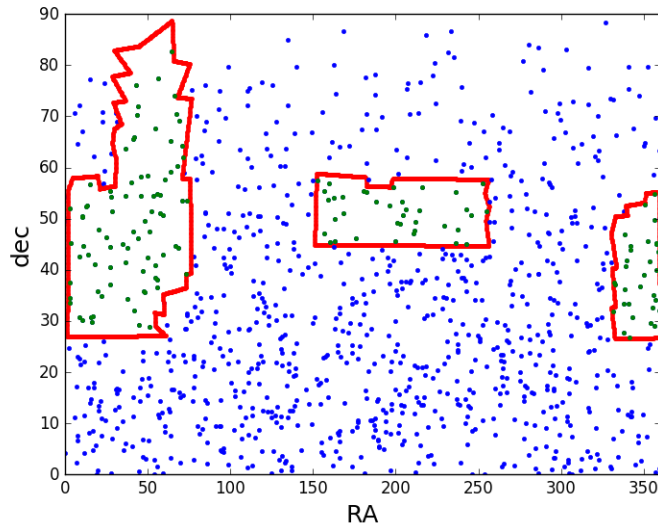

Figure 9: 1000 galaxies uniformly distributed across the sky, green galaxies are in the mask and the blue galaxies are outside the mask. The masks are from the shallow survey.

10

4 Determining the number of APERTIF detections as a function of distance

The number of detections made by APERTIF depends on two things. First using a constant survey solid angle, the volume increases as distance increases. Second the constant sensitivity of the surveys implies that only galaxies with a larger HI mass will be detected as the distance increases. To calculate the number of expected detections made by APERTIF we need to calculate the sensitivity of both surveys and take the distribution of HI mass and the HI profile width of galaxies into account. In this section the number of detections will be calculated for all parts of the sky using the HI mass-width function and the properties of the APERTIF system.

4.1 Survey sensitivity - the radiometer equation

In this section we calculate the sensitivity of the APERTIF surveys using the properties of the APERTIF FPA. The sensitivity of the FPA can be calculated directly using the radiometer equation for an interferometer and the properties of the APERTIF survey in table 1. The receiver noise, σ, for a single frequency channel is given by:

$$\sigma = \frac{2k_b T_{sys}}{A_e \sqrt{N_a(N_a - 1)N_p \Delta\nu\tau}} \tag{3}$$

In the equation above, T_{sys} is the system temperature, A_e is the effective area of one antenna, N_a is the number of antennas, N_p is the number of polarizations, $\Delta\nu$ is the bandwidth of a single channel and τ is the integration time.

Equation 3 is used to calculate the receiver noise in a single channel. However, the HI profile of a galaxy will span more than one channel, with the exact number depending on the spectral width of the galactic HI lineprofile. The minimum HI flux needed to detect a galaxy whose profile spans N spectral channels is:[1]

$$S_{min} = \frac{S}{N}\sqrt{N}\sigma\Delta\nu \tag{4}$$

Here $\Delta\nu$ is the bandwidth of a single spectral channel, σ is the noise over this bandwidth, and (S/N) is the signal to noise ratio required for a detection. The number of channels can be written as the total spectral width of the galactic profile divided by the spectral width of one channel. Since the width of a galaxy's profile is customarily expressed in terms of velocity, the Doppler equation allows us to convert profile width to a frequency range. This is how the velocity width of a galaxy and the bandwidth are related.

$$\frac{\Delta\nu}{\nu} = \frac{\Delta v}{c} \tag{5}$$

Since we are only looking at HI, thus the 21 cm line, ν can range from 1130 to 1420 MHz.

$$N = \frac{W}{dv} \tag{6}$$

Using both equation 5 and equation 6 equation 4 can be rewritten as equation 7

$$f = \sqrt{W \frac{c\Delta\nu}{\nu}} \sigma \tag{7}$$

In order to consider a signal as a detection we require that it be detected with a signal-to-noise ration of $\frac{S}{N} > 6$. Equation 7 now allows us to determine the minimum HI flux that a galaxy with a given profile width must have in order to be detectable by the survey. This will prove useful in the next section in order to calculate the minimum HI mass detectable at a given distance, for a galaxy with a given profile width.

4.2 Survey selection function - the HI mass-width function of galaxies

In the previous section we have calculated the minimum HI Flux that a galaxy with a given profile width W needs to have in order to be detectable by the APERTIF surveys. Given a distance, d, we convert the minimum flux requirement into a minimum HI mass requirement through the following equation:

$$M_{HI,min}(M_{sun}) = 2.356 * 10^5 \, D(Mpc)^2 \, S_{HI,min} \, (Jy\,km/s) \tag{8}$$

Using equation 27 the minimum HI mass can be plotted as a function of the HI profile width as shown in figure 10.
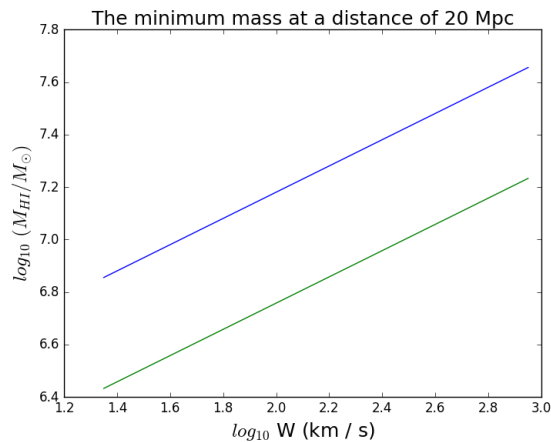

Figure 10: The minimum mass as a function of the log of the velocity width at a distance of 20 Mpc, the green line represents the medium-deep survey and the blue line the shallow survey.

Using number densities generated by the MWF for ALFALFA galaxies[4][5] an actual number of galaxies in the sky can be generated. The table that contains the number densities of galaxies is a 2D array with on one axis the log of the velocity width and on the other the log of the HI mass, which can be seen in figure 11. The log of the velocity width runs from 1.35 to 2.95 (17 bins) and the log of the HI mass runs from 7.1 to 10.9 (20 bins). Using the minimum mass calculated at each velocity width and distance the number densities of

galaxies can be extracted from this table. It is important to note that the APERTIF survey will be more sensitive than the ALFALFA survey, which means that galaxy below a log HI mass of 7.1 can also be detected. But since the ALFALFA survey does not provide any data on the number densities below 7.1, we will be showing here the results for galaxies with a log HI mass higher than 7.1 . In appendix B a method is described to obtain the number densities for HI mass values below 7.1, this is done by extrapolating the MWF of the ALFALFA survey.
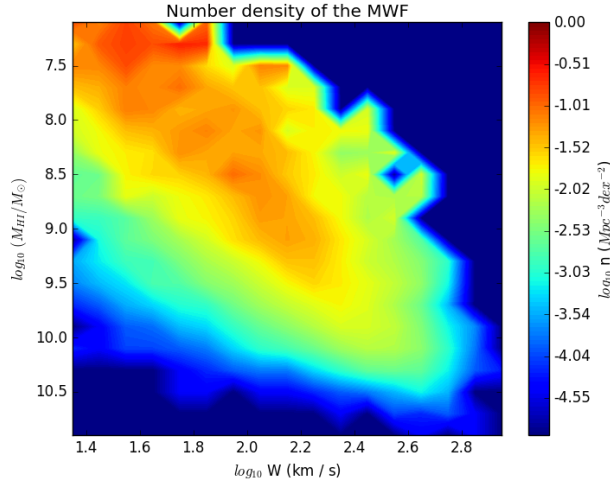

Figure 11: The log of the number densities of the MWF data

Equation 9 allows us the calculate the number of galaxies at each position in this data set.

$$N = n(D)R^2 dR d\Omega \tag{9}$$

where,

$$d\Omega = d\alpha \cos \delta \, d\delta \tag{10}$$

$$n(D) = \int_{W_{min}}^{W_{max}} \int_{M_{HI,min(W,D)}}^{M_{HI,max}} \phi(M_{HI}, W) \, dM \, dW \tag{11}$$

In apendix A the minimum detectable mass only depends on the distance by setting a detection limit. We use the properties of the Focal Array Plane of the APERTIF survey to calculate the sensitivity of the APERTIF surveys. For an HI survey, the minimum HI flux is determined by the spectral width of the HI profile of a galaxy. At any given distance, the variable minimum flux will give rise to a minimum detectable HI mass which can be calculate through 27. The minimum mass obtained will then be used to find the expected number of detectable galaxies using the appropriate integration limits of the mass-width function.

