Universiteit Utrecht

Opleiding Natuur- en Sterrenkunde

Intrinsic size correlations of galaxies

Bachelor Thesis

Dana Sophia Westbeek

Supervisors:

dr. H.S. Johnston First Supervisor
Institute for Theoretical Physics

dr. N.E. Chisari Second Supervisor
Institute for Theoretical Physics

June 16, 2021

Contents

1 Introduction

Of high interest to cosmologists is the structure of the Universe today, and how this arose from the processes that happened in the early Universe. With this, cosmologists gain information about the evolution of our Universe, and summarise this in a cosmological model. This is a model that describes the components of the Universe and their contribution to the expansion of the Universe [Ryden, 2017]. One of these models is the ΛCDM model. It claims that the Universe was created in a hot Big Bang, and has been cooling down and expanding in volume ever since. The energy density knows contributions from dark energy (the Λ in the name), cold dark matter (the CDM in the name), and in smaller amounts from baryons, radiation, and neutrinos. It is the model that best describes multiple independent measurements on the Universe while not requiring an excessive amount of fundamental parameters, which makes it the concordance model [Johnston, 2020].

By far the oldest observable to probe the Universe is light. In the time of Ancient Greece, people were observing the moon and stars in the night sky with the naked eye, until the first telescopes appeared around 1600. Then, also planets, galaxies and nebulae could be recorded [King, 2003]. Those are beautiful, but are all examples of ordinary baryonic matter, which actually is only a fraction of all matter in the Universe. Much more abundant is dark matter. We cannot observe dark matter directly, because it does not emit light, but we can still use light to look for dark matter. This is because it has mass, and therefore curves the space-time of the Universe. This means that dark matter causes light to follow a curved path instead of a straight path to our telescopes. If we measure changes in the shape or size of galaxies at certain positions in the sky, this tells us something about the distribution of dark matter in the Universe. The shear (shape-distortion) and magnification effects are called weak gravitational lensing [Johnston, 2020].

The fundamental assumption here, is that the shapes and sizes of galaxies in a certain position were uncorrelated in the first place. Otherwise, we would observe effects similar to those caused by weak lensing, while not being caused by the presence of dark matter, but by the intrinsic size correlations of the galaxies themselves. This possible contamination of measurements makes investigating whether intrinsic size correlations of galaxies are present in our Universe so important [Joachimi et al., 2015]. Previous research by Singh et al. [2020] pointed to the presence of intrinsic size correlations of galaxies, and also Joachimi et al. [2015] found these correlations for elliptical galaxies.

In our research, we will look for the intrinsic size correlations of elliptical galaxies, as in the two above studies, and in addition pay attention to the clustering and size correlations of spiral galaxies, which is a new line of inquest in cosmology. The clustering of matter, that forms the matter density field, will also be investigated. Clustering is quantified by use of two-point correlation estimators. We use simulated data, especially to quantify two variables, named galaxy bias and 'size bias'. The galaxy bias is the clustering of galaxies relative to the clustering of matter. The size bias is analogous to this, and describes the deviation from the matter density field that correlates with intrinsic size. It contributes to the galaxy bias, as this is a function of all galaxy properties.

To hypothesise how we expect the clustering of elliptical and spiral galaxies to be, we must look at the formation process of these galaxies. The oldest of the two types are the ellipticals, that formed in the densest regions of the Universe, also called the nodes of the cosmic web. Ellipticals are quite massive, and formed through the mergers of multiple smaller galaxies. We therefore expect elliptical galaxies to be more strongly clustered. Younger are the spirals, which live in filaments between the cosmic nodes. They accrete gas that was left over by the ellipticals. We thus expect spirals to also be clustered, and in addition that large spirals cluster together.

Our report starts with a description of the simulation data we use in our analysis. Then, a section follows in which we discuss our research method, including the manner of intrinsic size estimation, implementation of the correlation functions, and computation of the error in our measurements. In the results section, we list our findings and compare them to the results of other studies. We summarise our main results, name shortcomings of our research and ideas for further investigation in the conclusion. The report ends with a thanking word to everyone who guided and helped me in this project.

2 Data

For this research project, we limit ourselves to the data available from the Horizon-AGN simulation. Really, this is not a limitation at all. Simulations can provide us with very pure information, measured from a theoretical universe formed within the bounds of a certain cosmological model. In simulations, we do not have to be concerned about our measurements being contaminated by estimates of cosmological distances or the effects of gravitational lensing [Singh et al., 2020]. This is especially of use now, since it is difficult to distinguish weak lensing and intrinsic size correlations of galaxies in observations [Ciarlariello et al., 2014], and we thus want to disconnect the effects of weak lensing from observed galaxy sizes to assess the contamination of real measurements.

The cosmological model implemented in the simulation is the Λ cold dark matter (ΛCDM) cosmology. Details on the initial conditions and cosmological parameters chosen in accordance with the ΛCDM cosmology for the simulation can be found in Dubois et al. [2014]. A total of 1024^3 dark matter particles are placed in a box with dimensions 100 Mpc/h. This is a unit of length, Mpc, scaled with the dimensionless Hubble parameter h ($h = 0.704$). Then, the simulation starts running in `Ramses`, a deformable cell-based code that numerically solves hydrodynamic equations. The evolution of gas and particles is monitored, and more advanced structures and objects are allowed to form under specific conditions. So can a star be formed when the hydrogen number density in a region is higher than 0.1 H cm^{-3} [Dubois et al., 2014].

The aspect that makes this simulation of the Universe so special, is that it is hydrodynamical. The simulation includes feedback processes that dictate the interactions of particles. One of these feedback processes comes from black holes that are allowed to form within this simulation. We know that black holes are commonly found in the centers of galaxies, our own Milky way galaxy has one too. Galaxies with the possession of a black hole are given the name of active galactive nuclei (AGN), the AGN-part in the name Horizon-AGN originates exactly from this feature. Black holes are formed in regions where the Jeans criterion is violated [Dubois et al., 2012]. This happens if the mass of the region exceeds the Jeans mass, the critical mass at which the thermal energy and gravitational forces are balanced. Then, the outward pressure is insufficient in preventing a gravitational collapse, and a black hole is born [Ryden, 2017]. Their presence changes the world pertaining to ellipticals, because now, massive ellipticals may be formed. In the ΛCDM model, ellipticals are formed through a series of mergers. Without AGN-feedback, a galaxy would replenish its cold gas content to form again a disk, but this is now prevented from happening [Dubois et al., 2016]. Evidence for this can be provided by the stellar mass function, a function of the effective radius of galaxies against their stellar mass. We recreated this figure from Dubois et al. [2016] in figure 1, and confirm their findings. We see that at higher redshift, the same stellar mass is concentrated in a smaller radius, meaning that galaxies are more compact at higher redshift. This supports the conclusion that the older elliptical galaxies were formed through merges. Gas and stars both react differently to collisions. Where gas friction leads to cooling and contraction, stars are interacting only gravitationally. Combining this with the theory that younger galaxies (at higher redshift) have a higher gas content than the older ones (at lower redshift), we indeed expect higher redshift galaxies to be more compact [Johnston, 2020].

