

Hector Target Selection notes

Hector Target Selection Summary

Sam Vaughan, March 2024

Here's a short-ish summary of the target selection process for the Hector Galaxy Survey. Hopefully this will help when it comes to write the paper!

Note that much of the cataloguing work for the Clusters was done by Matt Owers (matt.owers@mq.edu.au), so you'll have to speak to him for the details there.

Any questions, you can contact me on samvaughan01@gmail.com.

Initial Catalogues

The initial catalogues for the WAVES North and WAVES South regions come from the WAVES team. Sabine Bellstedt (sabine.bellstedt@uwa.edu.au) has sent me catalogues which are the output from the Profound source extraction software. These were ~10s of Gb in size, containing every source within these regions of sky.

I manually corrected for galactic extinction using the maps of [Schlafly and Finkbeiner \(2011\)](#) and made a magnitude cut of $r < 19.5$ to remove faint objects. I also removed objects which had a class of "Star". These form the WAVES catalogues in the folder `HectorInputCatalogues/resources/InitialInputCatalogues/`.

The Cluster catalogue in there is saved in `resources/InitialInputCatalogues/` and named `COMBINED_ApMatch_MSTARV2.5_MattOwers_310123.parquet`. Matt will provide the details, and hopefully the date will help!

Our input redshift catalogue and the Hector Redshift Survey

WAVES

Our input redshift catalogue comes from a few places. For WAVES, it's mostly the 4HS survey input redshift catalogue which was collated by Ned Taylor (entaylor@swin.edu.au). He has combined together essentially every redshift survey in the literature into a catalogue called `hemispec.v0.91recommended`.

However, this catalogue is badly incomplete in certain areas of the WAVES South region. This is because there's never been an equivalent of the SDSS survey in the Southern hemisphere. The main workhorse instrument for measuring redshifts in the South has been 2dF, but 2dF has tended to concentrate on performing deep observations on relatively small patches of sky (e.g. the GAMA survey).

When it has performed wide area redshift surveys, such as 2dFGRS (ask Matthew and Scott!), the sky coverage has been patchy and the selection of galaxies hasn't been ideal- the selection of galaxies for 2dFGRS was a bJ-rF colour cut performed on data from photographic plates, which ended up preferentially selecting bluer objects at a given redshift. This has meant that the catalogue of redshifts from 2dFGRS is incomplete at all magnitudes- it doesn't just miss faint things, it misses some bright red things too.

To get around this issue, Scott, Julia and I devised the Hector Redshift Survey (HRS). We used the WAVES input catalogues and removed things which already had a redshift. We then tried to target all objects with an r -band magnitude of less than 17.7 (to match the depth of SDSS in the Northern Sky). Note that we *didn't* perform any colour cuts or photometric redshift selection here, which would probably have been useful.

We then observed galaxies in the H01 and H03 regions of sky. Our shallow limit of $r < 17.7$ meant that we had around ~ 200 targets per 2dF configuration, and our exposure times were only 20 minutes. This allowed us to get many observations performed per night. Matt Owers reduced the data and made the redshift measurements using MARZ (I believe).

As of March 2024, we've observed 7038 galaxies across H01 and H03. I performed some calculations to show that only $\sim 5\%$ of the galaxies we observed with the HRS would actually end up being part of the Hector Galaxy Survey proper, but that's how it goes sometimes! These observations were included in the main target selection pipeline in the folder `resources/HRS_Redshift_Observations`.

Clusters

Matt compiled the Cluster redshift catalogue from various literature studies of our chosen fields. We also measured many redshifts ourselves using 2dF, in a similar manner to the

Hector Redshift Survey for the WAVES fields (but with different magnitude limits- talk to Matt for details).

Catalogue preparation

WAVES

The script `prepapre_WAVES_Catalogues.py` then takes the initial input catalogue and makes some important changes. Firstly, it infers a stellar mass for each galaxy based in their $g-i$ colour (see Julia’s [2015 paper](#) for details) and assuming a flat Λ CDM cosmology (i.e. `from astropy.cosmology import FlatLambdaCDM`).

We then remove galaxies which are too close to bright stars. For WAVES North we use stars from the [PANSTARRS survey DR1](#) and for WAVES South we use stars from the [Skymapper survey DR2](#). Precise SQL queries can be found in `workflow/scripts/utils.py`. I’ve used slightly older data releases (e.g. PANSTARRS DR2 is out, and Skymapper DR4 has recently been released) because I download the data from the ESA GAIA portal (the easiest way I could find to actually access the data). GAIA have already performed cross-matches with PANSTARRS DR1 and Skymapper DR2, but not (at time of writing) to later data releases.