The number of galaxies are calculated within small three-dimensional cells. The reason for creating a 3D grid of cells for our calculation is that using these cells will enable us to take

density fluctuations in the survey volume caused by the large-scale structure into account to get the number of detectable galaxies. The limits of these three-dimensional cells are shown in table 2. As table 2 shows, each cell has an angular size of 5 deg x 5 deg in RA, Dec and spans 5 Mpc in distance out to a maximum of 300 Mpc. The number of galaxies are generated in two steps. First off it needs to be known which number densities of the MWF table can be used. For example, if the log of the minimum HI mass at a certain distance is 8.1 than the number densities in the table for mass bins below 8.1 cannot be used. Since both the distance and the velocity width are known the right number densities can be obtained.

	unit	start	end	bin size
RA (α)	deg	0	360	5
dec (δ)	deg	-90	90	5
R	Mpc	0	300	5

Table 2: Properties of the 3D grid used to create the mock APERTIF samples.

Second, the program loops over every bin, as defined by the limits in table 2. This means that there are 155520 bins for which the number of galaxies need to be calculated. Solving equation 9 for every bin gives this number. The minimum HI mass at a certain distance is not constant in every bin, since it also depends on the velocity width. This causes the minimum HI mass to increase with the velocity width as seen in the left figure of figure 12 . Integrating over all the number densities above this line gives the number of galaxies in this specific cell.

Once all the data is obtained it is saved into two tables which can later on be used to create a sample of mock galaxies. The first table contains the log of the minimum HI masses for each velocity width bin and for each distance. This table can be used to draw realist HI masses from the MWF of the ALFALFA survey later on. The second table contains the number of galaxies per bin. This table is saved in such a way that each row contains the distance, RA, Dec, volume and the number of galaxies of one cell.
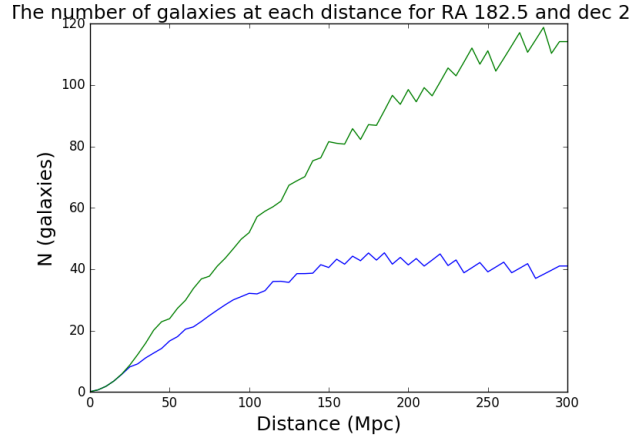
Figure 12: The number of galaxies for a specific cell(RA, dec) at each distance, the green line represents the medium-deep survey and the blue line the shallow survey. It however is important to note that a survey mask has not been applied. This means that in the end the medium-deep survey will not detect more galaxies than the shallow survey.

5 Large scale structure

The last piece of information that needs to be added to create a realistic sample of mock galaxies is the large scale structure. In the previous section it is assumed that the overdensity in each bin is equal to zero, thus there is no large scale structure involved. The known structure in the nearby universe can be described by calculating the overdensities at every position in the sky. Then these overdensities can be applied to the number of galaxies calculated assuming a homogeneous universe (see Sec 4.2) in order to create a more realistic mock galaxy sample.

We obtain the overdensities of the large scale structure by using two external catalogs of galaxies. We use the first to determine the overdensities for the nearby universe (out to about 10 Mpc), and the second, to determine the overdensities out to about 300 Mpc. We describe the two catalogs in detail in two subsections below.

The overdensities are calculated using equation 12.

$$\Delta = \frac{\rho}{\rho_{avg}} - 1 \tag{12}$$

Here ρ is the local density of galaxies at a certain position in space and ρ_{avg} is the average density within the entire survey volume. To calculate the overdensity the density of galaxies and the average density of that part of the universe needs to be known.

One complication related to the calculation of ρ_{avg} is that there are galaxies that cannot be observed in the galactic plane. The fraction of the volume missed because it is in the galactic plane can be determined using equation 13.

$$\gamma = \frac{\Omega_{gp}}{\Omega_{sky}} = \frac{\iint sin(b)\, db\, dl}{4\pi} \tag{13}$$

Then the average density is calculated using equation 14.

$$\rho_{avg} = \frac{N}{(1-\gamma)\, \iint_V R^2\, d\Omega\, dR} \tag{14}$$

5.1 Local Volume catalog

We use the Local Volume catalog[6] to determine the overdensities within 10 Mpc. This catalog is volume-complete for galaxies with $M_k <$ -16. The borders of the integration for equation 13 can be determined from a plot of the catalog in Galactic coordinates. The Galactic coordinates consist of the Galactic longitude, l, and the Galactic latitude, b. Ideally, the galactic plane should appear as a band with no galaxies around b = 0. Since the original reference for the Local Volume catalog does not provide Galactic coordinates (l, b) for the galaxies, these need to be calculated from its equatorial coordinates (α, δ). This can be done using equation 15, 16 and 17.

$$\sin(b) = \sin(\delta_G) \sin(\delta) + \cos(\delta_G) \cos(\delta) \cos(\alpha - \alpha_G) \tag{15}$$

$$\cos(b) \sin(122.9 - l) = \cos(\delta) \sin(\alpha - \alpha_G) \tag{16}$$

$$\cos(b) \cos(122.9 - l) = \cos(\delta_G) \sin(\delta) - \sin(\delta_G) \cos(\delta_G) \cos(\alpha - \alpha_G) \tag{17}$$

Here α_G and δ_G are the equatorial coordinates of the galactic pole. Figure 13 shows a plot of the position in the sky of Local Volume galaxies in Galactic coordinates. In figure 13 there is no obvious galactic plane present. As a result, we do not perform a galactic plane correction (this means that $\gamma = 0$ in equation 14).
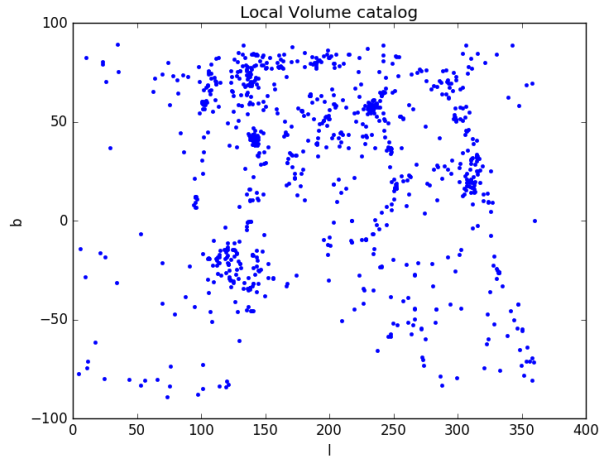

Figure 13: Galaxies of the Local Volume in galactic coordinates.