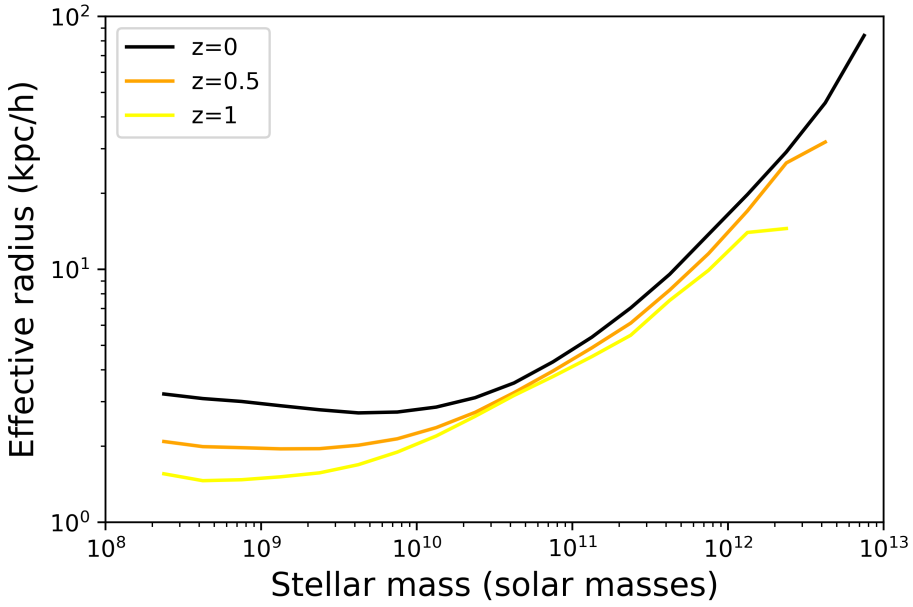

Figure 1: **The size-mass relation.** The effective radius of galaxies in the Horizon-AGN simulation is plotted against the mass, for different redshifts. We observe that galaxies are more compact at higher redshift. Figure recreated from Dubois et al. [2016].

While the simulation is running snapshots are taken at different redshifts, from z=0 to z=1.2. We will only use redshift 0 for this project, which is a recreation of the present-day Universe. The snapshots contain measurements on a collection of different properties of galaxies. As size estimator for our fundamental planes (I will talk about that later), we will use the effective radius as well as the virial radius. The effective radius is a mean half-mass radius, and is measured from the geometric mean of three half-mass radii, estimated from two dimensional projections of the stellar density along each axis of the box [Dubois et al., 2016]. The virial radius, the radius within which gravitational potential and kinetic energy are linearly proportional, is probably found by measuring these two energies on particles in a certain radius, while increasing this radius until the virial theorem is satisfied. The two size estimators are different in their range of values, yet suitable for comparison in our analysis, as we shall see later on.

Parts of our analysis will use a subsampled matter catalogue from the Horizon-AGN simulation, containing stars, gas, dark matter, and black hole particles. The subsampling was done randomly. To classify elliptical and spiral galaxies in this, we use the V/σ criterium, where V is the average velocity of objects in the galaxy, and σ the dispersion in this velocity. Elliptical galaxies are dispersion-dominated and thus have a low V/σ, while spiral galaxies are rotation-dominated and show a high V/σ. We chose our threshold to be $V/\sigma = 0.6$. Furthermore, we made a selection of the data from the simulation with which we will perform our analysis. We want to exclude any substructures that may occur in a galaxy from our

measurements. Then, due to the resolution of the simulation being 1 kpc, we demand the number of stellar particles to be above 3000 and the measured radius above 6 kpc/h. This resulted in 2432 elliptical galaxies and 795 spiral galaxies to work with. Figure 2 shows a projection of the simulation box onto the xy-plane. All galaxies that were present in the simulation are coloured as light salmon dots. The galaxies we selected, are coloured darker, with blue dots for selected ellipticals and pink dots for selected spirals. We observe that our galaxies are not uniformly distributed over the box, and are positioned largely within the regions of highest density in the simulation, but do lie in all corners of the box, which otherwise would have been problematic in analysis.

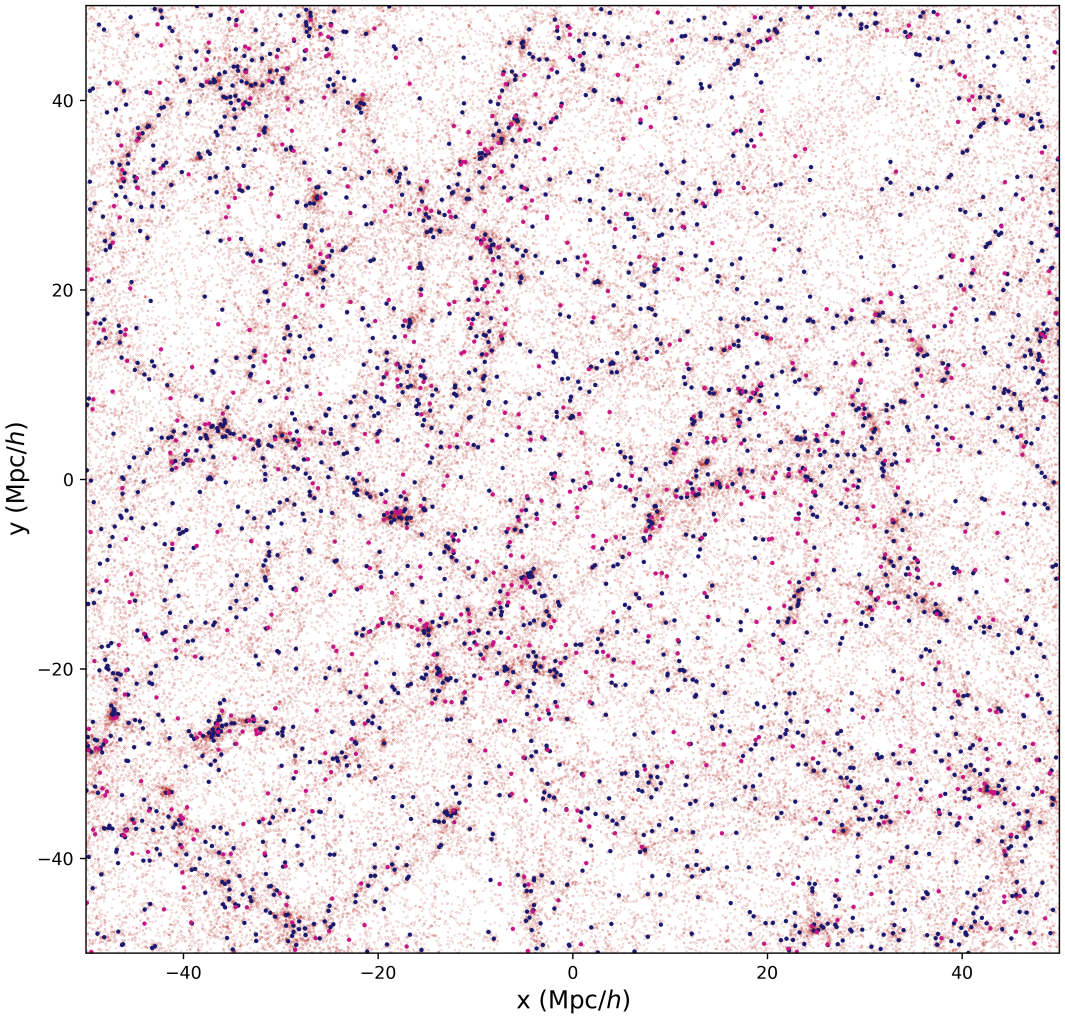

Figure 2: **The distribution of galaxies in the simulation box.** A two-dimensional projections of the positions of the galaxies is shown, with the y-coordinates on the vertical axis and the x-coordinates on the horizontal axis. Unselected galaxies are represented by salmon coloured dots, selected ellipticals are blue dots, and selected spirals are pink dots.