Assuming a galaxy bundle size of 20 arcseconds (our largest bundle), we mark as ‘bad’ a galaxy which:

- Has any star within 5 arcseconds of the galaxy coordinate
- has a 16th magnitude star within the bundle radius
- has a 15th magnitude star which is within 2 bundle radii
- has a 12th magnitude star or brighter which is within 5 bundle radii

Finally, for WAVES North, any SAMI galaxies in the field are removed from the catalogue. This leaves us with the files in `results/FinalInputCatalogues`. These catalogues haven’t had the target selection steps applied yet- that is performed later.

Clusters

The script `workflow/scripts/prepare_cluster_catalogue_including_foreground.py` takes the initial Cluster input catalogue and performs many of the same steps as for the WAVES catalogues (see above). There are two important differences, however.

Firstly, we remove SAMI galaxies (among the clusters which have been observed before by SAMI) and then only keep either galaxies which are either:

- members of their parent cluster (using Matt’s cluster membership flag, `mem_flag`) and are within 2.5 times the cluster r_{200} value.
- are in the foreground of the cluster and have a stellar mass below $10^{9.5}$ solar masses.

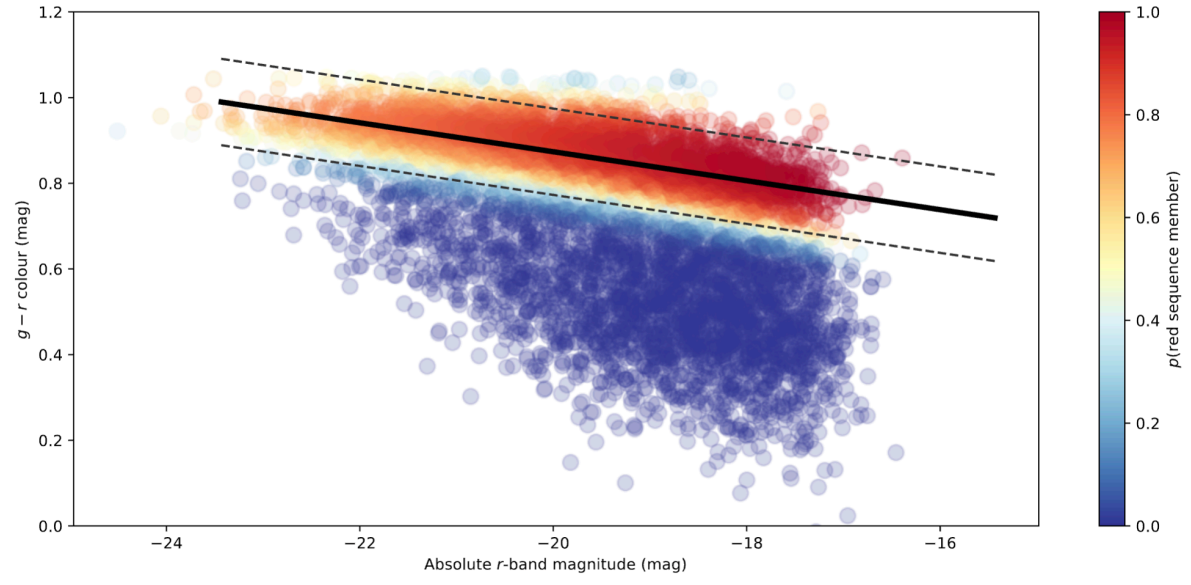
The next important step is to find the probability that each cluster member is part of the cluster Red Sequence. To do this, we build a Gaussian mixture model based on the distribution of galaxies in the absolute R band magnitude vs $g - r$ colour space.

Firstly, we calculate a K-corrected absolute R -band magnitude for each galaxy assuming a Flat Λ CDM cosmology. We then fit a Gaussian mixture model using the program [Stan](#), which performs full Bayesian inference of the posterior using Hamiltonian Monte Carlo sampling. Note that this mixture model is fit to all cluster galaxies at the same time- we don't do this on a cluster-by-cluster basis.

