



Utrecht University

Intrinsic size correlations of elliptical galaxies in the Horizon-AGN simulation

Sjoerd Hendrik-Jan Weide

5975905

Institute for Theoretical Physics
dr. N.E. Chisari & dr. H.S. Johnston

Utrecht University
20 January 2021

Contents

1	Introduction	1
2	Data	3
2.1	Horizon-AGN Simulation	3
2.2	Data Cuts	4
3	Method	9
3.1	The Fundamental Plane	9
3.2	The Pipeline	10
3.2.1	The Fundamental Plane	10
3.2.2	Fitting the Fundamental Plane	10
3.2.3	Correlation Estimators	12
3.2.4	Comparison Between Methods	15
3.2.5	Error Calculation	17
4	Results	21
4.1	Correlation Estimators	21
4.2	Concerns	22
5	Conclusion	25
	Bibliography	28
	Appendix A Acknowledgements	29

Chapter 1

Introduction

In modern cosmology weak lensing has been used to constrain cosmological parameters. Weak lensing is a phenomenon where small distortions in how galaxies look are caused by the gravitational field in between the source of the light and the receiver of that light. These distortions are measured as shape (shear) and size (magnification) correlations between galaxies that lie in the same region, because their light travels through the same field. These correlations are then used to tune cosmological parameters. Similar shape correlations can, however, be caused by the intrinsic alignment of galaxies (Ciarlariello and Crittenden 2016). In much the same way, the *size* correlations of these galaxies can also be different from the weak lensing signal by having intrinsic size correlations between galaxies (Joachimi, Singh, and Mandelbaum 2015). These correlations would, in that case, find their origin in other factors than gravitational lensing, and thus contaminate the tuning of parameters.

A good way of seeing whether an elliptical galaxy is magnified is the Fundamental Plane (FP). This is a relation between two measurable factors and the size of a galaxy, which can be used as a guideline for actual size and compared to observations. Even so, if a galaxy deviates from the FP because of other factors than the lensing, lensing measurements might be impacted. As has been seen in previous research intrinsic size correlations do cause galaxies to significantly deviate from the FP (Joachimi, Singh, and Mandelbaum 2015).

In this thesis we aim to see whether this effect that has been measured in observations from the Sloan Digital Sky Survey (Joachimi, Singh, and Mandelbaum 2015) can also be measured in simulations. We use data obtained from the Horizon-AGN simulations to recreate results from Joachimi, Singh, and Mandelbaum 2015 and compare these with the observations. We find evidence for a positive correlation between the position of galaxies and the deviation of their size from the FP in the simulation, but not for intrinsic size auto-correlation.

This is in disagreement with measurements of the real sky (Joachimi, Singh, and Mandelbaum 2015), which might be explained by systematic errors or difference in datasets.

Chapter 2

Data

2.1 Horizon-AGN Simulation

Our data are gathered from the Horizon-AGN simulation. This gave us the advantage of not having to take in to account several factors that might impact the values measured for certain properties when using real life measurements. For example we did not have to compensate for gravitational lensing or for peculiar velocity when looking at distances. The first would change the observed size (and shape) of galaxies, the second would make line-of-sight distance measurements by redshifts imprecise. The simulation comes with a downside however, because its minimum resolution is 1 kpc (Y. Dubois et al. 2014). Sizes nearing that value are thus not reliable.

The Horizon-AGN simulation is run in a periodic cube where each side is 100Mpc/h long. In the simulation a Λ CDM model is used, so a universe dominated by a cosmological constant and cold dark matter. The total matter density $\Omega_m = 0.272$, the dark energy density $\Omega_\Lambda = 0.728$, the baryon density $\Omega_b = 0.045$ and the Hubble constant $H_0 = 70.4\text{kms}^{-1}\text{Mpc}^{-1}$, so that $h = \frac{H_0}{100\text{kms}^{-1}\text{Mpc}^{-1}} = 0.704$. In total there are 1024^3 dark matter particles and 6.6×10^9 gas particles. These are started off at the time of emission of the cosmic microwave background and are evolved until now, with snapshots taken at different ages (Yohan Dubois et al. 2016). The snapshot that was used in this thesis is snapshot 761 which has a redshift of 0.06.

When enough gas particles are close enough to each other, stars are formed which can lead to both type Ia (where the energy source is fusion of carbon (and oxygen) to iron) and type II (where the energy source is gravitational potential energy) (Ryden and Peterson 2010) supernovae and to the creation of black holes (Yohan Dubois et al. 2016). From which active galactic nuclei (AGN) are

stimulated. An AGN is a black hole whose surroundings send out energy lost by falling into the black hole. In the simulation this is done by either isotropic injection of thermal energy into the surrounding gas, or by sending a cylinder of energy in two directions with a speed of 10^4kms^{-1} . These supernovae and AGN provide the otherwise empty intergalactic medium with warm and metal-rich gas (Y. Dubois et al. 2014). The AGN are necessary to form elliptical galaxies, without AGN these galaxies reform discs instead (Yohan Dubois et al. 2016).

A program called the AdaptaHOP finder (Aubert, Pichon, and Colombi 2004, updated by Tweed et al. 2009) was used to make a catalogue of the galaxies. In this, 20 neighbours of each star particle are used to compute the local density. If the density is high enough (178 times the average total matter density), the structure big enough (2 kpc radius) and it consists of at least 50 stellar particles, it is added to the galaxy catalogue (Y. Dubois et al. 2014). In the snapshot we used 126361 galaxies were identified.

The information about these galaxies includes their mass (expressed in amount of stellar particles), their galaxy level (whether the galaxy is a substructure of a larger one or not), the position in the box, the effective radius, the radial velocity divided by the radial velocity dispersion, the r-band absolute magnitude, the velocity dispersion itself and the length of the three axes of the galaxy. The effective radius being equal to the half-mass radius since each stellar particle emits approximately the same amount of light (as can be seen in Figure 2.1) and has the same weight. These physical quantities are known exactly whereas data from observations is contaminated by light or dust, projection on our night sky or gravitational lensing.

2.2 Data Cuts

In order to put this data to use, we needed to select which galaxies we wanted. We wanted to select the elliptical galaxies because we could use the Fundamental Plane to estimate their size (more on this in section 3.1). Elliptical galaxies are galaxies that look roughly like an ellipse, in contrast to spiral or irregular galaxies. Ellipticals consist mainly of older stars and are thus redder. They can withstand their own gravitational pull by being pressure supported. This means that the random movements of the stars within the galaxy keep it from collapsing. Since pressure supported galaxies have a more random distribution of star speeds than other types of galaxies, a good way to find them is selecting for a lower average speed over velocity dispersion (v/σ) (Carroll and Ostlie 2013).