3 Method

For measuring galaxy bias (which was the clustering of galaxies relative to the underlying density field), all information we need, i.e. the x-, y- and z-positions of the galaxies and other objects, is directly accessible from the Horizon-AGN simulation data [Dubois et al., 2016]. However, for measuring size bias (which was the clustering of intrinsic galaxy sizes relative to the underlying density field), we need an idea of the distribution of intrinsic galaxy sizes that are expected from theory. To realise this, we use the fundamental plane, which is an equation that predicts the intrinsic size of a galaxy from other known properties Singh et al. [2020]. For elliptical galaxies, we can start from (one of the many formulations of) the virial theorem, a well-known relation in mechanics connecting the kinetic and potential energy of a system:

$$m = \frac{K_V \sigma^2 R}{G}, \tag{1}$$

in which the mass m is expressed as a function of a constant K_V taking into account the conversion from three-dimensional to two-dimensional projected data, the velocity dispersion σ, the size of a galaxy R, and the gravitational constant G. Writing the surface brightness I as $I = \frac{L}{2\pi R^2}$ using the luminosity L and size R, we find:

$$R = \frac{K_V}{2\pi G} \left(\frac{m}{L}\right)^{-1} (I)^{-1} \sigma^2, \tag{2}$$

which we map onto logspace as:

$$\log(R) = 2\log(\sigma) - \log(I) + \log(K_V) - \log\left(\frac{m}{L}\right) - \log(2\pi G). \tag{3}$$

Here, the latter three terms on the right hand side of the equation are (approximately) constant under the assumption of homology, and can be summarised in a constant called the zero-point ZP of the equation, which is constrained through observations of galaxy samples:

$$\log(R) = 2\log(\sigma) - \log(I) + \mathrm{ZP}. \tag{4}$$

This defines a plane in $(\log R - \log\sigma - \log I)$-space [D'Onofrio et al., 2017]. In practice however, data tends to deviate somewhat from this virial theorem derivative. This is because every galaxy has a unique history of formation that influences its properties and prevents an exact replication of the plane for all galaxies [Johnston, 2020]. It is therefore that we allow for the plane to tilt by fitting the coefficients in $(\log R - \log\sigma - \log I)$-space to our sample of galaxies. In other words, we are going to use the following equation for the fundamental plane:

$$\log(R_e) = a\log(\sigma) + b\log(I) + c, \tag{5}$$

and let a, b, and c be so that the plane fits the data best, as also done by Joachimi et al. [2015]. The effective radius R_e served as an estimate for the size R here, and the velocity dispersion σ was measured from stellar particle motions in the simulation. The surface brightness I of galaxies was not measured directly in the Horizon AGN-simulation, however could be computed from their magnitude using the bolometric magnitude of the sun:

$$L = L_{\odot} \cdot 10^{0.4\left(M_{bol} - M_{bol,\odot}\right)}, \tag{6}$$

where we approximate the bolometric magnitude M_{bol} of the galaxies by their absolute magnitude M. The surface brightness I then follows from dividing by the surface area A of an ellipse, $A = \pi ab$, with a and b half the lengths of the minor and major axis respectively [Knapp, 2011]. Thinking of the dependence of the surface brightness I on the luminosity L, we can see the elliptical fundamental plane as an extended version of the Faber-Jackson relation [Shen et al., 2001]:

$$L \propto \sigma^4, \tag{7}$$

a law that states that the luminosity L of an elliptical galaxy is proportional to the velocity dispersion σ to the power of 4, where in the fundamental plane the size R is introduced to reduce scatter.

That brings us to fitting our elliptical galaxies to the fundamental plane in equation 4, from which the results can be seen in the top two panels of figure 3. The first panel shows the relation between the effective radius R_e on the horizontal axis and the radius as predicted by equation 4 on the vertical axis. Galaxies are plotted as blue dots whose size is scaled by V/σ, the grey line is the one-to-one relation between measured sizes (from the simulation data) and predicted sizes (from the fundamental plane). This plot reveals two outliers, where the effective radius R_e is much smaller than predicted by the fundamental plane. For both of the points the value of V/σ is low and not near our boundary of $V/\sigma = 0.6$, which rules out the possibility that they are misplaced spirals. We also see that the size distribution of galaxies is not uniform, there are relatively more smaller galaxies than larger ones. This is expected given the theory on structure formation in the Universe. Elliptical galaxies accrete mass by merging with other galaxies, and only rare peaks in the matter density field could have produced massive ellipticals, which makes that these large objects are few in number Johnston [2020]. The second panel presents the amount of correlation between the fitted parameters a, b, and c. We found a = 0.96, b = -0.31, and c = 11.44, from which only b and c were strongly correlated. This implies that the surface brightness I and the size R are degenerate. It also supports the fact that surface brightness I and velocity dispersion σ are uncorrelated at redshift $z = 0$ Sheth and Bernardi [2012]. The coefficient of the surface brightness I is also the one that deviates the most from the virial theorem, which predicted a = 2 and b = -1 and is 52% and 69% off. Two out of three of the values of our coefficients, namely b and c, deviate from the ones found by Joachimi et al. [2015] on another sample of ellipticals, a = 0.95, b = -0.73, and c = -7.37, again underlining the importance of fitting a separate fundamental plane to each unique galaxy sample, with variable properties and selection effects.

Since in our study, we are also gathering information about the intrinsic size correlations of spiral galaxies, something that has often been slid to the side in previous research on this matter, we want to find a fundamental plane for spirals. Spiral galaxies do not obey the Faber-Jackson relation (equation 7), therefore, the above approach of deriving a plane from the virial theorem is not valid for spirals. They do however obey the Tully-Fisher relation:

$$L = \mathrm{ZP} \cdot V^C, \tag{8}$$

that writes the luminosity L of a galaxy as a function of circular velocity V to the power of some constant C scaled by a zero-point ZP. With these two quantities alone, one can define a plane in $L - V$-space, but it would not give a prediction for the intrinsic size of a galaxy. Shen et al. [2001] scrutinized a third variable, simulating the formation of spirals in haloes with specific parameters to predict their properties. They found that adding the disk scale length R_d to the equation substantially reduced the scatter of data around the plane, making it a valuable addition. The reason for this is that the gamut of values for R_d is wide, which makes that a certain value provides a considerable amount of information on the galaxy. The proposed plane is as follows:

$$M = \alpha \log(R_d) + \beta \log(V) + \gamma, \tag{9}$$

where a tilt of the plane is allowed by introducing the constants α, β, and γ, in the same way as was done with a, b, and c for the elliptical fundamental plane [Shen et al., 2001]. The substitution of the luminosity L for the magnitude M is a matter of preference, since these can be interchanged using equation 6. Unfortunately, we did not have the disk scale length R_d of the spiral galaxies at our disposal and were obliged to use a different galaxy size estimate. The effective radius, as we used for elliptical galaxies, is not a good choice here since the definition assumes a circular symmetric distribution of mass within the object, a condition that is not met for spirals, with their swirling stretched out arms. Instead, we utilise the virial radius R_v, which was a size estimate based on the average density of the object at that size within which the virial theorem (equation 1) is satisfied:

$$\log(R_v) = \alpha M + \beta \log(V) + \gamma. \tag{10}$$

We changed the order of terms in the above equation to be able to predict sizes and not magnitudes. This equation provides us with a reasonable, evenly scattered fit, as can be seen in the middle-left panel of figure 3. The virial radius R_v on the horizontal axis was plotted against the radius as predicted by equation 3 on the vertical axis. From here, it also becomes visible that the magnitude of the virial radius R_v is different from that of the effective radius R_e in the top left panel of figure 3. Galaxies are plotted as pink dots whose size is scaled by V/σ, the grey line is the one-to-one relation between measured sizes (from the simulation data) and predicted sizes (from the fundamental plane). The distribution of sizes is more uniform than was the case for the elliptical galaxies. This is what we expect given the theory on structure formation in the Universe. In the evolution of spirals, the galaxies are kept relatively compact during gas accretion, and if a spiral experiences so many merges that it has become massive, the spiral structure would have been destroyed so that the remaining galaxy is classified as an elliptical. Of a few galaxies the deviation from the perfect fit is significant, but no obvious outliers can be detected. The middle-right panel in figure 3 shows the amount of correlation between the fitted parameters α = -0.26, β = -0.50, and γ = -4.69. Opposite to the correlations we found for ellipticals, now all parameters correlate to a greater or lesser extent with each other. This points to somewhat degenerate information characterising the fundamental plane. Regardless, we proceed to use this plane in our further analysis, since the scatter around the fundamental plane is fairly symmetric and closely comparable to that

from ellipticals.

By comparing the measured size R with the one predicted by the fundamental planes R_{FP}, we can define a variable that quantifies how much larger or smaller a galaxy is than expected:

$$\lambda = \frac{R}{R_{FP}} - 1. \tag{11}$$

This λ is the first order fundamental plane residual, and can be used for both ellipticals and spirals. If sizes are precisely as predicted by the plane, then the graph of λ would look like the Dirac delta function. That is clearly not the case, as we already have seen scatter in the fit of the planes. The bottom panel of figure 3 shows the density distribution function of λ for ellipticals in blue and for spirals in pink. The grey curve is the probability density function of a proper Gaussian with mean $\mu = 0$ and standard deviation $\sigma = 0.15$. Looking at the plot, we find the λ densities to be approximately Gaussian. This symmetry around 0 is validating our choice of the parameters in the fundamental planes. We also observe that the shapes of the blue and pink curves are similar, so that we can qualitatively compare the values of λ and therefore intrinsic size correlations of the elliptical and spiral galaxies.

This will become important now, when we move on to the actual correlation estimators. These estimators should give us a measure of how much more clustered the galaxies are in our sample compared to randomly distributed ones, and eventually give us an idea of the galaxy bias and size bias in our universe. We will use two-point correlation functions, which means as much as that we are going to measure quantities on pairs out of two input samples. Following the approach of Anderson et al. [2012], Singh et al. [2020] and Joachimi et al. [2015], we start with the Landy-Szalay estimator [Landy and Szalay, 1993]:

$$\xi_{gg}\left(r_p, \Pi\right) = \frac{SD - DR_S - SR_D + R_S R_D}{R_S R_D}, \tag{12}$$

which measures the clustering of points in sample S to that of sample D relative to the clustering of random points R. The terms of this equation are paircounts, as a function of transverse separation r_p and line of sight separation Π. If ξ_{gg} is larger than zero, then S is positively clustered with D, if on the other hand ξ_{gg} is smaller than zero, S is anti-clustered with D. This estimator is applied to various galaxy samples. We can measure a similar quantity on the matter field, by weighting matter particles by their mass. This brings us to the following:

$$\xi_{\delta\delta}\left(r_p, \Pi\right) = \frac{m_S m_D - m_D R_S - m_S R_D + R_S R_D}{R_S R_D}. \tag{13}$$

Here, S and D are equal, namely the subsampled matter catalog from the Horizon-AGN simulation. Taking the ratio of $\xi_{\delta\delta}$ and ξ_{gg} results in the squared galaxy bias function.

These functions thus give information about the clustering of galaxies of type S with type D and their relation to the matter field, but they do not yet tell us anything about the clustering of galaxies of different intrinsic sizes. To examine this, we measure ξ_{gg} on galaxies in S and D for which the fundamental plane residual λ is larger than zero separately from