We then go ahead and calculate local overdensities within the same three dimensional cells as described in Sec 4.2. The densities for each cell in the Local Volume can be calculated relatively easily. A program loops over all the bins which are defined by the limits of table 2 and counts for each cell how many galaxies are inside of it. Then we divide by the volume of that bin the density is calculated. Because the catalog is volume-complete for $M_k < -16$ the average density is the number of galaxies in the catalog divided by the total volume (0 - 10 Mpc) of the catalog. The overdensities that are obtained are shown in figure 14.
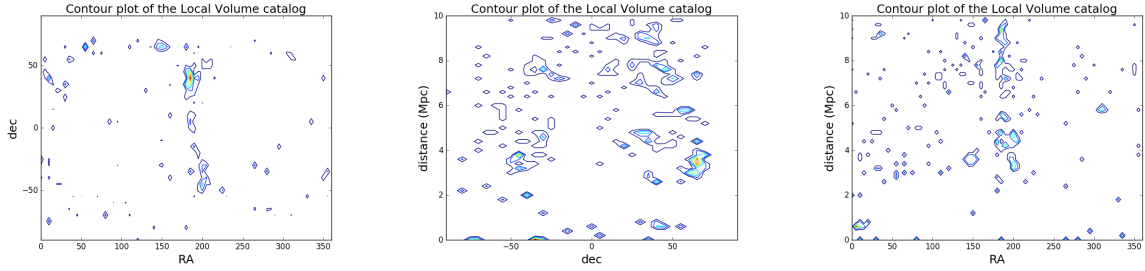
Figure 14: Contour plots showing the overdensities of the Local Volume catalog (in logarithmic scale). Left panel: the overdensities in the RA, Dec plane. Middle panel: the overdensities in the distance, Dec plane. Right panel: the overdensities in the distance, RA plane.

5.2   2MRS catalog

We use the 2MASS Redshift Survey (2MRS) catalog[7] to calculate overdensities between 10 and 300 Mpc. This catalog contains about 45000 galaxies out to a distance of 400 Mpc. The original reference for the 2MRS catalog does provide galactic coordinates for all galaxies, which means a conversion will not be necessary. The sky distribution of 2MRS galaxies are plotted in figure 15.
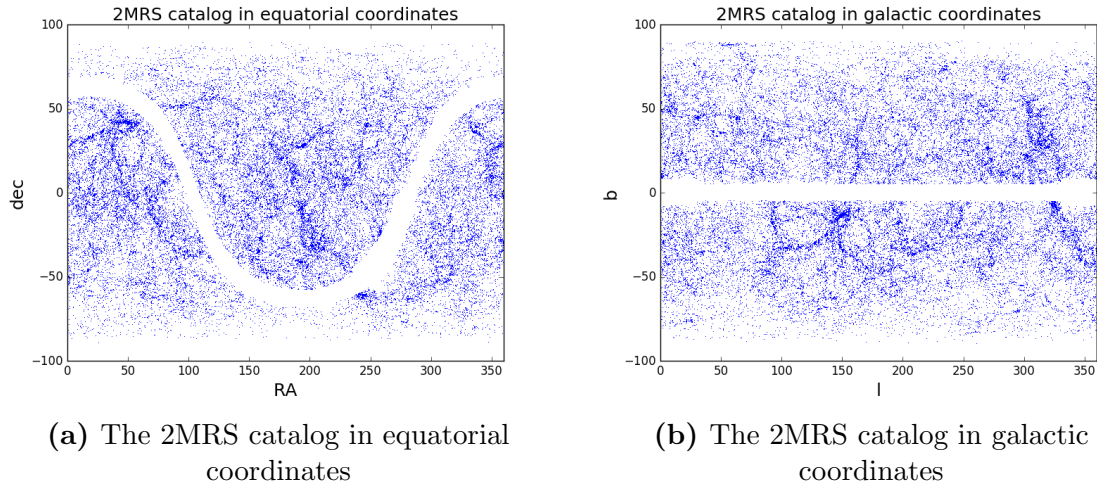
(a) The 2MRS catalog in equatorial coordinates

(b) The 2MRS catalog in galactic coordinates

Figure 15: Plots of the 2MRS catalog.

In these figures the Galactic plane is clearly visible, stretching from approximately b = -5 to b = 5 degrees. Using equation 13 the correction factor γ can be calculated. For the 2MRS catalog $\gamma = 0.087$.

The 2MRS catalog does not provide distances for the galaxies, but rather redshifts. This means that the distance first needs to be calculated using Hubbles law (eq 18) with a Hubble constant of $H_0 = 70$ km/s/Mpc.

$$D = \frac{v}{H_0} \tag{18}$$

The 2MRS catalog is a flux limited catalog, which means that the average density can not be calculated in the same way as for the volume limited catalog of galaxies in the Local Volume (section 5.1). However, volume-limited subsamples of the flux limited 2MRS survey can be extracted from the catalog and then be used in the same way as for the Local Volume to calculate the overdensities. Figure 16 illustrates the process we use to define volume-limited subsamples of the 2MRS sample. The volume limited samples can be extracted from the 2MRS catalog in the following way.

The catalog only provides the apparent magnitude of the galaxies. these need to be converted to absolute magnitudes to be useful. This can be done with equation 19.

$$M = -25 + m - 5\log_{10}(\frac{d}{Mpc}) \tag{19}$$

The catalog has a cutoff at a apparent magnitude of m = 11.75. Using this cutoff volume limited samples can be created. The volume limited samples are created by setting a distance limit, this way a magnitude limit can be obtained using equation 19. For the determining the overdensities 11 volume samples are created of which the limits are shown in table 3 and the samples are shown in figure 16.

Sample	Radial extent (Mpc)	Limiting abs. K-bad magnitude (mag)	Number of galaxies
1	10-50	-21.74	2741
2	50-75	-22.63	4776
3	75-100	-23.25	5455
4	100-125	-23.73	5390
5	125-150	-24.13	4764
6	150-175	-24.46	3507
7	175-200	-24.75	2166
8	200-225	-25.01	1639
9	225-250	-25.24	1169
10	250-275	-25.44	682
11	275-300	-25.63	433

Table 3: Volume-limited samples taken from the 2MRS catalog.

For each of these volume limited samples an average density is created and the densities of galaxies in each bin are calculated. It is important to note that these are not the real average density and densities of galaxies, but since they are both calculated in the same volume with
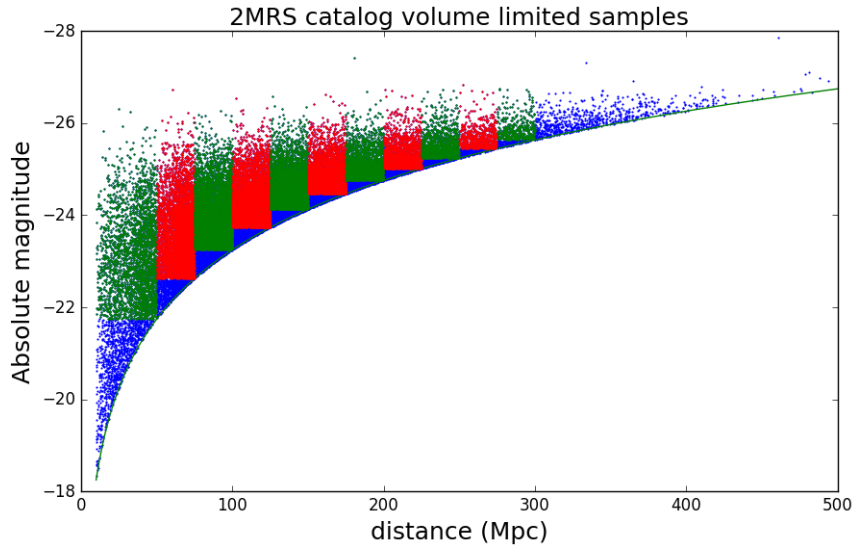
19

Figure 16: Volume limited samples from the 2MRS catalog. 2MRS catalog. The blue points show the entire 2MRS sample in a plot of absolute K-band magnitude vs. distance. The colored datapoints correspond to the 5 volume-limited subsamples (as seen in table 3) used here to calculate the large-scale structure overdensity.