The model assumes that the cluster red sequence lies along a straight line in absolute R vs $g - r$ colour space. The slope and intercept of this straight line are left free. We then state that cluster red-sequence members are drawn from Student-t distribution centred on this line, with a width (σ) which is also allowed to vary.

We also model the remaining galaxies to be drawn from a Student-t distribution which follows a linear relationship between R and $g - r$ colour. The width of this distribution is also free to vary. The ν parameter of the Student-t distribution is also free to vary, but is the same for both mixture components.

An example of the output is shown below. The colour bar gives the probability that a given galaxy is part of the red sequence. The thick black line is the relationship between the galaxy's $g - i$ colour and its absolute R -band magnitude: the red sequence itself. The dashed lines show the 1σ width of the red sequence.



Going forward, we say that a red sequence member is any galaxy which is within 2σ of the scatter around the main sequence.

The Selection Function

The next step for each of the WAVES catalogues and the Cluster catalogue (including foreground objects) is to run the Hector target selection code. This code does a few things behind the scenes (including renaming various columns), but its main role is to only keep galaxies which meet our target selection criteria.

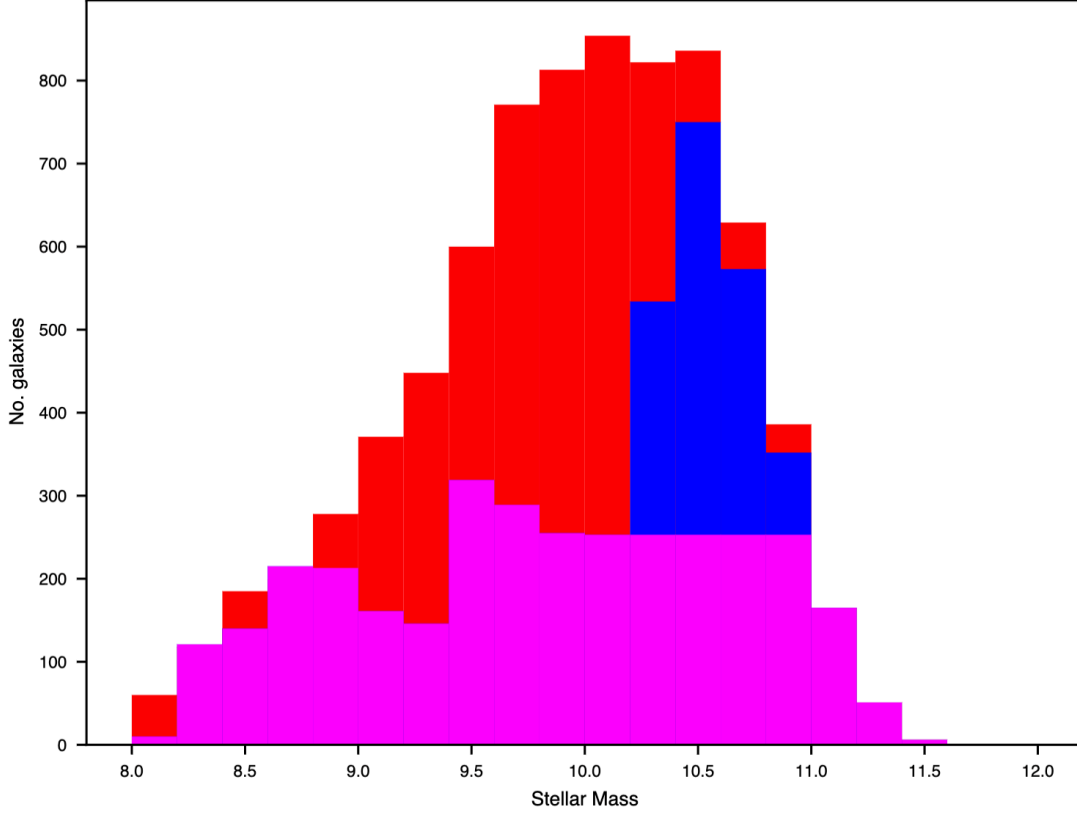
The parameters which get fed into the target selection code are saved in the files `resources/TS_config_files/*`. The main aspects of the code are to:

- Only keep galaxies with $z < 0.1$.
- Keep galaxies with a mass greater than 10^8 solar masses and less than 10^{12} solar masses.
- Have an effective radius of greater than 1.5 arcseconds
- Lie above the Hector Steps in the mass/redshift plane. The Hector Steps are very similar to the SAMI steps, with a small difference at the high mass end. The actual definition is in the file `HectorObservationPipeline/SelectionFunctions/custom_selection_function.yaml`.
- Sparsely select galaxies to either be “flat in mass” or sparsely select galaxies from the cluster red sequence.