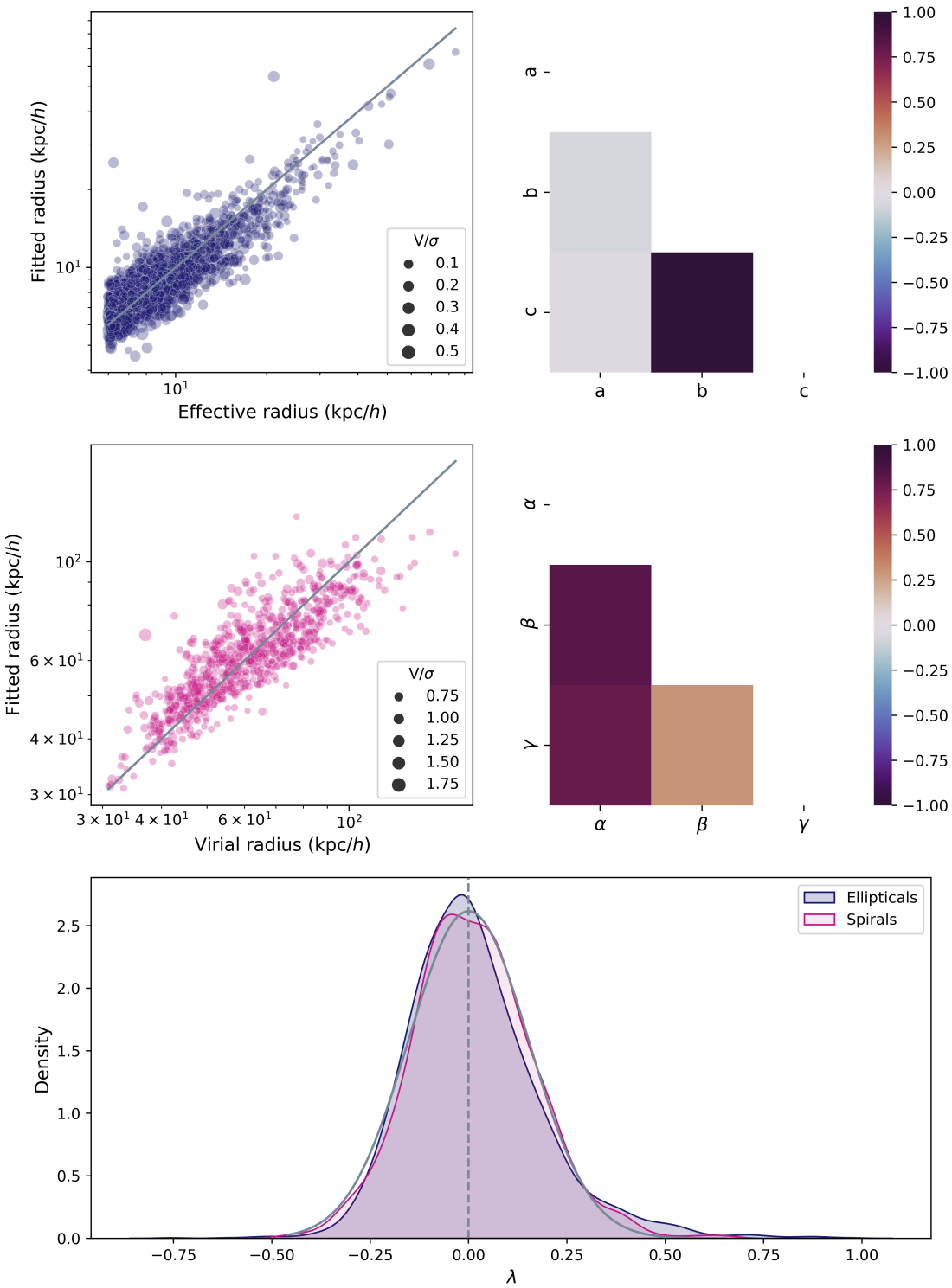

Figure 3: **The fundamental planes of elliptical and spiral galaxies**. The elliptical fundamental plane is shown in the top two panels, with the scatter around the plane on the left plot and the correlation between the parameters on the right plot. The same is done for the spiral fundamental plane in the middle two panels. The bottom panel shows histograms of the distribution of values of the fundamental plane residuals, overlaid with a normal distribution of mean zero and standard deviation 0.15.

the galaxies for which λ is smaller than zero, and take the ratio between them to define a quantity Δ_{gg}:

$$\Delta_{gg} = \frac{\xi_{gg}(\lambda > 0)}{\xi_{gg}(\lambda < 0)} - 1. \tag{14}$$

A Δ_{gg} larger than zero indicates that larger galaxies are clustered together more strongly than smaller galaxies, for a Δ_{gg} smaller than zero this is the other way around. Since we are in possession of an elliptical galaxy sample as well as a spiral one, we can look at other ratio's as well, and match the λ selection in the numerator and denominator:

$$\Delta_{gg}^{(1)} = \frac{\xi_{gg,\text{elliptical}}(\lambda > 0)}{\xi_{gg,\text{spiral}}(\lambda < 0)} - 1; \tag{15}$$

$$\Delta_{gg}^{(2)} = \frac{\xi_{gg,\text{spiral}}(\lambda > 0)}{\xi_{gg,\text{elliptical}}(\lambda < 0)} - 1; \tag{16}$$

$$\Delta_{gg}^{(3)} = \frac{\xi_{gg,\text{elliptical}}(\lambda > 0)}{\xi_{gg,\text{spiral}}(\lambda > 0)} - 1; \tag{17}$$

$$\Delta_{gg}^{(4)} = \frac{\xi_{gg,\text{elliptical}}(\lambda < 0)}{\xi_{gg,\text{spiral}}(\lambda < 0)} - 1. \tag{18}$$

Another way of incorporating sizes into the correlation function is by weighting one or both of the points in a pair by the fundamental plane residual λ, yielding:

$$\xi_{g\lambda}\left(r_p, \Pi\right) = \frac{\lambda_S D - \lambda_S R_D}{R_S R_D}; \tag{19}$$

$$\xi_{\lambda\lambda}\left(r_p, \Pi\right) = \frac{\lambda_S \lambda_D}{R_S R_D}. \tag{20}$$

Using this last equation, we can define the 'size bias' as the ratio of $w_{\lambda\lambda}$ to $w_{\delta\delta}$, analogous to the galaxy bias. In this approach, we assume the intrinsic size field to be a linear transformation of the matter density field [Joachimi et al., 2015].

However, in these functions it could happen that the signal from positive and negative λ's are cancelling each other out in the final estimator. To reveal whether this is the case, we evaluate the functions multiple times on different galaxy subsamples, measuring galaxies with a positive and negative λ separately in the lambda weighted part of the function. We also measure $\xi_{\delta\delta}$ on the galaxies (equation 13), using their fundamental plane residual λ as weight, to investigate the effect of using λ as a correlation function weight even further.

Now that we know what our correlation estimators look like, it is time to discuss their way of operating. Every two-letter term loops over the points in each of the two samples. For every point in the first sample, we loop over every point in the second sample, and compute the transverse separation r_p, after which we add the weight of the pair to that specific separation bin. We use separations between 0.62 and 16.26 Mpc/h, equally distributed over nine bins in logspace. In the calculation of the separation, periodic boundary conditions are

incorporated to connect opposite sides of the simulation box to each other and account for the periodic forces acting throughout the box. The above functions are also dependent on the line of sight separation Π. To incorporate this, we only added a pair if $|\Pi| < 20$ Mpc/h. That way, we do not have to integrate over Π and transform the functions from $\xi(r_p, \Pi)$ to $w(r_p)$ after projection onto the xy-plane only. We arrive at the following correlation functions:

$$w_{\delta\delta}(r_p) = \frac{m_S m_D - m_D R_S - m_S R_D + R_S R_D}{R_S R_D}; \tag{21}$$

$$w_{gg}(r_p) = \frac{SD - DR_S - SR_D + R_S R_D}{R_S R_D}; \tag{22}$$

$$w_{g\lambda}(r_p) = \frac{\lambda_S D - \lambda_S R_D}{R_S R_D}; \tag{23}$$

$$w_{\lambda\lambda}(r_p) = \frac{\lambda_S \lambda_D}{R_S R_D}. \tag{24}$$

We coded these functions up in the Python language. There is one problem with the evaluation of the functions, which is that the computation of the terms, only using brute force code to compute the transverse separations between all points, is quite expensive. The most time-consuming term is RR. The computational time for this increases with the square of the sample size, and even more since we are oversampling the number of randoms. Therefore, it would be interesting to evaluate RR analytically, instead of looping over all points. We estimate RR by comparing the volume of the hollow cylinder enclosed by the annulus between two transverse separations r_p and a line of sight separation Π to the total volume of the simulation box. Self-pairs are excluded subtracting the number of duplicated objects (denoted by δ_{DS}) from the length of second sample in the case of correlating two of the same samples. Our formula is as follows:

$$\hat{RR} = \frac{\pi \cdot ((r_{p,2})^2 - (r_{p,1})^2) \cdot (N_D N_{R_D}) \cdot ((N_S N_{R_S}) - \delta_{DS}) * 2\Pi_{\max}}{100^3}, \tag{25}$$

where N stand for the number of points present in the subscripted sample. To check whether this is a valid substitution, and also to validate the pair-counting routine, we computed RR and equation 25 for all transverse separations used, for representative galaxy sample sizes and compared it with the analytical expectation (figure 4). We observe an excellent match, and decide to substitute equation 25 for RR in all our correlation estimators.