the same amount of galaxies the overdensities are real. Then using the same method as used for the Local Volume catalog the overdensities are calculated. These can be seen in figure 17.
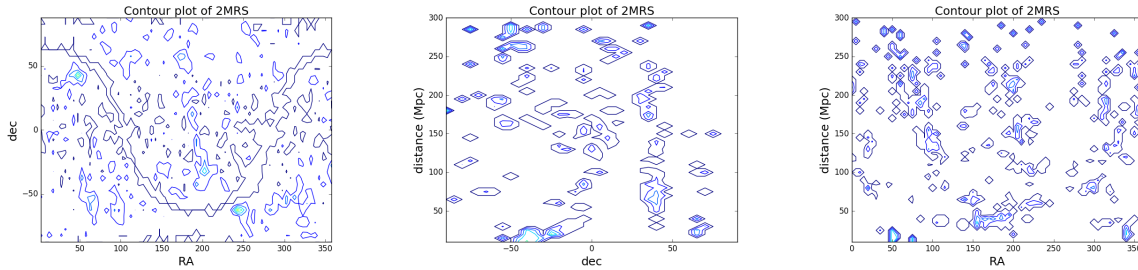

Figure 17: Contour plots showing the overdensities of the 2MRS catalog (in logarithmic scale). Left panel: the overdensities in the RA, Dec plane. Middle panel: the overdensities in the distance, Dec plane at RA = 50. Right panel: the overdensities in the distance, RA plane at dec = 50.

6 Giving realistic properties to the mock galaxies

In the previous sections we computed realistic estimates of the number and spatial distribution of APERTIF detections (which also includes the effect of large scale structure). In this section a sample of mock galaxies will be created and also be given realistic HI properties. The HI properties that will be determined for each of the galaxies are the HI mass, the HI profile width and the HI radius.

6.1 HI Mass and HI profile width

The determination of the mass, M_{HI}, and the profile width, W, go hand in hand with each other. This is because of two reasons: First, the distribution of galactic profile widths depends on the HI mass of a galaxy. Galaxies with larger M_{HI} values have larger average values of W (see figure 11). Second, the minimum detectable HI mass at a given distance depends on the profile width of a galaxy (see figure 10).

Realistic HI masses and widths for our mock galaxies are obtained in the following way. Using the sensitivity limit of each APERTIF survey the minimum detectable mass at each distance can be calculated. Using this minimum detectable mass a HI mass is drawn from the HI mass function of the ALFALFA survey. The HI mass function measures the number density of galaxies as a function of their HI mass, and therefore can be obtained by the MWF by marginalizing along the width direction (see figure 11). Then a HI profile width is calculated based on this HI mass, according to the distribution of W at fixed M_{HI} determined by the MWF of the ALFALFA survey. Using both the distance and the HI profile width the real minimum detectable mass can be obtained and then it can be checked whether the HI mass is detectable or not. This process is repeated until the number expected APERTIF detections of galaxies in a specific cell is reached (the cell that are used in section 4 and 5).

6.2 Radius

The HI radius of a galaxy can be estimated by using the HI mass-radius relation. Equation 20 is the best fit for finding this HI radius, given the HI mass of a galaxy (see above).[8]

$$\frac{M_{HI}}{M_\odot} = 12 \left(\frac{r_{HI}}{pc} \right)^2 \tag{20}$$

Which can be rewritten as,

$$\frac{r_{HI}}{pc} = \sqrt{\frac{1}{12} \frac{M_{HI}}{M_\odot}} \tag{21}$$

6.3 Applying the mask

To get a sample of mock galaxies for a specific survey the masks that have been calculated in section 2 have to be applied. The masks were saved as a true-false table for each RA-dec position on the sky. The way the program checks if it wants to keep a galaxy is that it uses

the galaxies RA-dec coordinates and checks in the table if there is a true or a false value. Is it a true value the galaxy will be kept and can be used for that specific survey. If it is a false value the galaxy will be discarded.

6.4 Spatially resolved HI data

One more extremely important aspect of the performance of the APERTIF surveys is the expected number of detections that will have resolved HI data. Estimates of the number of spatially resolved galaxies can be obtained using out catalog of mock APERTIF galaxies, and taking into account the resolving power of the APERTIF interferometer.

First we calculate the angular extent of each mock galaxy on the sky by using equation 22.

$$\theta_g = \frac{r}{10^3 \, D} \tag{22}$$

Where θ_g is the angle in radians, r is the radius of the mock galaxy in kpc and D is the distance to that galaxy in Mpc. This equation can be used because it is an approximation for small angles.

The angular resolution of APERTIF is determined by the separation and orientation of the baselines among the 12 Westerbork antennas. The Westerbork interferometer is constructed in a linear layout, and this results in different angular resolutions along the Dec and RA directions. As a result, the degree to which the HI emission from a galaxy is resolved depends not only on the angular extent of the galaxy on the sky, but also on the galaxy's position angle, PA. The position describes the orientation of the major axis of a galaxy with respect to the equatorial coordinates.

$$\theta_A = \sqrt{x^2 + y^2} \tag{23}$$

Where,

$$x = 15'' \cos(PA) \tag{24}$$

$$y = \frac{15''}{\sin(\delta)} \sin(PA) \tag{25}$$

This will give the angular resolution of APERTIF in arcseconds. PA is the position angle of the mock galaxy, which is generated randomly between 0 and 90, and δ is the Dec position of the mock galaxy. We can now calculate how many Westerbork resolution elements fit within the major diameter of each mock galaxy. This is done by using equation 26.[3]

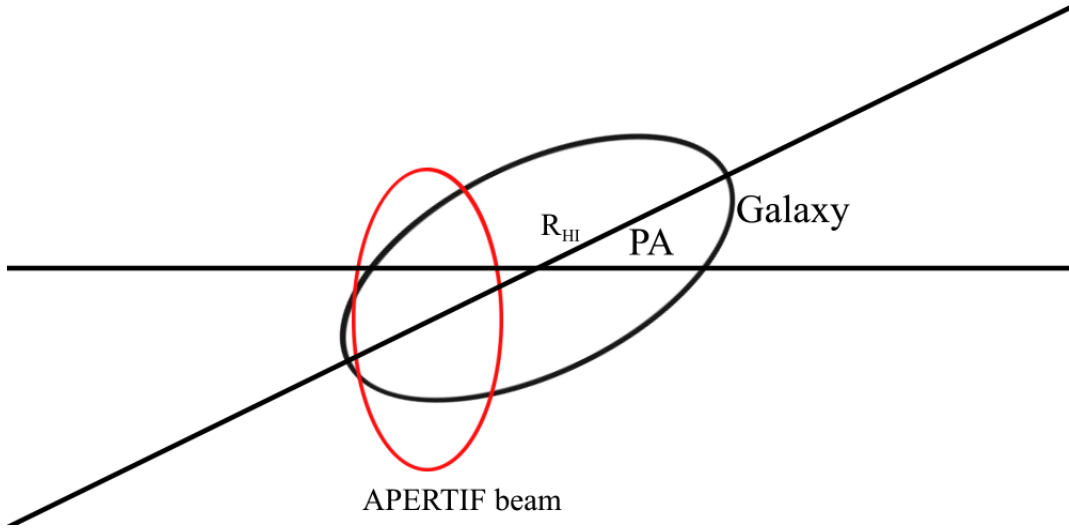$$R = \frac{206264.8 \, \theta_g}{\theta_A} \tag{26}$$

Figure 18: Illustration of how many APERTIF beams can be fit along the major axis of a mock galaxy.