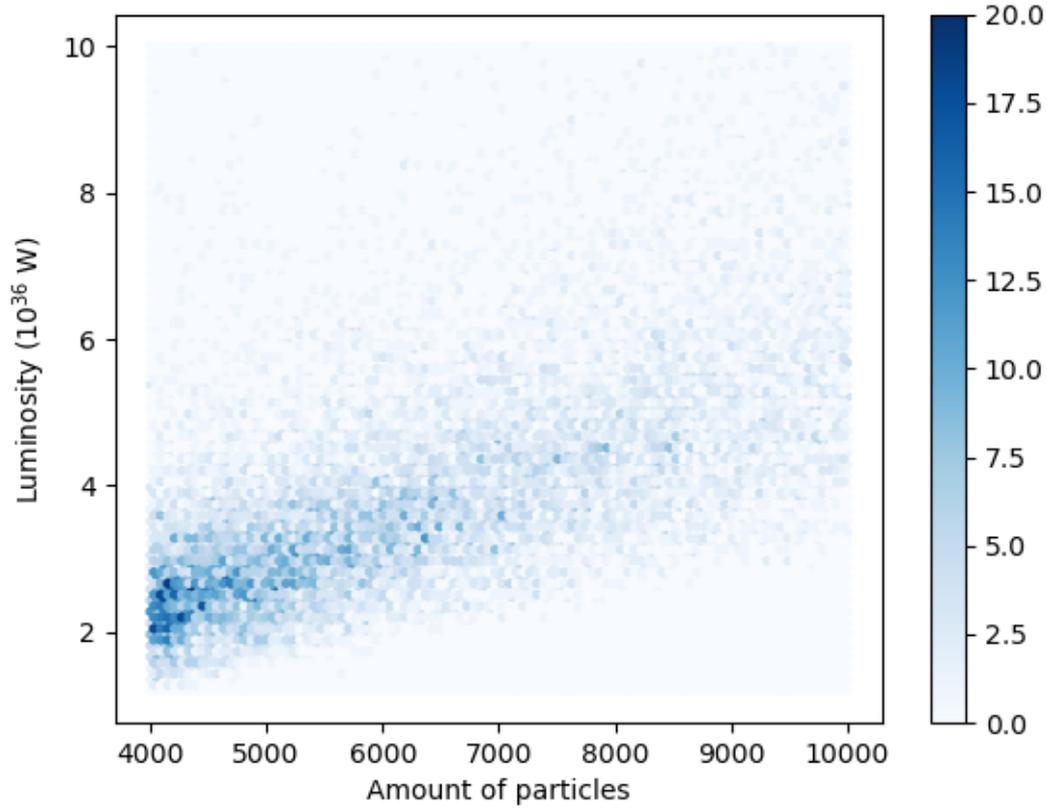


Figure 2.1: Amount of particles vs luminosity. Only galaxies with more than 4000 particles are shown, which is only one of the cuts made to the general computation. Luminosity is calculated via the formula $L = L_0 \times 10^{-0.4 \times M}$, with L_0 the zero point of the absolute magnitude scale $3.01 \times 10^{28} \text{W}$, and M the r-band absolute magnitude of the galaxy, obtained from the simulation. This graph shows an approximate linear relation between the amount of particles and the luminosity of a galaxy in the simulation.

Other factors that needed selection were the galaxy level (no substructures), minimal mass (amount of particles) and minimal size (radius). The reason for the last two cuts is the resolution of the simulation being 1 kpc (Y. Dubois et al. 2014). Any galaxies nearing that could not be fairly measured.

The following data cuts were eventually implemented, histograms of all galaxies and the selection can be seen in figure 2.2:

Galaxy level < 2

Circular velocity divided by velocity dispersion < 0.6

Amount of stellar particles > 3000

Measured radius $> 6\text{kpc}/h$

These parameters were chosen because they best selected elliptical galaxies and gave a good fit for the fundamental plane. As can be seen in image 2.3, the selected galaxies are mostly distributed over galaxy dense regions. This means the selection is not a good representation of the whole space, but this is not an issue we could avoid. After these cuts 2431 galaxies remained.

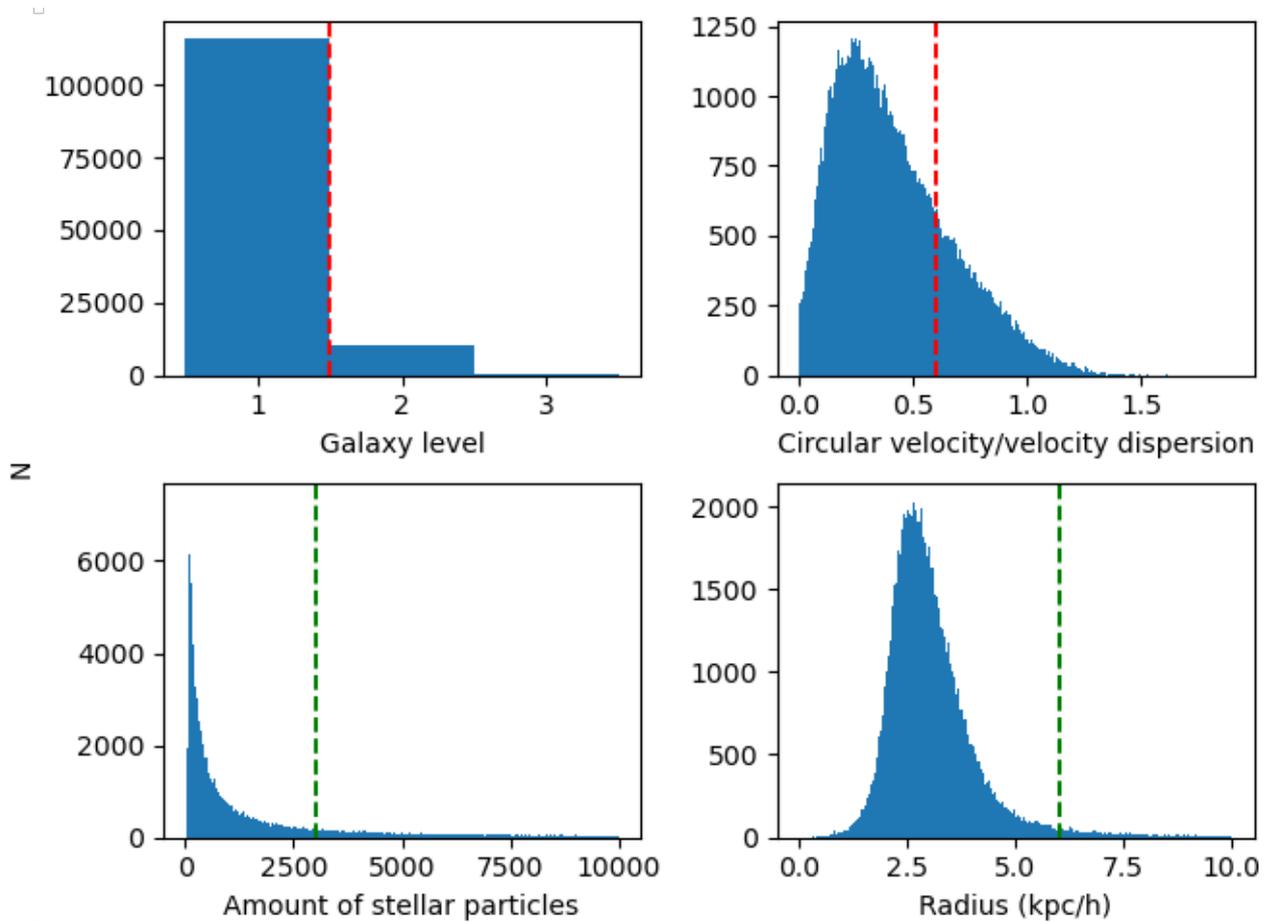


Figure 2.2: Histograms of the properties on which was selected. All 126361 galaxies in the simulation are graphed, except the 10016 galaxies which contained more than ten thousand particles which are excluded from the bottom left histogram. The red dotted lines show a maximum and the green dotted lines show a minimum for the selection. It is clear to see that elliptical, pressure supported galaxies are relatively rare.