Not unnecessary is an estimation of the error in these measurements, something that we will discuss here. We will use the jackknife estimation method, one that is useful for estimating errors in the absence of a theoretical model behind the measurements [McIntosh, 2016]. Even though theoretical models for the correlation functions do exist (Anderson et al. [2012], Joachimi et al. [2015], Singh et al. [2020]), they are beyond the scope of this work. The method is based on the one-by-one removal and reinsertion of individual observations, while in between measuring the quantity of interest from the remaining observations [McIntosh, 2016]. Multiple Universes would definitely come in handy now if we were to follow this approach

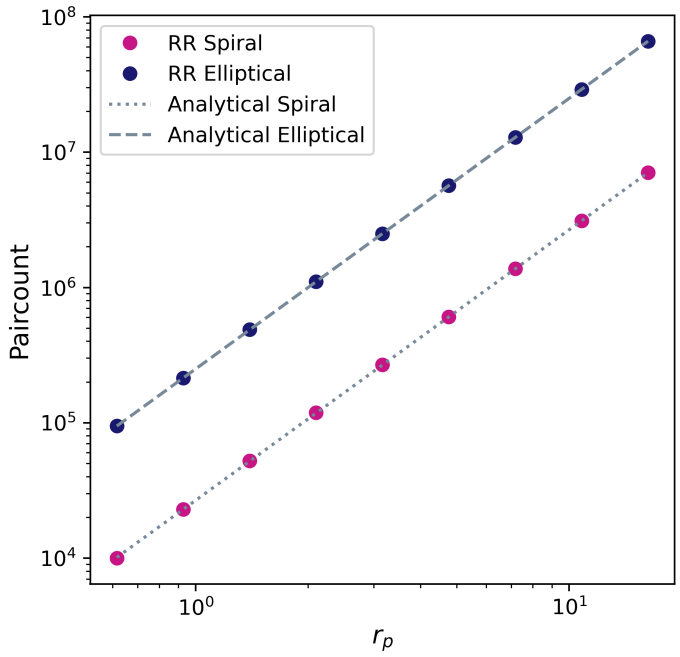

Figure 4: **A comparison between RR and the analytical paircount.** The number of paircounts on the vertical axis is plotted against the transverse separation r_p in Mpc/h on the horizontal axis. The true value of RR is computed for 20 times as many randoms as galaxies, for ellipticals in blue and spirals in pink. The grey dashed and dotted lines represent the analytically estimated counts. The two estimates match very well.

exactly, because only then would we have access to multiple independent observations of our quantities of interest, i.e. the correlation estimators. Since we do not, we look at the simulation box as if it were an ensemble of independent smaller universes, and each of these was an pseudo-independent observation. We divide the simulation box into 27 of these smaller universes, by slicing it in three in all spatial directions x, y and z. That way, each measurement will have 27 jackknife replicates $\hat{\theta}_{(i)}$. Following McIntosh [2016], we compute the empirical average $\hat{\theta}_{(\cdot)}$ of the replicates:

$$\hat{\theta}_{(\cdot)} = \frac{1}{27}\sum_{i=1}^{27}\hat{\theta}_{(i)}. \tag{26}$$

From there, the standard error is constructed as:

$$SE\left(\hat{\theta}\right)_{jack} = \left\{\frac{26}{27}\sum_{i=1}^{27}\left(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)}\right)^2\right\}^{1/2} \tag{27}$$

Of course we want to make sure that we actually measure the sample variance here and not just 'shot noise'. Therefore, we oversample the number of randoms in equations 21-24 by taking 20 times as many randoms as data points in the case of the elliptical and spiral galaxy samples and 5 times as many in the case of the subsampled matter catalogue.

To assess whether our measurements are significantly different from 0, and are thus an actual correlation signal, we collect some supporting statistics. The first one is the reduced chi-squared, χ^2_{red}, or the goodness-of-fit parameter. We construct a full covariance matrix M from the jackknife replicates, with the same dimensions as our data X (which is 9) and elements s_{mn} defined as:

$$s_{mn} = \frac{26}{27}\sum_{i=1}^{27}\left(x_{im} - \bar{x}_m\right)\left(x_{in} - \bar{x}_n\right), \tag{28}$$

where $\bar{x}_m = \frac{1}{27}\sum_{i=1}^{27} x_{im}$ is the mean of the m-th data point over all jackknife replicates. The χ^2 statistic now follows from:

$$\chi^2 = X \cdot M^{-1} \cdot X \cdot HF, \tag{29}$$

where we have corrected for the bias of the inverse covariance matrix, caused by a finite number of jackknife replicates, by multiplying with the Hartlap factor HF, which in our case is equal to $\frac{8}{13}$ [Hartlap et al., 2007]. The reduced chi-squared is obtained by dividing χ^2 by the degrees of freedom, which equals 9, the length of our data vectors. Without consulting statistical tables, the χ^2_{red} can give us a rough estimate of the significance of the measurement: a value much larger than 1 indicates that the data is inconsistent with the model, a value close to one means that the model gives a sufficient description of the data, and a value close to zero points towards the model resembling the data too much, and thus the conclusion that the data is over-fitted, or the variance is over-estimated. That's why in physics people often report significance in units of σ, computed directly from the p-value and the number of degrees of freedom. The p-value is the probability of finding a result as extreme as, or

more extreme than, our observation given the proposed model. In other words, it gives the probability that our observation can be attributed to random fluctuations. In this thesis, we refer to deviations from the null hypothesis at $\geq 2\sigma$ as significant, corresponding to a p-value of 0.0455 [Whitlock and Schluter, 2015].

Summarising, we effectively measure w_{gg}, $w_{g\lambda}$, $w_{\lambda\lambda}$, and $w_{\delta\delta}$ on ellipticals versus ellipticals, spirals versus spirals, and ellipticals versus spirals, and observe the results. We also measure $w_{\delta\delta}$ on the subsampled matter catalogue. Comparing w_{gg} of the galaxies with the matter density field $w_{\delta\delta}$ will give us an estimate of the galaxy bias of ellipticals and of spirals, i.e. the clustering of galaxies relative to the clustering of matter. Looking at $w_{g\lambda}$ and comparing $w_{\lambda\lambda}$ of the galaxies with the matter density field $w_{\delta\delta}$ will provide an estimate of the 'size-bias' of ellipticals and of spirals, i.e. the deviation of the density field that correlates with intrinsic size. Then, we question the validity of weighting the correlation functions by λ, and measure the different quantities after splitting the galaxy samples into two, with λ positive and negative separate from each other.

4 Results

After carrying out everything we mentioned in the Method section, we have collected an enormous set of data on the correlations of galaxies and the matter field. Our findings and their comparison to the results of others will be the focus of this chapter. We begin with discussing the matter density field $w_{\delta\delta}$ and galaxy clustering w_{gg}. Then, Δ_{gg}, $w_{g\lambda}$ and $w_{\lambda\lambda}$ are considered per galaxy sample combination (elliptical-elliptical, spiral-spiral, and elliptical-spiral), using scatter plots and statistics. A few notes on the words and numbers we use to discuss our results are in place here. Whenever I say 'large' or 'small', I do not refer to the absolute size of the galaxy, but rather to the deviation from the fundamental plane, as described by the residual λ. In addition, any numbers reported are rounded off to two decimals.

We first start with the clustering of matter and galaxies as a whole. Figure 5 shows on the vertical axis our measurements of $w_{\delta\delta}$ on the subsampled matter catalog (top left), and w_{gg} on ellipticals only (bottom left), spirals only (bottom right), and a combination of elliptical and spiral galaxies (top right), against the transverse separation r_p on the horizontal axis. We assumed the signal of $w_{\delta\delta}$ to be noiseless. The overall shape of all four of the signals is similar, with the most positive correlation on the smallest separation scales, which then decreases exponentially with larger separations.

For the galaxies, we see positive clusterings on all scales. This means that ellipticals, spirals, and ellipticals and spirals do cluster together in the simulation. The galaxy signals seem to asymptote to zero on the larger scales, while the matter field does not show this behaviour. Galaxy clustering and matter clustering thus work on different scales. This is also what we observe for the galaxy bias, the square root of $w_{gg}/w_{\delta\delta}$, which evolves with transverse separation. For the ellipticals, it is decreasing from 0.74 on the smallest scale to 0.15 on the largest one. This range is a tad smaller for the spirals, where it decreases from 0.70 to 0.27. Since the linear model for the galaxy bias is not a good approximation on the smaller scales, we can only say that the clustering of spiral galaxies dominates the clustering of ellipticals on larger scales (order of 10 Mpc/h). This is opposite to what we would expect, since elliptical galaxies are older and thus followed the underlying density field more in their formation. What is also strange, is that the amplitude of all of the galaxy w_{gg} signals is lower than that of $w_{\delta\delta}$. One explanation for these observations are the selection criteria we applied to the dataset. The prerequisite that the number of stellar particles should be over 3000, could have a different impact on the behaviour of the elliptical sample and the spiral sample.