7 Results

In this section we show several interesting results obtained from our sample of mock APER-TIF galaxies. First we show the properties of the sample of mock galaxies for each survey (in both the fall and spring sky). The figures are listed in the following order: a histogram of detections (figures 19 and 24, the log_{10} of mass a function of distance (figures 20 and 25), histogram of the log_{10} of the HI mass (figures 21 and 26), the log_{10} HI profile width as a function of distance (figures 22 and 27)and the distribution of mock galaxies in a RA, Dec plane (figures 23 and 28). Then in a table we summarize the properties for each of the planned surveys.
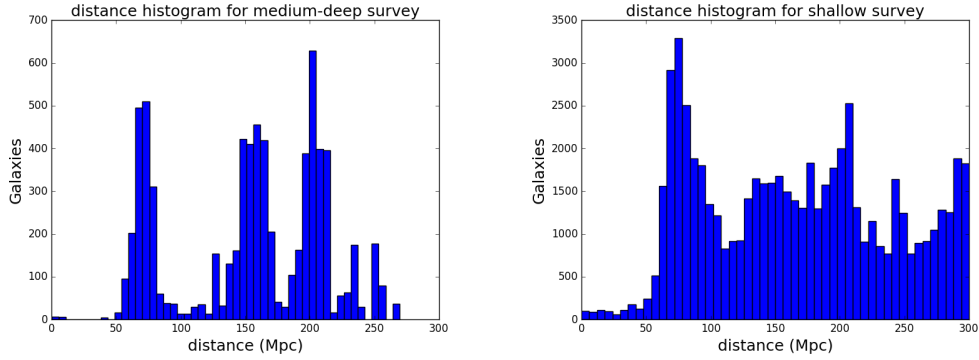

Figure 19: Left Figure: the histogram of the number detection for the medium-deep survey in the fall sky. Right Figure: the histogram of the number detection for the shallow survey in the fall sky.
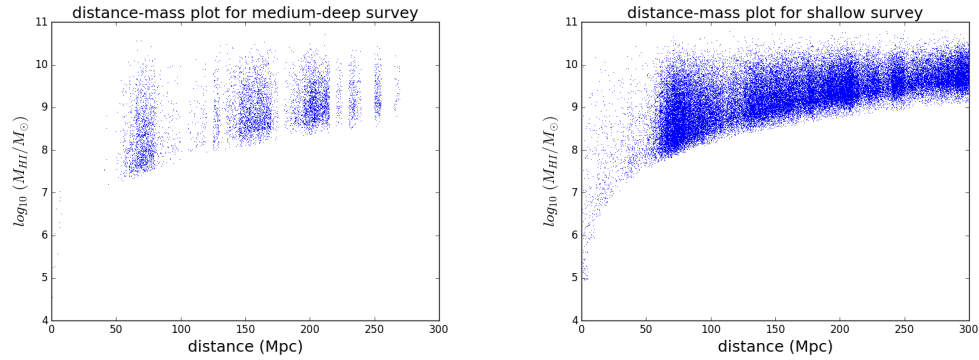

Figure 20: Left Figure: the log_{10} of the HI mass as a function distance for the medium-deep survey in the fall sky. Right Figure: the log_{10} of the HI mass as a function distance for the shallow survey in the fall sky.
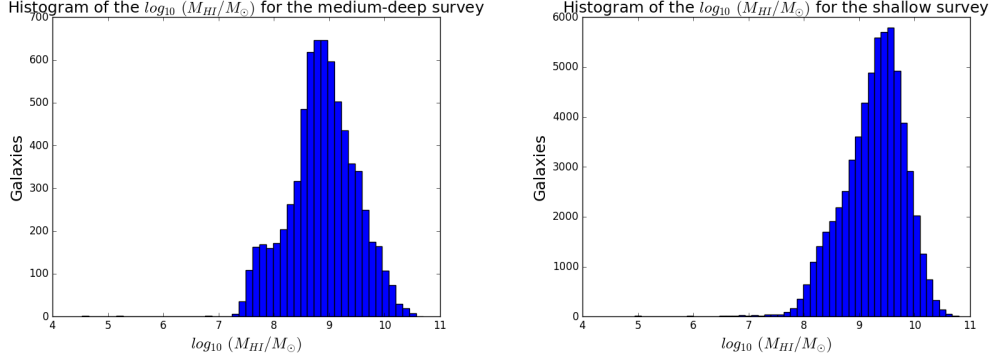
Figure 21: Left Figure: histogram of the log_{10} HI mass for the medium-deep survey in the fall sky. Right Figure: histogram of the log_{10} HI mass for the shallow survey in the fall sky.
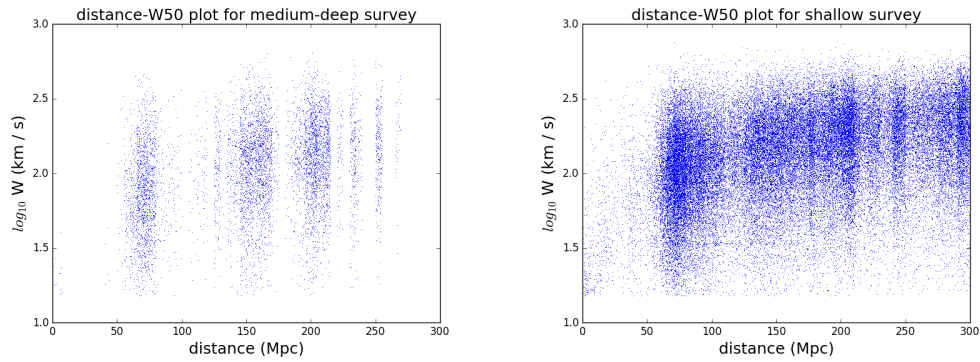

Figure 22: Left Figure: the log_{10} HI profile width as a function of distance for the medium-deep survey in the fall sky. Right Figure: the log_{10} HI profile width as a function of distance for the shallow survey in the fall sky.
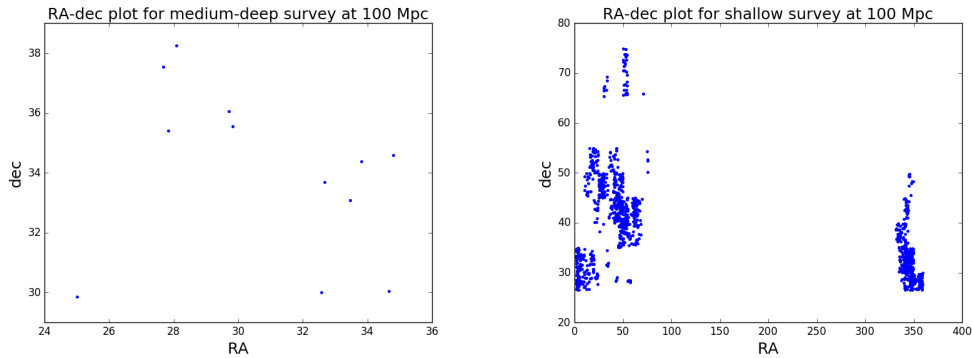