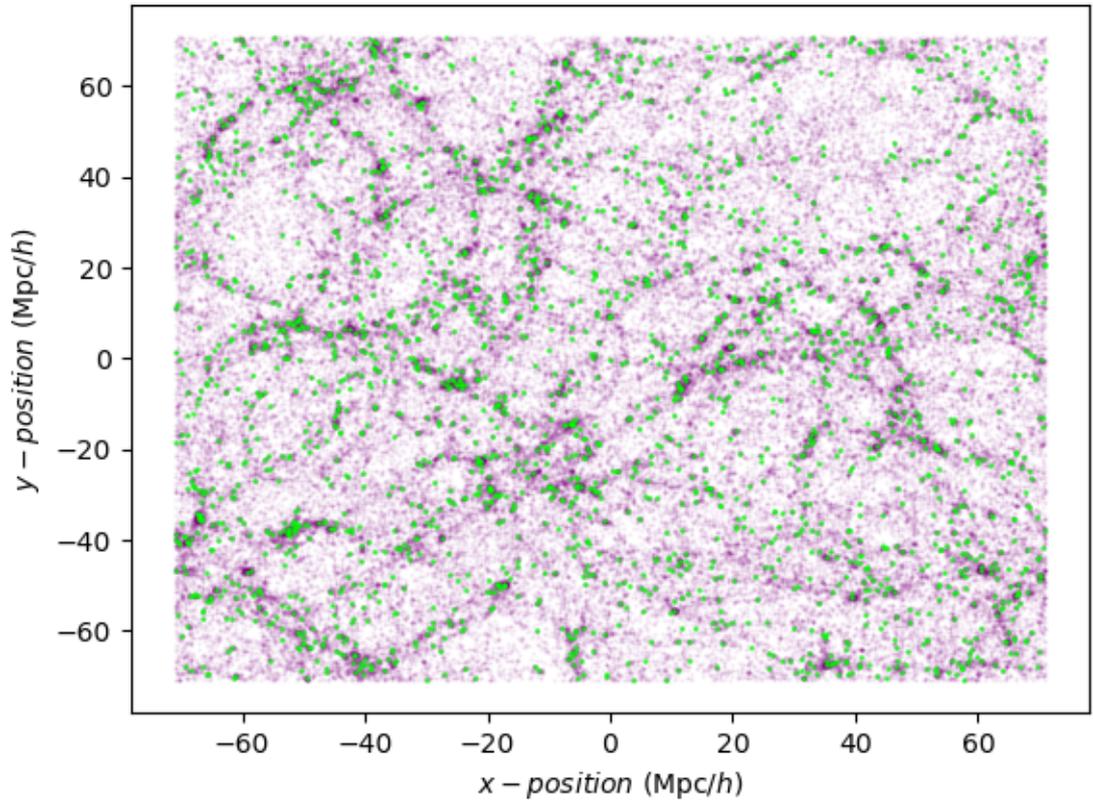


Figure 2.3: All galaxies in the simulation and the selected ones. All galaxies in the simulation are plotted in purple and the selected galaxies are plotted in lime. This is ignoring the z-axis and is thus a projection on the x-y plane. This is also clearly showing that the selected galaxies only cover the galaxy dense regions.

Chapter 3

Method

3.1 The Fundamental Plane

In order to explain how these selected galaxies were further processed, some more background information on how the Fundamental Plane works is first needed. The Fundamental Plane (FP) is a relation between the effective radius (R_{eff}), the central radial velocity dispersion (σ) and the surface brightness of a galaxy (I). It can be used to predict the size of a galaxy based on these measurable parameters. Size correlations can be measured by comparing the measured size of a galaxy to the size predicted by the FP. Because we wanted to see whether these size correlations are also present in an unlensed sample, we need to also calculate the FP, which is the following relation:

$$\log R_{\text{eff}} = a \log \sigma + b \log I + c \quad (3.1)$$

This relation has been used to describe elliptical galaxies because their pressure supported nature lends weight to the assumption that their sizes can be represented by the virial theorem. Derivations from the virial theorem can lead to a similar plane in $R - \sigma - I$ -space. These derivations would give us $a = 2$ and $b = -1$ for a and b in equation 3.1 (Saulder et al. 2013). In practice these values are an inaccurate representation of the actual relations between the quantities. Departures from the virial theorem expectation could be due to underlying assumptions like having a galaxy in virial equilibrium, the effective radius being the same as the virial radius or the line-of-sight velocity dispersion being equal to the velocity dispersion (Bernardi et al. 2003). This is however still a matter of debate (Saulder et al. 2013). Selection of the samples also impacts the parameters and this is why the Fundamental Plane is generally fitted to the sample, and sample specific. (Saulder et al. 2013)

3.2 The Pipeline

To convert the data obtained from the simulation into useful information we have written a Python file to process the data, a so called pipeline. The pipeline consists of a few parts: reading and selecting the data, fitting the Fundamental Plane, calculating the terms of the correlations for different measurements, calculating the correlation functions, estimating the error bars for the correlations and finally plotting the correlations. The selection of the data was already discussed in section 2.2. In this section I will elaborate on most other parts of this pipeline.

3.2.1 The Fundamental Plane

A few approximations were made when fitting the FP, mostly for ease of computing and because of time constraints. To start with, although the FP uses the surface brightness (I) instead of the surface mass (μ), the mass to light ratio is linear (as can be seen in Figure 2.1). So these are the same except for a factor which will be absorbed by the fitted coefficient, so we used μ . Furthermore, the approximation was made that all galaxies are circular, so surface mass was calculated by dividing the mass of the galaxy by π times the square of the galaxy radius. As can be seen in figure 3.1, most ellipticities are on the lower end, so we feel this was not too bold of an assumption. These ellipticities were calculated via the formula $e = \frac{a-b}{a}$ where e is the ellipticity, a is the major axis and b is the intermediate axis of the galaxy.

3.2.2 Fitting the Fundamental Plane

The Fundamental Plane was fitted using a function which uses non-linear least squares to fit a function to data. This fit was done with the formula for the FP, equation 3.1, with surface mass (μ) instead of surface brightness (I). The parameters from this fit were then used to calculate the expected sizes of the galaxies, based on the Fundamental Plane. The ratio λ between the actual size and the size predicted by the Fundamental Plane was then calculated and centered on zero. A positive (negative) deviation from $\lambda = 0$ would then signify a galaxy that is bigger (smaller) than predicted by the FP. This method of quantifying a deviation from the FP is taken directly from Joachimi, Singh, and Mandelbaum 2015.

$$\lambda = \frac{R_{meas}}{R_{FP}} - 1 \quad (3.2)$$

The parameters obtained for the fit of the FP are:

$$a = 38.99 \pm 0.41$$

$$b = -12.97 \pm 0.22$$

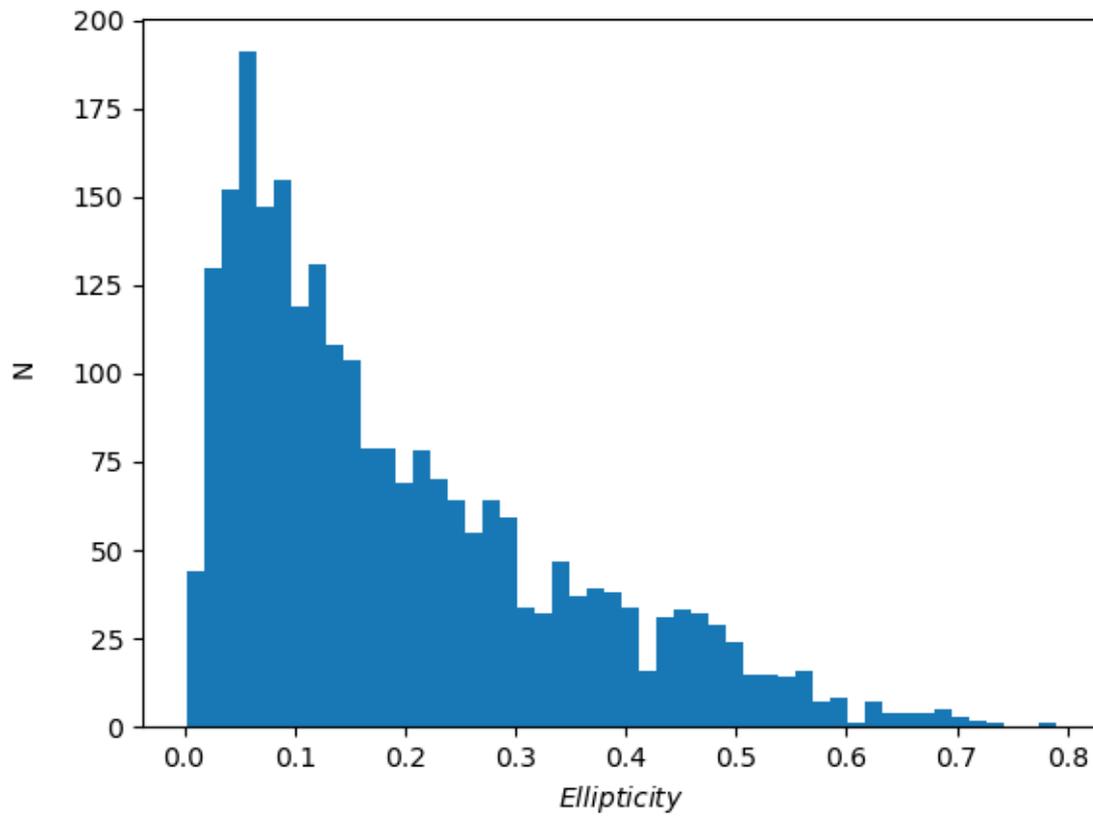


Figure 3.1: Histogram of the ellipticities of the selected galaxies. Less than five percent of selected galaxies have an ellipticity higher than 0.5, more than 77 percent have ellipticities has an ellipticity lower than 0.3. An ellipticity of 0.3 would mean the surface mass is off by a factor of $\frac{1}{7}$.

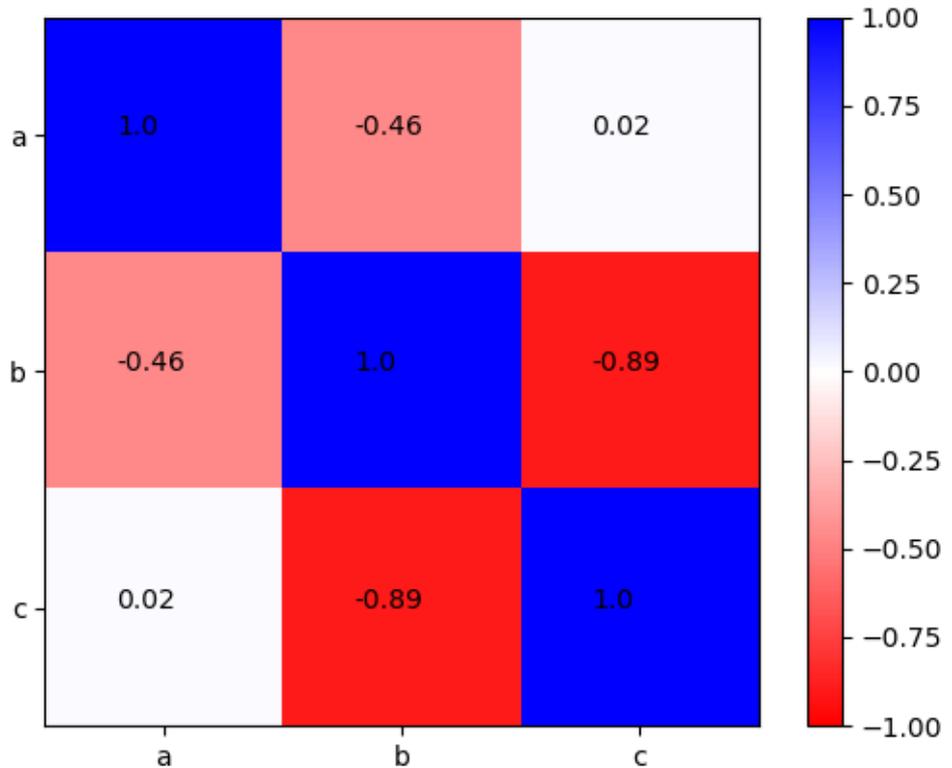


Figure 3.2: Correlation matrix for the three parameters of the Fundamental Plane fit.

$$c = 39.57 \pm 1.72$$

The correlation matrix between these parameters can be seen in image 3.2. Comparing this to the expectation we see that, like we expected from the virial theorem derivation, there is a negative correlation between the radius and the surface brightness. We also see that there is a positive correlation between the radius and the velocity dispersion. These are very different parameter values than predicted via the virial theorem (which were $a = 2$, $b = -1$ (Saulder et al. 2013)), but still give a good fit as can be seen in figure 3.3. The corresponding λ values can be seen in figure the inline graph in figure 3.3.

3.2.3 Correlation Estimators

We were interested in how this measure λ correlates both with itself ($w_{\lambda\lambda}$) and with the positions of galaxies ($w_{g\lambda}$) to see whether galaxies with certain λ s cluster more. In other words: we wanted to see whether there was a statistical relation

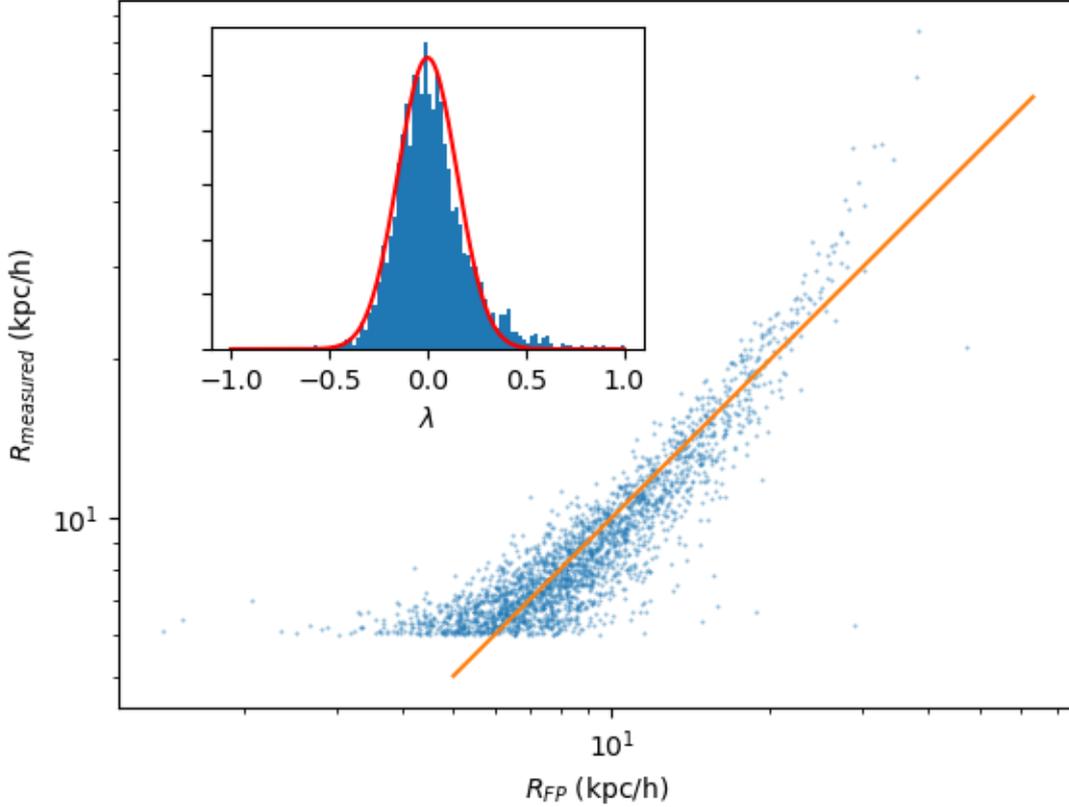


Figure 3.3: Radius as predicted by the Fundamental Plane versus radius measured. Each blue point is a galaxy that adheres to the selection made in section 2.2. The orange line indicates the $x = y$ line we expect the points to be centered around. The reason for the flat bottom is the requirement for the selected galaxies to at least have a radius of 6 kpc/h. The inset graph shows the associated λ values as calculated in equation 3.2. Nine outliers that had λ bigger than 1 are not shown, but are included in analysis. This distribution of λ s can be approximated by a gaussian distribution with $\sigma = 0.15$ as shown with the red line.

between these statistics. We also looked at the clustering of galaxies (w_{gg}) to help with interpretation of the results. To calculate the correlation functions, we followed Joachimi, Singh, and Mandelbaum 2015 and used the following estimators:

$$w_{\text{gg}} = \frac{DD - 2DR + RR}{RR} \quad (3.3)$$

$$w_{\text{g}\lambda} = \frac{F_{\lambda}D - F_{\lambda}R}{RR} \quad (3.4)$$

$$w_{\lambda\lambda} = \frac{F_{\lambda}F_{\lambda}}{RR} \quad (3.5)$$

Where D is the set of galaxies used and R are a set of points which are randomly distributed over the same area D . The random points are used to compare the possible clustering signal to the random chance that that many pairs were close together and to normalise the signal to make it independent of the sample size. To reduce noise, five times as many random points as galaxies are used which was then re-normalised.