Next, we discuss other measurements on elliptical galaxies, from which the results are depicted in figure 6 and the statistics are listed in table 1. The top panel of figure 6 plots Δ_{gg} (ratio of large ellipticals over small ellipticals) on the vertical axis against the transverse separation r_p on the horizontal axis. The statistics conclude that this signal is not significantly different from zero with $\sigma = 0.01$. We conclude that the clustering of larger and that of smaller elliptical galaxies are similar. Our results here are in apparent disagreement with earlier findings of Joachimi et al. [2015], who report a significant negative signal for Δ_{gg} up to 20 Mpc/h. Their signal amplitudes however are fairly small (0.1-0.2), so that their measuremets fall within our level of uncertainty, and it may be that we lack the statistical

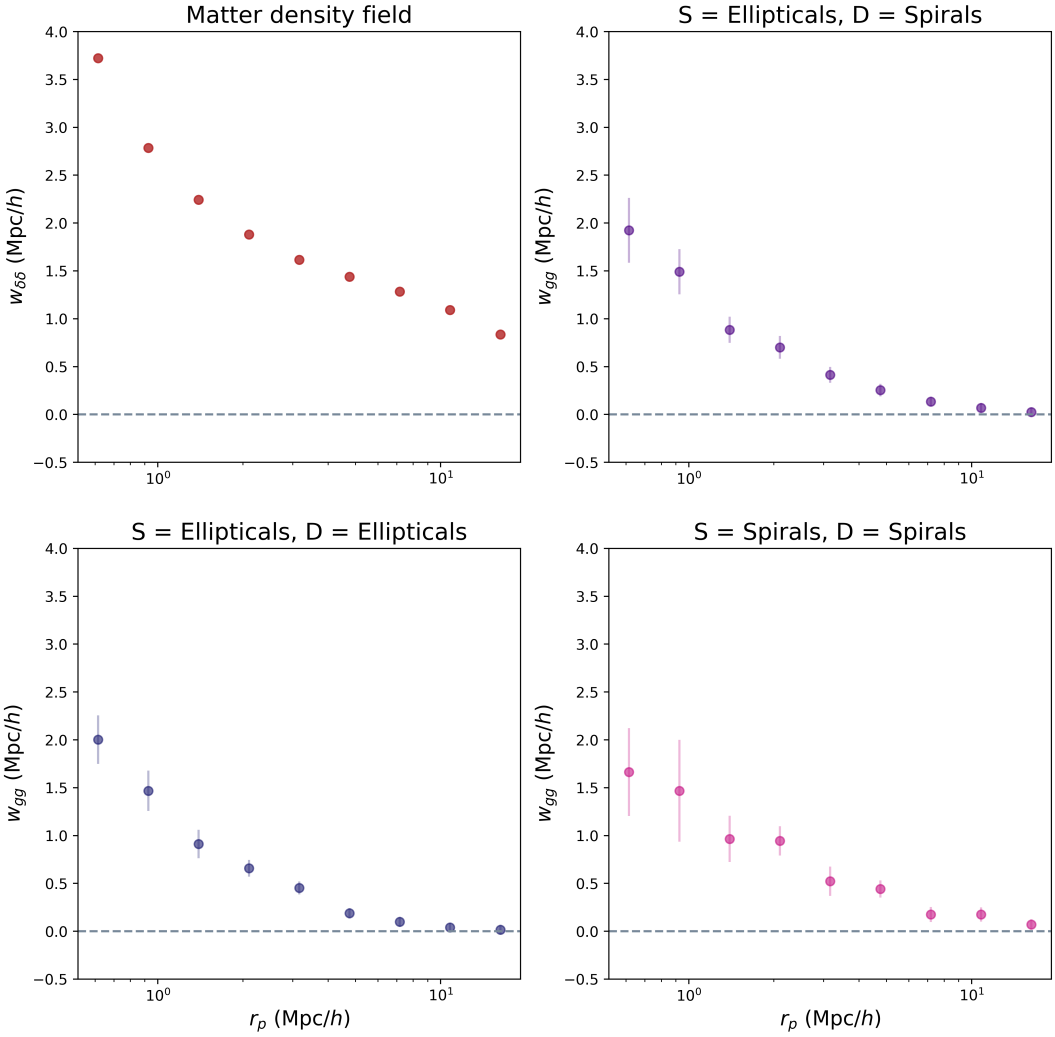

Figure 5: **The clustering of matter and galaxies.** The top left panel shows the matter density field as measured from the subsampled matter catalogue. The other three panels show w_{gg} for different galaxy sample combinations, against the transverse separation r_p.

Table 1: **Elliptical galaxy auto-correlation statistics.** Columns are the measured quantity, the first input sample of the correlation function S, the second input sample of the correlation function D, the reduced chi-squared of comparing the signal to a null-model, and the corresponding σ.

Quantity	S	D	$\chi^2_{red}(0)$	$\sigma(0)$
Δ_{gg}	Ellipticals	Ellipticals	0.24	0.01
$w_{g\lambda}$	Ellipticals	Ellipticals	1.46	1.41
$w_{\lambda\lambda}$	Ellipticals	Ellipticals	0.63	0.28

power to retrieve the signal from the noise here.

The next question would be whether large and small elliptical galaxies cluster separately, or if elliptical galaxies of all sizes cluster together. For this we look at $w_{g\lambda}$, depicted in the middle panels of figure 6. Judging from the statistics, the results are inconclusive, no deviation from zero could be found for the signal $w_{g\lambda}$ ($\sigma = 1.41$). This is again different from Joachimi et al. [2015], and now even the complete oppositve, which we cannot blame onto the size of the errors in our measurements. It could be that there is a discrepancy between simulated data, like we are looking at, and observational data, as used by Joachimi et al. [2015]. This would deserve further examination. What we can say, is that the shape of $w_{g\lambda}$ resembles the one from Singh et al. [2020]. It is curious that the two studies do not report the same result, as they were using the same dataset, namely the Sloan Digital Sky Survey (SDSS). The difference could lie in the fact that Singh et al. [2020] use a sample with low redshift, while Joachimi et al. [2015] incorporated a redshift correction in the fundamental plane.

Moving on to our last measured quantity on ellipticals, $w_{\lambda\lambda}$, we can look at the two bottom panels of figure 6. The plot on the left is very noisy, and does not yield a significant result. From our measurements, we find a 'size bias' of elliptical galaxies between 0.05 and 0.14. Joachimi et al. [2015] found a smaller bias here, 0.038.

Let us turn to the spiral galaxies, for which we created the same plots in figure 7 as for the elliptical galaxies. The statistics are listed in table 2. Since we have the honour of being the first to analyse the intrinsic size correlations of spiral galaxies, we are unable to compare our findings to other literature. What we can do, is make a comparison to similar statistics we measured on elliptical galaxies. The top panel of figure 7 displays the behaviour of Δ_{gg}, which is not significantly different from zero ($\sigma = 0.58$). It looks like the clustering of large spiral galaxies overpowers the clustering of small ones on medium scales of separation (1-5 Mpc/h), yet the signal is not significant.