Figure 23: Left Figure: galaxies distribution in the RA, Dec plane at 100 Mpc for the medium-deep survey in the fall sky. Right Figure: galaxies distribution in the RA, Dec plane at 100 Mpc for the shallow survey in the fall sky.
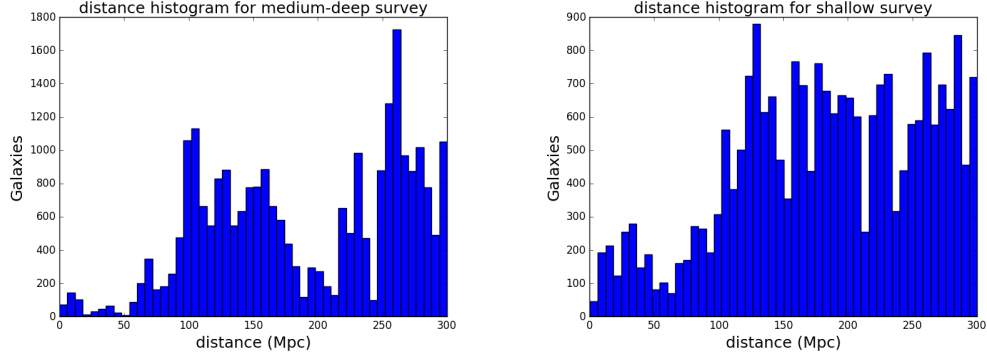
Figure 24: Left Figure: the histogram of detection for the medium-deep survey in the spring sky. Right Figure: the histogram of detection for the shallow survey in the spring sky.
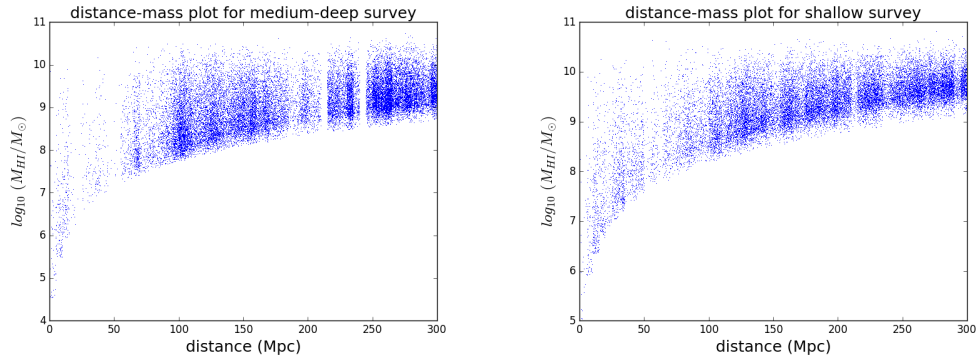

Figure 25: Left Figure: the log_{10} of the HI mass as a function distance for the medium-deep survey in the spring sky. Right Figure: the log_{10} of the HI mass as a function distance for the shallow survey in the spring sky.
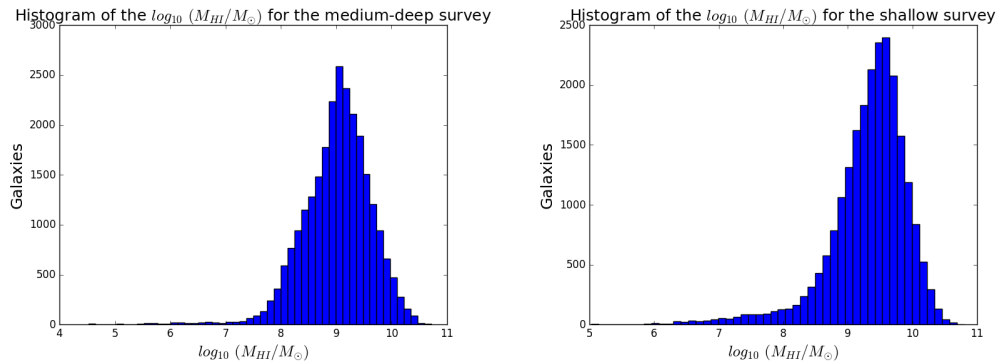

Figure 26: Left Figure: histogram of the log_{10} HI mass for the medium-deep survey in the spring sky. Right Figure: histogram of the log_{10} HI mass for the shallow survey in the spring sky.
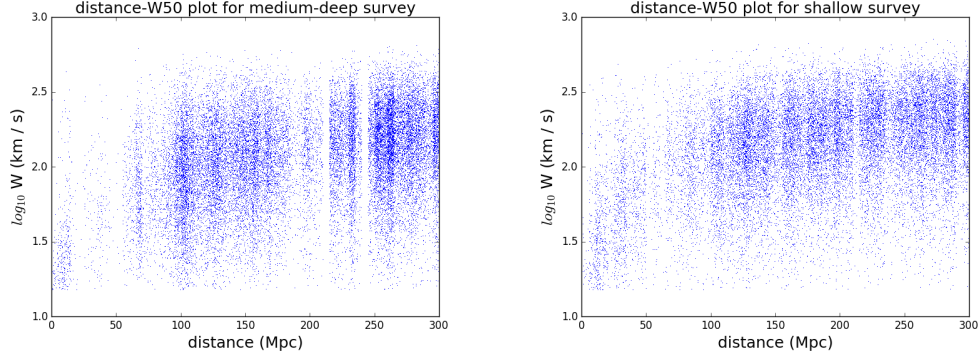
Figure 27: Left Figure: the log_{10} HI profile width as a function of distance for the medium-deep survey in the spring sky. Right Figure: the log_{10} HI profile width as a function of distance for the shallow survey in the spring sky.
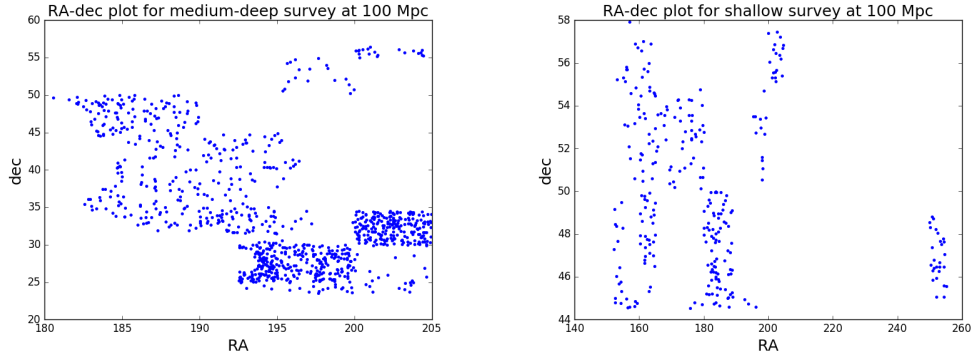

Figure 28: Left Figure: galaxies distribution in the RA, Dec plane at 100 Mpc for the medium-deep survey in the spring sky. Right Figure: galaxies distribution in the RA, Dec plane at 100 Mpc for the shallow survey in the spring sky.

Survey	Number of detections within 300 Mpc[1]	Number of dwarf galaxies[2]	Number of resolved galaxies[3]	Number of resolved dwarfs
Shallow (spring sky)	49682	6620	207	10
Shallow (fall sky)	127685	24707	459	11
Medium-deep (spring sky)	25761	4571	86	10
Medium-deep (fall sky)	7063	1707	24	2

Table 4: Detection statistics of the shallow and medium-deep APERTIF HI surveys. [1] Refers to the number of detectable galaxies with $M_{HI} >= 10^7 \, M_{sun}$ and W $>= 23$ km/s. [2] Dwarfs are here defined as galaxies width $M_{HI} = 10^7 - 10^{8.5} \, M_{sun}$. [3] Resolved here refers to galaxies with $>= 4$ resolution elements along their major diameter.

8 Conclusion

In this thesis we create a realistic sample of mock galaxies of APERTIF HI detections. First, We create a mask for each of the APERTIF surveys which allows us to create mock galaxies over the planned sky footprint of each survey. We store this mask as a table, which contains the information (true or false) of the sky coverage for every RA, Dec point. Second, we determine the number of expected APERTIF detections at each point of the survey volume, assuming a spatially homogeneous galaxy distribution. We use the technical specification from each of the APERTIF surveys the sensitivity of FPA can be determined an used to calculate the minimum detectable HI mass at each distance. We then obtain a count of galaxies for each cell by integrating over the mass-width function (MWF) of the ALFALFA survey. Then the large-scale structure is added by using 2 catalogs, the volume-limited Local Volume catalog (0 - 10 Mpc) and the flux-limited 2MRS catalog (10 - 300 Mpc). The overdensities are computed from each of these catalogs and stored into an array of overdensity values for each cell in a 3D grid covering the survey volume. Overdensities from the 2MRS can not be extracted in the same as for the Local Volume catalog. The Local Volume catalog is volume-limited, which means the overdensities are simple the densities in a specific cell divided by the average density. The 2MRS catalog is flux-limited, which means volume-limited samples need to be extracted from this catalog to extract the overdensities. To do this we used 11 samples ranging from 10 to 300 Mpc to extract to overdensities. Finally we assign realist HI properties. We use the MWF measured by ALFALFA to assign HI masses and HI profile widths to our mock galaxies. We furthermore assign HI radii by using a scaling relation between the HI mass of a galaxy and the radius. Estimates of the galactic HI radii allow us to also compute if APERTIF can resolve the galaxy on the sky. From the results in section 7 we can see that the most HI detection will be made with the shallow survey, which can easily be explained due to its large sky coverage. However at larger distances the medium-deep survey plays a bigger role because of its higher sensitivity. Also the most detections, which can be spatially resolved will take place in the shallow survey. The most dwarf galaxies can also be found in the shallow survey. The software used to create the mock catalog can also be used for different radio telescopes in helping finalizing there designs. This can be done by changing the variables in the radiometer equation to get a different sensitivity and apply a different sky mask to get the required mock catalog.