XY means a count of all pairs with one galaxy from X and one galaxy from Y that fall within a certain bin in transverse separation. For XY with two sets of random points (RR) equation 3.9 was used for computing efficiency.

$F_{\lambda}X$ and $F_{\lambda}F_{\lambda}$ are defined by the formulas below where the sums are over all galaxies in the respective samples and where $[j|i]$ is equal to one if two galaxies are within the transverse separation bin and zero otherwise.

$$F_{\lambda}X = \sum_{i \in F, j \in X} \lambda_i [j|i] \quad (3.6)$$

$$F_{\lambda}F_{\lambda} = \sum_{i \in F, j \in F} \lambda_i \lambda_j [j|i] \quad (3.7)$$

All of these bins are in practice logarithmic bins of 0.32 in logarithmic space so we can see the correlations at different scales. The precise bin size of 0.32 was chosen to more closely resemble the figures in Joachimi, Singh, and Mandelbaum 2015.

The reason why transverse separation is used is because the positions on the projection on the sky are easier to obtain than the distance. The distance would have to be computed by measuring the redshift of a galaxy and converting that to distance within the context of an assumed cosmological model. Transverse separation is also not influenced by redshift space distortion, which line-of-sight separation is influenced by. The line-of-sight distance to a galaxy

might be bigger or smaller than the redshift suggests. This can be caused by the peculiar motion of the galaxy, which is called redshift space distortion.

Since we use a simulation we can get the actual positions and shapes and thus would not experience these problems. We however chose to use the transverse distance because this makes our results more easily comparable to the results in Joachimi, Singh, and Mandelbaum 2015 which uses observations. Using three-dimensional separation might be an interesting study for future work.

The transverse distance is taken as the distance in the x-y plane of the simulation, but each pair count described above also takes into account a maximum line-of-sight distance, which is the separation in the z direction of the simulation. Previous work has employed linear bins for the line-of-sight distance, which are then integrated over (Joachimi, Singh, and Mandelbaum 2015), but since this is equivalent to using a maximum and minimum distance, we simply restrict to pairs within a maximum line-of-sight separation. This binning followed by integrating would however be necessary for non-linear (e.g. logarithmic) bins.

The correlation functions were calculated using *k-dimensional trees*, which are a data structure used to order points. These can be used to find which points are within a certain distance of each other more rapidly than a brute force method would. An example of the code can be seen in figure 3.4.

The clustering signal w_{gg} was taken for both positive and negative λ to see any systematic differences in clustering between galaxies with $\lambda < 0$ and $\lambda > 0$, using the statistic:

$$\Delta_{gg} = \frac{w_{gg}(\lambda > 0)}{w_{gg}(\lambda < 0)} - 1 \quad (3.8)$$

3.2.4 Comparison Between Methods

To make sure these computations were done correctly we compared the amount of points counted within bins of transverse separation for randomly distributed points counted by three methods: using the *tree* functions, using a brute force method we built and using the value one would expect analytically from randomly distributed points.

The function for the analytical expectation was constructed as follows: It was assumed that, with enough points, a random distribution would net the same result as a uniform distribution and thus the latter could be assumed. This meant that the amount of pairs in a certain bin would be equal to the total amount of points times the average amount of other points within a circle around a point

```

DeDList = np.array([])
for i in np.arange(low,high,stepsize):
    #look at which pairs are within two ranges of each other
    shell1 = tree1.query_ball_tree(tree2, h * 10**float(i))
    shell2 = tree1.query_ball_tree(tree2, h * 10**float(i+stepsize))
    DeD = 0
    #loop over all galaxies
    for k in range(tree1.n):
        #see to which galaxies galaxy k has a separation between i and i+stepsize
        #zs is to check if they are not too separated in LOS distance
        indx = list((set(shell1[k]) ^ set(shell2[k])) & zs[k])
        #count how many pairs have the right separation
        DeD += len(indx)
    #add the counts to the list of counts per separation
    DeDList = np.append(DeDList, DeD)
return DeDList

```

Figure 3.4: A code snippet showing the process of counting how many pairs of galaxies are within a certain transverse separation bin. It returns a list of the amount of pairs that have a separation between 10^i and $10^{i+stepsize}$, where i goes from low to $high$ in steps of $stepsize$. *DeDList* is used to store the count of pairs for each bin in a list. The variables *shell1* and *shell2* both store a list of lists of indices. Each of the sublists refers to a galaxy in the group of galaxies in *tree1*. Each sublist consists of the indices of galaxies in the group of galaxies in *tree2* that is closer than $h \times 10^i$ or $h \times 10^{i+stepsize}$ respectively to the galaxy in *tree1* that sublist is associated with. *indx* then stores which galaxies are only in either *shell1* or *shell2* (where *shell1* is always a subset of *shell2*) and are also within the maximum line-of-sight distance. *zs* is a list of sets where each set corresponds to a galaxy in the group of galaxies in *tree1* and each set contains the galaxies in the group of galaxies in *tree2* that are within maximum line-of-sight separation (30Mpc) of the corresponding galaxy from *tree1*. *DeD* is a counter for the amount of pairs within a certain bin and is added to the end of *DeDList* after a loop over all galaxies in *tree1* is completed..

with radius equal to the upper limit of the bin, minus the same calculation but with the radius being the lower limit of the bin. We wanted this function to also be useful when calculating how many objects of one sample lie within a bin of transverse separation with another sample and we wanted to make exclude the point itself in the average amount of points within a circle around that point. This led us to equation 3.9:

$$C = \frac{(N_1 - 1) \times N_2}{d^2} \times \pi(r_1^2 - r_2^2) \quad (3.9)$$

Where C is the amount of pairs in the bin, N_1 the amount of points in sample 1, N_2 the amount of points in sample 2, d the length of the side of the box, r_1 the maximum range of the bin and r_2 the minimum range. This equation is also used for computing RR in functions 3.3, 3.4 and 3.5.

The comparison of different counting methods at first led us to believe the *tree* method and the analytical method were in agreement, as can be seen in figure 3.5. This prompted us to conclude we were computing the amount of pairs in the bins correctly and that the brute force method would need more development. We continued using the *tree* method for everything except the counts of pairs which used two random samples (RR in equations 3.3, 3.4, 3.5). In those cases we used equation 3.9.

However, a later recreation of this graph, with updates to most parts of the code used, gave us figure 3.6. As the figure shows, the *by tree* and *analytically* method are no longer in agreement. The fact that the deviation between both methods is a constant logarithmic separation suggests a systematic error in either our usage of the *tree* functions or our analytical calculations. We used both of those methods to calculate our results and thus conclude there is most likely a systematic error in our results. Since both methods seem to scale correctly with the size of the bins, we still believe our results to have value, just that they might be off by a factor.