Sparse Selection

The sparse selection merits a bit of explanation. If we observe all galaxies which lie above the Hector Steps in the mass/redshift plane, we’d end up with thousands of galaxies at a stellar mass around $10^{10.5}$ solar masses (roughly L_* in the stellar luminosity function). These galaxies would be predominantly quiescent fast rotators, which aren’t too interesting... We therefore select galaxies to be “Flat in Mass” above a stellar mass of 10^{10} , which means that we chop off the top of the stellar mass function.

This is best explained with an example. The plot below shows the mass functions for different selections. The red histogram shows all galaxies in the master catalogue, before we apply the Hector Steps or any target selection criteria. The blue histogram (which is identical to the pink one below 10^{10}) shows all the galaxies which are above the Hector steps in the mass/redshift plane. The pink histogram shows the final selection: our “Flat in Mass” selection.



We have drawn a horizontal line across at the height of the 10^{10} bin, then taken a random sample of the galaxies in the next bin up to that line. For example, if there are 1000 galaxies in the 10^{10} bin and 1500 galaxies in a higher mass bin, we randomly sample 1000 of them to match the number in the 10^{10} bin. If there are fewer galaxies in a bin than the 10^{10} bin, we keep all of them.

For the cluster targets, we instead use the red sequence membership from before (i.e. any galaxy within 2σ of the red sequence) and randomly sample 70% of these galaxies. This percentage was chosen to give us the right number of cluster targets overall (\sim roughly 3000).

The catalogues we save

Finally, we then save the sparsely sampled catalogues in a folder in `results/MasterCatalogues/`. These catalogues are the ones we use for the tiling, observing, etc. We also save the master catalogue *with* the stepped selection function but *without* the sparse sampling, and *without* removing the “bad class” galaxies mentioned previously (i.e. those which are too close to a

star, etc). This master catalogue without the sparse sampling is useful to act as the overall source-of-truth for the survey.

The actual catalogues with the sparse sampling (which we observe) will change every time the catalogue code is run, i.e. every time we get more redshifts from the HRS. This means that there have been situations in the past where we had observed galaxies which technically weren't in our master catalogue. After a discussion with Scott, Julia and Stefania in March 2024, we decided the best thing to do was save the un-sampled catalogue somewhere, and use this as the “source-of-truth” which we'd match against in the Hector database.

Splitting up the Catalogues by Region

We then take our master catalogue (which has passed the target selection criteria) and separate them by their individual regions (e.g. H03, A3376, etc). This is done by their RA and Declination, using the file `resources/RegionInformation/all_regions.csv`. The catalogues themselves are saved in `results/RegionCatalogues/{region_name}`, with the filename `{region_name}_Hector_target_galaxies.csv`.

Selecting Stars

The last important step is to make catalogues of guide stars and standard stars for the WAVES regions of the sky. For the clusters, Matt has us sorted and has selected the star catalogues already.

Similarly to before, we use the Panstarrs catalogue for WAVES North and Skymapper for WAVES South. The script `workflow/scripts/select_stars.py` uses the `astroquery` package to get data from the ESA Gaia archive.

We remove stars which are too close to one another, correct for proper motion (from GAIA) and assign different priorities to each one (which are the same as the SAMI star priorities- see [Bryant et al. 2015](#)).

We select guide stars which have a g -band magnitude between 14 and 15. This is the ideal brightness for the Hector guider.

We then try and select guide stars which are most like F-type stars. To help with this, I took some observed F-type star spectra from the X-Shooter Spectral Library, and “observed” them through various filters from the Skymapper and PANSTARRS data. I measured their colours, and have stored the averages in the files `resources/Misc/PANSTARRS_F_star_colours.csv` and `resources/Misc/SkyMapper_F_star_colours.csv`. I then compare the star colours from the catalogue to these “template” F-star colours, and assign each star an “X” value (in the same way Julia did in her 2015 paper). I then sort by this value and keep stars which have smaller values of X- these are most likely to be F stars.

The star catalogues are then saved in `results/RegionCatalogues/{region_name}` and are called `{region_name}_standard_stars.csv` and `{region_name}_guide_stars.csv` respectively.