9 References

[1] Wim van Cappellen. Tom Oosterloo, Marc Verheijen. Apertif - the focal-plane array system for the WSRT. October 2010.

[2] van der Hulst (private communication). The apertif survey layout. 2016.

[3] George Heald Wim van Cappellen Marc Verheijen, Tom Oosterloo. Hi surveys with apertif. June 2009.

[4] M.P. Haynes E. Papastergis, R. Giovanelli and F. Shankar. Is there a "too big to fail" problem in the field? May 2015.

[5] et al Giovanelli, R. The arecibo legacy fast alfa survey.

[6] Dmitry I. Makarov Igor D. Karachentsev and Elena I. Kaisina. Updated Nearby Galaxy Catalog. March 2013.

[7] et al John P. Huchra, Lucas M. Macri. The 2MASS Redshift Survey - Description and data release. December 2011.

[8] G. De Lucia S. Khochfar D. Obreschkow, D. Croton and S. Rawlings. Simulation of the Cosmic Evolution of Atomic and Molecular Hydrogen in Galaxies. June 2009.

[9] Riccardo Giovanelli Martha P. Haynes Christopher M. Springob Ann M. Martin, Emmanouil Papastergis and Sabrina Stierwalt. The Arecibo Legacy Fast ALFA Survey: X. The HI Mass Function and OmegaHI from the 40Survey. October 2010.

10 Appendix

10.1 Appendix A: Survey selection function in a simple idealized case - the HI mass function of galaxies

In this section we compute the expected number of APERTIF detections in an idealized case, where the detection limit of the survey does not depend on the profile width, W, of the galaxy (see section 4)

Since this setup is not realistic for the APERTIF surveys, we denote the minimum detectable HI flux by $S_{HI,min}$, and we assign to it the arbitrary value of $S_{Hi,min} = 0.1$ Jy km/s. The minimum mass detectable at distance D can be calculated with equation 27.

$$M_{HI,min}(M_{sun}) = 2.356 * 10^5 \, D(Mpc)^2 \, S_{HI,min}(Jykm/s) \qquad (27)$$

In order to calculate the number of expected APERTIF detections as a function of distance in this simplified case, we need to know the number density of galaxies in the universe, as a function of their HI mass. This distribution is referred to as the HI mass function (HIMF) of galaxies.

The HI mass function is parameterized as a Schechter function in the form of equation 28. This function consists of 2 parts, a part described by a power law and a part described by an exponential law. It switches between the two at its characteristic mass, M^*. This is illustrated in figure 29.

$$\Phi(M_{min}) = \phi_* \frac{M_{min}}{M_*}^{\alpha} e^{-\frac{M_{min}}{M^*}} \qquad (28)$$

In equation 28 there are 3 parameters, namely the faint-end slope α, the characteristic mass M_* and the normalization parameter ϕ_*. The Schechter function in equation 28 parameterizes the number of galaxies per unit volume per unit logarithmic mass, and has units of $Mpc^{-3} \, dex^{-1}$. Here, we use the values of the parameters for the HI mass function shown in table 5, as determined by the ALFALFA survey[9]

M_*	$10^{9.96}$
ϕ_*	0.0048
α	-1.33

Table 5: Parameters of the Schechter function as determined by the ALFALFA HI survey.

Given the minimum detectable mass at distance D calculated from equation 27 the number of galaxies per Mpc^3 with HI mass greater than the minimum mass. This is done by integrating the Schechter function:

$$n_{det}(D) = \int_{M_{min}(D)}^{M_{max}(D)} \Phi(M_{HI}) dM \qquad (29)$$

This integral can be expressed analytically in terms of the inverse gamma function.

30

$$n_{det}(D) = \Gamma(\alpha, M_{HI,min}) \qquad (30)$$

We show the result in the right of figure 29. This figure show the selection function of the survey. We set the number density of galaxies detected by the survey at the lowest distance (here $D_{min} = 1$ Mpc) included in the computation as the reference number density. We then calculate the number density of detections at different distances, and express the result as a percentage of the reference number density. Using this a fraction of the galaxies that is detected at each distance is obtained from the HI mass function, which is referred to as the selection function of the survey.
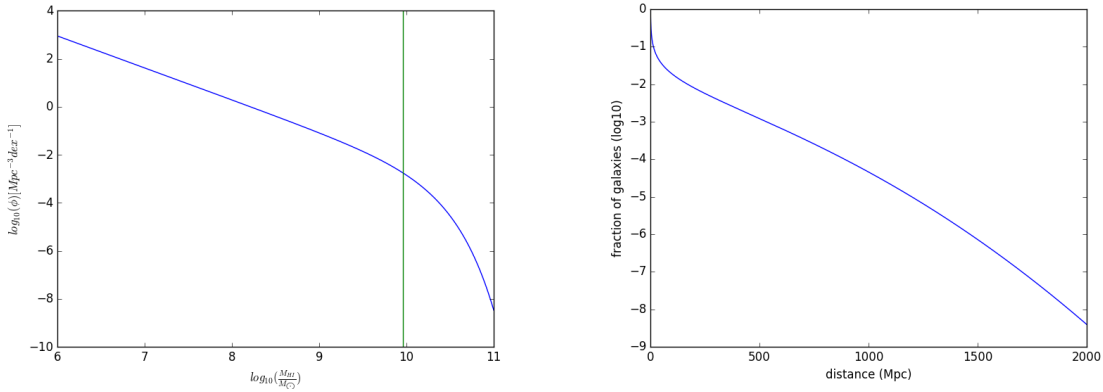

Figure 29: Left Figure: the HI mass function plotted as a function of mass. The green line shows the charateristic HI mass M^*, the value of HI mass where the transition between a power law and an exponential behavior takes place. Right Figure: the survey selection function, i.e., the fractions of galaxies that are detected by the survey as s function of distance, with respect to the galaxies detected at the smallest distance considered here, $D_{min} = 1$ Mpc.