3.2.5 Error Calculation

The standard error of the correlations was calculated using jackknife resampling. Jackknife resampling is a technique in which a part of the data is removed, after which the statistic of interest (x) is calculated with the available data and then the data are placed back. This is repeated for all subsamples of the data and these are then treated as individual measurements (x_i in the error calculation.

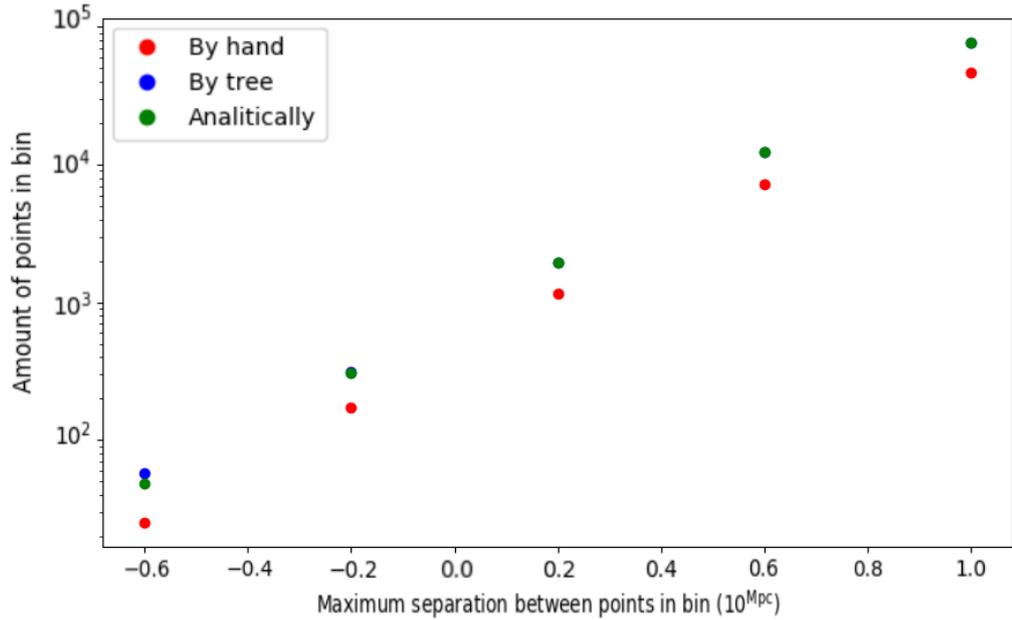


Figure 3.5: The amount of points within bins of transverse separation, the upper limit of which is on the x-axis, calculated by three different methods. An arbitrary amount of points were randomly distributed over a $100/h$ by $100/h$ square. The three different methods are the brute force method we designed ourselves, the method by using *tree* functions as shown in figure 3.4 and the analytical expectation as described in equation 3.9. The blue points cannot be seen for the last three points as they are under the green ones, which shows us agreement between the two associated methods. This is an older version of the graph shown in figure 3.6 which uses updated code and a different amount of points.

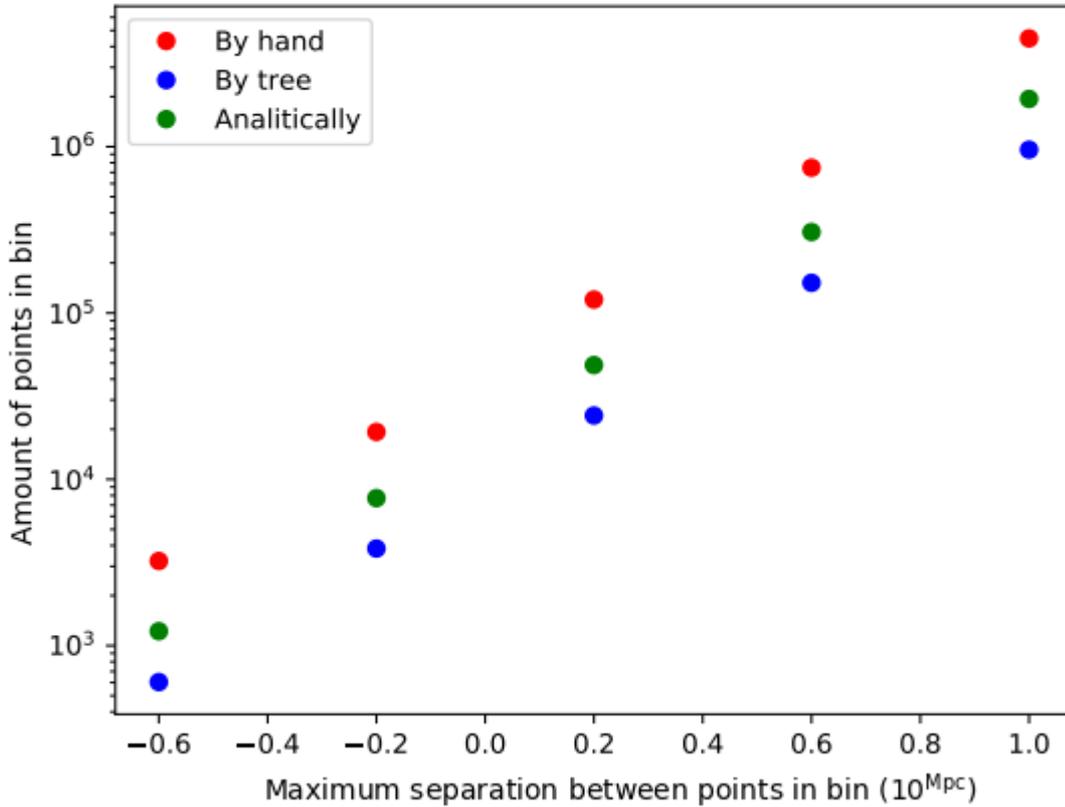


Figure 3.6: The amount of points within bins of transverse separation, the upper limit of which is on the x-axis, calculated by three different methods. An arbitrary amount of points were randomly distributed over a $100/h$ by $100/h$ square. The three different methods are the brute force method we designed ourselves, the method by using *tree* functions as shown in figure 3.4 and the analytical expectation as described in equation 3.9. The blue and green points are not in agreement and thus show an inconsistency in computation. The inconsistency suggests a bug in either computation method (or in both). This is a newer version of the graph shown in figure 3.5 which uses older code and a different amount of points.

The Jackknife error is expressed as: (McIntosh 2016)

$$SE(x)_{jack} = \left(\frac{n-1}{n} \sum_{i=1}^n (x_i - x_{(\cdot)})^2 \right)^{1/2}$$

Where $x_{(\cdot)}$ is the average of the Jackknifed samples and n is the amount of Jackknifed samples.

In our case, eight subsamples were taken by removing a cube with sides of 50 Mpc/ h from a different corner of the simulation box (which has sides of 100 Mpc/ h) for each subsample. Correlations were then computed for each jackknifed set of data after which the covariance matrix between the subsets was computed. The diagonal of this matrix is the variance (P.J.S. van Capel 2020) and this easily gave us the jackknife errors by normalising the covariance and then taking the square root of the variance.

Chapter 4

Results

4.1 Correlation Estimators

The correlation estimators described in equations 3.3, 3.4 and 3.8 are plotted in figure 4.1. The last seven points of the bottom graph are also plotted in figure 4.2 to better see their actual values.

$w_{g\lambda}$ is positive, which suggests that intrinsically large galaxies ($\lambda > 0$) are more likely to be located in galaxy-dense regions than intrinsically small galaxies ($\lambda < 0$). Comparing this to the graph showing the same estimators in Joachimi, Singh, and Mandelbaum 2015, we see that they concluded the opposite: Their $w_{g\lambda}$ is negative and shows a much more significant deviation from zero.