Figure 30 shows the effect of the survey selection function in shaping the distribution of number of survey detections as a function of distance. Assuming a spatially homogeneous galactic population, the amount of galaxies in a volume-limited sample scale with the square of the distance, (since the volume of a radial shell increases as $dV = 4\pi R^2 dR$). The number of galaxies at eacht distance that are bright enough to be detected by the survey is obtained by applying the selection function calculated equation 29 to the volume-limited number of galaxies in the survey volume.
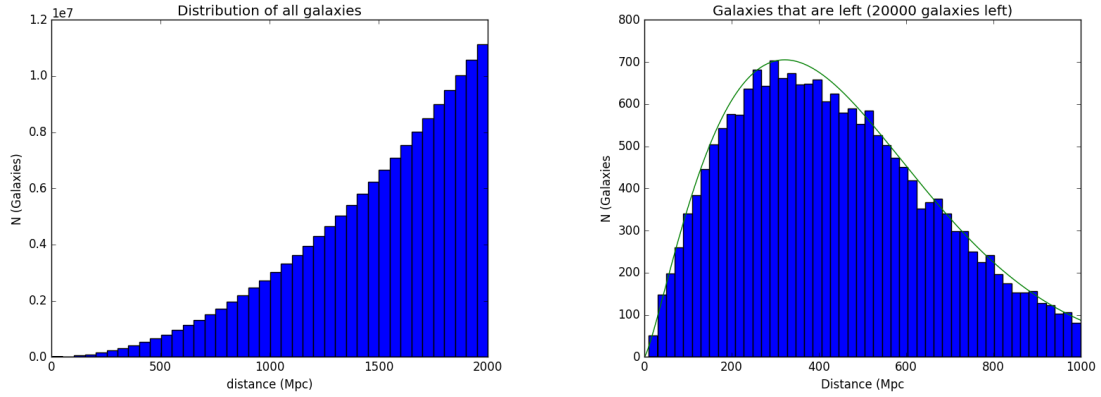
Figure 30: Left Figure: the distribution of number of galaxies as a function of distance for a volume-limited galactic sample. The distribution follows a square law, reflecting the growing volume of more distant radial shells. Right Figure: the distance distribution of galaxies that can be detected by the survey, according to the survey selection function (equation 29

In practice, the values in the right figure of figure 30 were obtained in the following way. First galaxies are generated reflecting the distance distribution of the volume-limited sample. This is achieved by assigning distance such that their cube power is uniformly distributed between 1 and in this case a 1000 Mpc. Since the fraction is known at each distances of galaxies that should be left the program uses this fraction as odds if the galaxies generated at a certain distance should be kept or should be discarded. The program does that in the case of figure 30 until there are 20000 galaxies left.

10.2 Appendix B: Extrapolating the mock sample to very low HI masses

A problem that arises with determining the number of galaxies in each bin, is that at nearby distances the minimum mass is actually lower than the first HI mass bin of the table. This is caused by the fact the APERTIF surveys are more sensitive than the previous generation survey,ALFALFA. This means that number densities below a HI mass bin of 7.1 still need to be generated. The smallest galaxies that have been seen have a mass of around 10^5 solar masses, so the missing HI mass bins are between 5.1 and 7.1.

The way in which detections with HI mass in the range of M_{HI} $10^5 - 10^7 M_{sol}$ can be generated is by extrapolating the ALFALFA measurements of the MWF below its low-mass limit. We do this by taking slices of the ALFALFA MWF along constant values of mass, and then fitting an analytical form to the distribution of velocities in each mass slice. The missing HI mass bins for the MWF can be generated is by fitting a probability distribution to the know data sets and than extrapolation this probability distribution for the missing HI mass bin. The probability distribution used for fitting the velocity distributions is a left-skewed Gumbel distribution (equation 31).

$$pdf(x) = \frac{1}{\beta} e^{\frac{x-\mu}{\beta} - e^{\frac{x-\mu}{\beta}}} \tag{31}$$

The Gumbel distribution has three parameters, μ the location parameter, β the scale parameter and δ, an overall normalization parameter. To generate the distribution for the lower mass bins these values need to be obtained.

The location parameter μ is the position of the maximum value. To find the location parameter for the missing data sets a linear function can be fitted to the location parameters of the MWF table. Obtaining the maximum value for each of the missing data sets is important because this value is used to normalize the distribution. The maximum for the missing data sets can also be found by fitting a linear function. The last piece of the puzzle is to determine the scale parameter β, which determines the shape of the distribution. Finding the scale parameter for each distribution is a little bit more trickier. It can be done by summing up all the number density values in each mass bin and then extrapolate this linearly downwards for lower mass bins. Now the sum of the density values is known for the missing mass bins. This can be used to find the scale parameter in the following way. A program creates a list of 100 scale parameters ranging from 0.05 to 2. For every of these scale parameters it create a Gumbel distribution using the location parameter and normalizes the distribution using the maximum values. Then it sums up the densities of each of the created distribution and the program checks which sum is the closest to the expected value of the sum of densities. This is the scale parameter it will use in creating the distribution for the missing mass bin. The process is repeated for every missing mass bin.
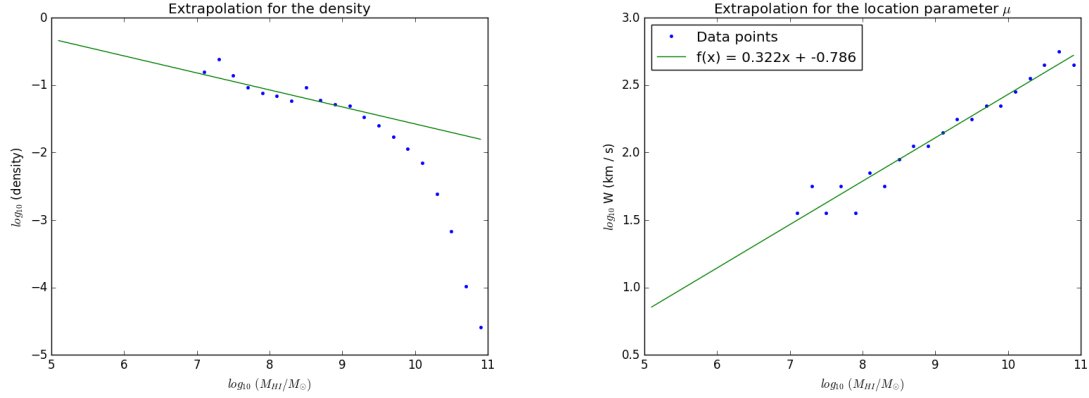
Figure 31: Left Figure: fit for the sum of the densities. Right Figure: fit for the location parameter

The number densities for mass bins 5.1 to 7.1 can now be generated. Since the MWF table has a dM = 0.2 ten additional bins need to be generated. The densities seem to be a bit higher than expected, this can be explained by the fact that in each mass bin 17 number densities are generated (dW = 0.1). But the data sets for low mass bins(7.1 to 8.1) only contain about 3 to 5 data points, the rest is zero. Thus a lot of extra number densities are added in the generated mass bins which could explain this.
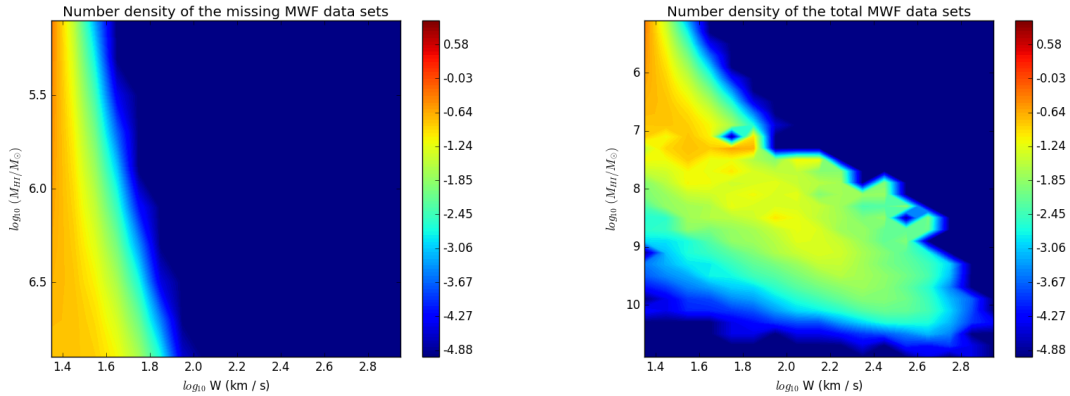
(a) The number densities of the MWF of the missing data sets

(b) The number densities of the MWF of all data sets

Figure 32: The MWF data sets.

10.3   Appendix C: Code used in this thesis

All the code used to create to the sample of mock galaxies and the results can be found at the following location:

www.astro.rug.nl/˜roelenga