We see that $R \times w_{\lambda\lambda}$ does not clearly deviate from zero and fluctuates to both sides of zero. We thus conclude there is no evidence that smaller-than-expected ($\lambda < 0$) galaxies are closer to other smaller-than-expected galaxies or that bigger-than-expected ($\lambda > 0$) galaxies are closer to other bigger-than-expected galaxies. When comparing this to Joachimi, Singh, and Mandelbaum 2015 we see that they did find a significant positive signal that does suggest galaxies with λ values of the same sign are closer together.

We do see similar trends between $R \times w_{\lambda\lambda}$ and $R \times w_{g\lambda}$ in that from the fourth point, they trend upwards with significant errors, then briefly back down, and then up again. This might suggest that there is one or two specific ranges in which there is a relation between λ and clustering. But since we are unsure whether this signal is significant, we are hesitant to conclude this.

The most notable thing about Δ_{gg} is that it clearly trends towards zero, as can also be seen in 4.2. Except for the first three bins, it can be concluded that there is no significant difference in clustering between galaxies with positive and negative λ s. This is consistent with the absence of a high amplitude signal for either $w_{\lambda\lambda}$ or $w_{g\lambda}$. Comparing this to Joachimi, Singh, and Mandelbaum 2015

again, we see that they did find a positive correlation between clustering and negative λ_s .

4.2 Concerns

The three bins with smallest R have most likely too small of a maximum distance between galaxies to have them be useful. For these bins only 2, 20 and 80 pairs were counted respectively. This is in line with expectations according to the analytical solution for randomly distributed galaxies of equation 3.9, which gives us 3, 15 and 80 respectively. In future work this might be remedied by using a broader selection of galaxies, thus also giving significant data for closer pairs.

The last bin, with biggest R , might also have to be ignored, since its maximum R (100Mpc) is nearing the box size of the simulation ($100\text{Mpc}/h \approx 142\text{Mpc}$). This might lead to pairs being counted more than once.

All of these results are preliminary because the inconsistency between the expectation of what our counting methods should result in and what we analytically expect to see, as outlined in section 3.2.4. A continuation of this study might find more conclusive results or results closer to those found in Joachimi, Singh, and Mandelbaum 2015. For now, we can only speculate on the impact of the difference between the analytical formula and our counting method.

As mentioned in section 3.2.1, we assumed circular galaxies to calculate the surface mass because of time constraints. Since the shapes of the galaxies are obtainable, future work might want to calculate the actual surfaces of galaxies. This would in theory improve the FP fit and with it, the precision of this whole study.

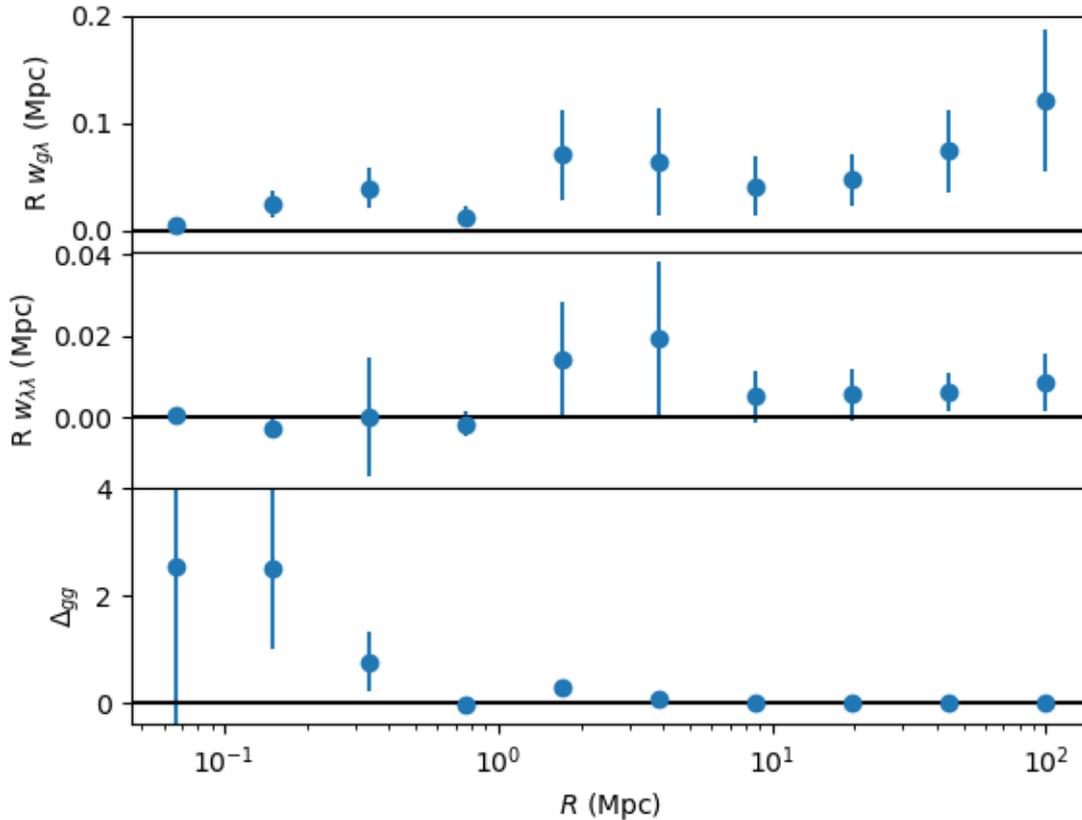


Figure 4.1: *Top/Middle:* Correlation estimators between galaxy density and galaxy size $w_{g\lambda}$ as described in equation 3.4 and correlation estimators between the galaxy size and the galaxy size of nearby galaxies $w_{\lambda\lambda}$ as described in equation 3.5 respectively. Both are in bins of transverse separation between $10^{-1.5} - 10^2$ Mpc and multiplied by the upper limit of their respective bins. *Bottom:* Ratio between the clustering correlation of galaxies that are bigger than expected ($\lambda > 0$) and galaxies that are smaller than expected ($\lambda < 0$), as described in equation 3.8, in bins of transverse separation between $10^{-1.5} - 10^2$ Mpc. The last seven points are also plotted in figure 4.2. These graphs are an attempt at recreating figure 4 in Joachimi, Singh, and Mandelbaum 2015 with a different data set.

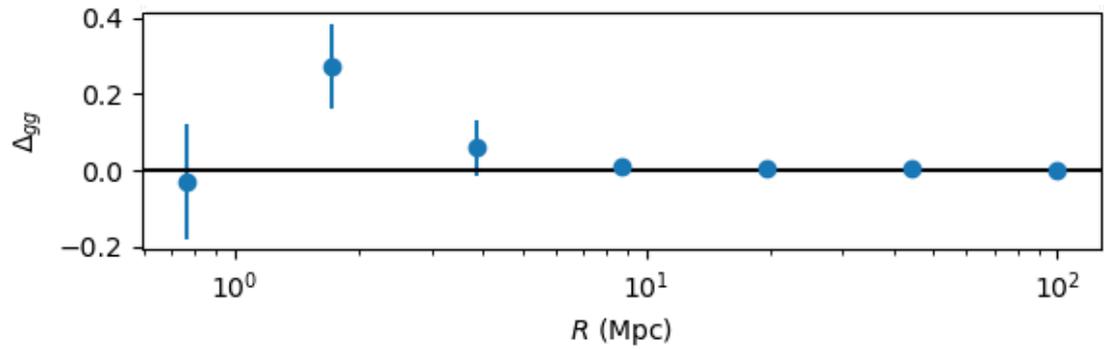


Figure 4.2: Ratio between the clustering correlation of galaxies that are bigger than expected ($\lambda > 0$) and galaxies that are smaller than expected ($\lambda < 0$) as described in equation 3.8 in bins of transverse separation between $0.76 - 10^2$ Mpc. These and three points with smaller separations are also plotted in the bottom graph in figure 4.1.

Chapter 5

Conclusion

We have shown, with reservation, there is a positive correlation between the clustering of galaxies and the deviation of their size from the FP in the Horizon-AGN simulation. This is in contrast with real observations where this correlation is negative (Joachimi, Singh, and Mandelbaum 2015). We have also found no evidence for auto-correlation between deviations of size from the FP of nearby pairs of galaxies in the Horizon-AGN simulation. This again contrasts with real observation where this correlation was found to be present and positive (Joachimi, Singh, and Mandelbaum 2015).

To do this we have applied several techniques including: Choosing data cuts, fitting functions to data, automating processes while still being able to change parameters, developing analytical calculations for random pair counts, comparing different methods of computing the same statistic for validation, computing correlation functions and estimating errors.

Our results suggest the Horizon-AGN simulation is not a good way to re-search the impact of intrinsic sizes on measurements. Alternatively, it suggests intrinsic sizes do not necessarily adhere to the correlations as found in Joachimi, Singh, and Mandelbaum 2015, but that they might vary more per dataset. More realistically, it means the bugs in our computations have made our results unreliable and this study is simply a good step in answering whether the above statements are true.

In future work one could repeat this study without the bugs and see whether there are intrinsic sizes in the Horizon-AGN simulation. Another good way to improve upon our results would be to calculate the surface masses using elliptical surfaces for the galaxies instead of the circular surface approximation we have used. We believe the difference to be small, but it would be an improvement because it is closer to reality.

Future work could also repeat this study with different data constraints. Looser data constraints (and thus more galaxies) might lead to more detectable signal, especially at smaller separations.

It might also be interesting to try this method with galaxies that are not necessarily elliptical. The theory does not suggest other galaxies will follow the FP, but finding a similar relation or fitting a FP anyway might still lead to finding intrinsic sizes in other types of galaxies.

Another interesting extension of this study might be to repeat these measurements for snapshots of the simulation with higher redshifts and include a redshift term in the FP, as done in Joachimi, Singh, and Mandelbaum 2015, and see how that would impact the computed correlations.

Seeing as a new Horizon simulation with a box size of 1049 comoving Mpc^3 was ran (Lee et al. 2020), this study might bear repeating for the new simulation. The bigger box size would decrease the possible influence of the periodicity and provide more data for measurements of large scales, such as in this study.

Since we computed our correlations in transverse separation bins, one could also expand upon our results by using bins that use three dimensional separation. Using transverse separation is very useful for real life observations as explained in 3.2.3, but these advantages are not present for simulations. Fully using the third available dimension will give more information.

Lastly, more information might be gathered on the relation between the correlations and the matter density in the region around the galaxies, not just the galaxy density. Some model is proposed in Joachimi, Singh, and Mandelbaum 2015, but where they are using theoretical models, one could also use the abundance of information in a simulation to their advantage and measure the matter density in the simulation directly and compare that to the theoretical density and look at the correlation between these densities and intrinsic sizes.

Bibliography

- Aubert, D., C. Pichon, and S. Colombi (Aug. 2004). “The origin and implications of dark matter anisotropic cosmic infall on L haloes”. In: *Monthly Notices of the Royal Astronomical Society* 352.2, pp. 376–398. ISSN: 1365-2966. DOI: 10.1111/j.1365-2966.2004.07883.x. URL: <http://dx.doi.org/10.1111/j.1365-2966.2004.07883.x>.
- Bernardi, Mariangela et al. (Apr. 2003). “Early-Type Galaxies in the Sloan Digital Sky Survey. III. The Fundamental Plane”. In: *The Astronomical Journal* 125.4, pp. 1866–1881. ISSN: 1538-3881. DOI: 10.1086/367794. URL: <http://dx.doi.org/10.1086/367794>.
- Carroll, Bradley W. and Dale A. Ostlie (July 2013). *An Introduction to Modern Astrophysics*. 2nd ed. Edinburgh Gate, Harlow, Essex CM20 2JE, England: Pearson Education, Limited. ISBN: 978-1-292-02293-3.
- Ciarlariello, Sandro and Robert Crittenden (Aug. 2016). “Modelling the impact of intrinsic size and luminosity correlations on magnification estimation”. In: *Monthly Notices of the Royal Astronomical Society* 463.1, pp. 740–755. ISSN: 1365-2966. DOI: 10.1093/mnras/stw2052. URL: <http://dx.doi.org/10.1093/mnras/stw2052>.
- Dubois, Y. et al. (Aug. 2014). “Dancing in the dark: galactic properties trace spin swings along the cosmic web”. In: *Monthly Notices of the Royal Astronomical Society* 444.2, pp. 1453–1468. ISSN: 0035-8711. DOI: 10.1093/mnras/stu1227. eprint: <https://academic.oup.com/mnras/article-pdf/444/2/1453/24039446/stu1227.pdf>. URL: <https://doi.org/10.1093/mnras/stu1227>.
- Dubois, Yohan et al. (Sept. 2016). “The Horizon-AGN simulation: morphological diversity of galaxies promoted by AGN feedback”. In: *Monthly Notices of the Royal Astronomical Society* 463.4, pp. 3948–3964. ISSN: 1365-2966. DOI: 10.1093/mnras/stw2265. URL: <http://dx.doi.org/10.1093/mnras/stw2265>.
- Joachimi, Benjamin, Sukhdeep Singh, and Rachel Mandelbaum (Sept. 2015). “Detection of spatial correlations of Fundamental Plane residuals, and cosmological implications”. In: *Monthly Notices of the Royal Astronomical Society* 454.1, pp. 478–488. ISSN: 1365-2966. DOI: 10.1093/mnras/stv1962. URL: <http://dx.doi.org/10.1093/mnras/stv1962>.

- Lee, Jaehyun et al. (2020). *The Horizon Run 5 Cosmological Hydrodynamic Simulation: Probing Galaxy Formation from Kilo- to Giga-parsec Scales*. arXiv: 2006.01039 [astro-ph.GA].
- McIntosh, Avery (2016). *The Jackknife Estimation Method*. arXiv: 1606.00497 [stat.ME].
- P.J.S. van Capel, dr. (2020). *Data-acquisitie Toegepaste analyse: Verwerking Dictaat*.
- Ryden, Barbara and Bradley M. Peterson (2010). *Foundations of Astrophysics*. 1301 Sansome St., San Francisco, CA 94111, USA: Addison-Wesley. ISBN: 978-0-321-59558-4.
- Saulder, Christoph et al. (Aug. 2013). “Calibrating the fundamental plane with SDSS DR8 data”. In: *Astronomy & Astrophysics* 557, A21. ISSN: 1432-0746. DOI: 10.1051/0004-6361/201321466. URL: <http://dx.doi.org/10.1051/0004-6361/201321466>.
- Tweed, D. et al. (Aug. 2009). “Building merger trees from cosmological N-body simulations”. In: *Astronomy Astrophysics* 506.2, pp. 647–660. ISSN: 1432-0746. DOI: 10.1051/0004-6361/200911787. URL: <http://dx.doi.org/10.1051/0004-6361/200911787>.

Appendix A

Acknowledgements

I would firstly like to thank my thesis supervisors Elisa and Harry. They helped me a lot over the last few months by providing me with a subject and resources about that subject, guiding me at every step and helping me with every problem I encountered. Even though I have never met either in person, just via video calls, they provided me with all the help I could want, Elisa did so even while being nine months pregnant and again a few weeks after her child was born.

Secondly, I would like to thank the people who worked on the *SciPy*, *NumPy* and *Matplotlib* libraries and everyone who worked on Python in general. The open source resources provided by these people makes theses like mine possible.

Lastly, I would like to thank my girlfriend, Eline Kempkes, who helped me tremendously with motivating me. She helped me plan when I was overwhelmed and calmed me down when I was stressed. It would be unfair to not also mention the times she also helped me with Python or the physics behind it all. Without her, I would still not be finished as of writing this (8 hours before the deadline).