This same behaviour can be observed for $w_{g\lambda}$ in the center panels of figure 7, but here it is significant ($\sigma = 3.38$). We deduce that, mainly at medium scales, spiral sizes correlate with spiral positions. We observed something similar in $w_{g\lambda}$ for ellipticals, but here the signal was not significant. This could be because the datapoints in the elliptical signal are more

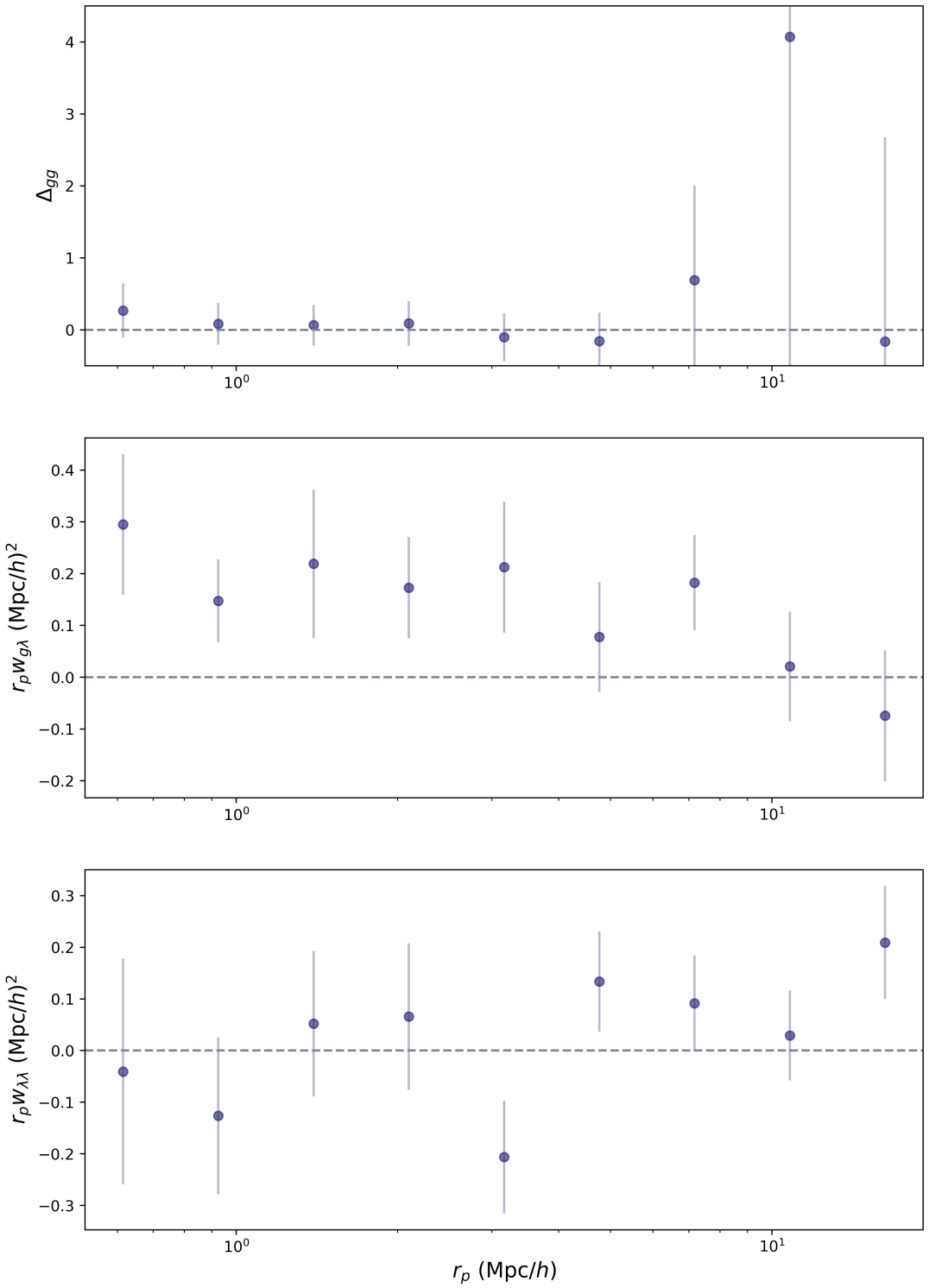

Figure 6: **Correlation estimators for elliptical galaxies.** Here, Δ_{gg} is shown in the top panel, $w_{g\lambda}$ in the two center panels, and $w_{\lambda\lambda}$ in the bottom panel. On the horizontal axis, the transverse separation r_p is set out. Mind that the latter two signals are scaled by r_p on the vertical axis.

Table 2: **Spiral galaxy auto-correlation statistics.** Columns are the measured quantity, the first input sample of the correlation function S, the second input sample of the correlation function D, the reduced chi-squared of comparing the signal to a null-model, and the corresponding σ.

Quantity	S	D	$\chi^2_{red}(0)$	$\sigma(0)$
Δ_{gg}	Spirals	Spirals	0.86	0.58
$w_{g\lambda}$	Spirals	Spirals	3.19	3.38
$w_{\lambda\lambda}$	Spirals	Spirals	0.60	0.26

correlated.

Even though the amplitude of the auto-correlation of λ for spirals is larger than for ellipticals, the signal is very noisy and not significantly different from zero ($\sigma = 0.26$), as was the case for ellipticals. From our measurements, we find a 'size bias' of spiral galaxies between 0.14 and 0.39, which is higher than what we found for ellipticals.

Lastly, we discuss our measurements on elliptical galaxies versus spiral galaxies, depicted in figure 8. In the top panel, we see Δ_{gg} for different λ cuts on the elliptical and spiral galaxies. The statistics tell us that only $\Delta^{(3)}_{gg}$ is significantly different from zero. The datapoints are positioned under the null-line, meaning that large spirals are clustered to a higher degree than large ellipticals.

Moving on to $w_{g\lambda}$, the center two panels in figure 8, we see a large overlap in shape of the correlation function between the galaxy density of spirals and the sizes of ellipticals (left panel), and the galaxy density of ellipticals and the sizes of spirals (right panel). Both of the signals are not significant ($\sigma = 1.49$ and $\sigma = 1.28$). In figure 5 we saw a positive correlation between the positions of ellipticals and the positions of spirals, especially at smaller scales of separation. We deduce that the sizes of the galaxies are to lesser extend subject to clustering than the positions are.

Last up are the measurements on $w_{\lambda\lambda}$ for ellipticals versus spirals. It is difficult to see from the figure, but the signal is highly significantly different from zero ($\sigma = 3.24$), because the error in the measurement is very small. We can make a very important conclusion here, namely that large ellipticals cluster with large spirals, and small ellipticals with small spirals.

The faintness of many of the signals lead to the suspicion that the weights of the positive and negative values were cancelling each other out in the total signal of the measured quantities. This could be the case if the larger and smaller galaxies behaved in a similar manner, since the distribution of λ was symmetric around zero. As a first step to investigate this, we measure $w_{\delta\delta}$ (equation 13) for galaxies, weighting the paircounts by their fundamental plane residual λ instead of by their mass m. The results are depicted in figure 9, where different parts of $w_{\delta\delta}$ on the vertical axis are plotted against the transverse separation r_p on

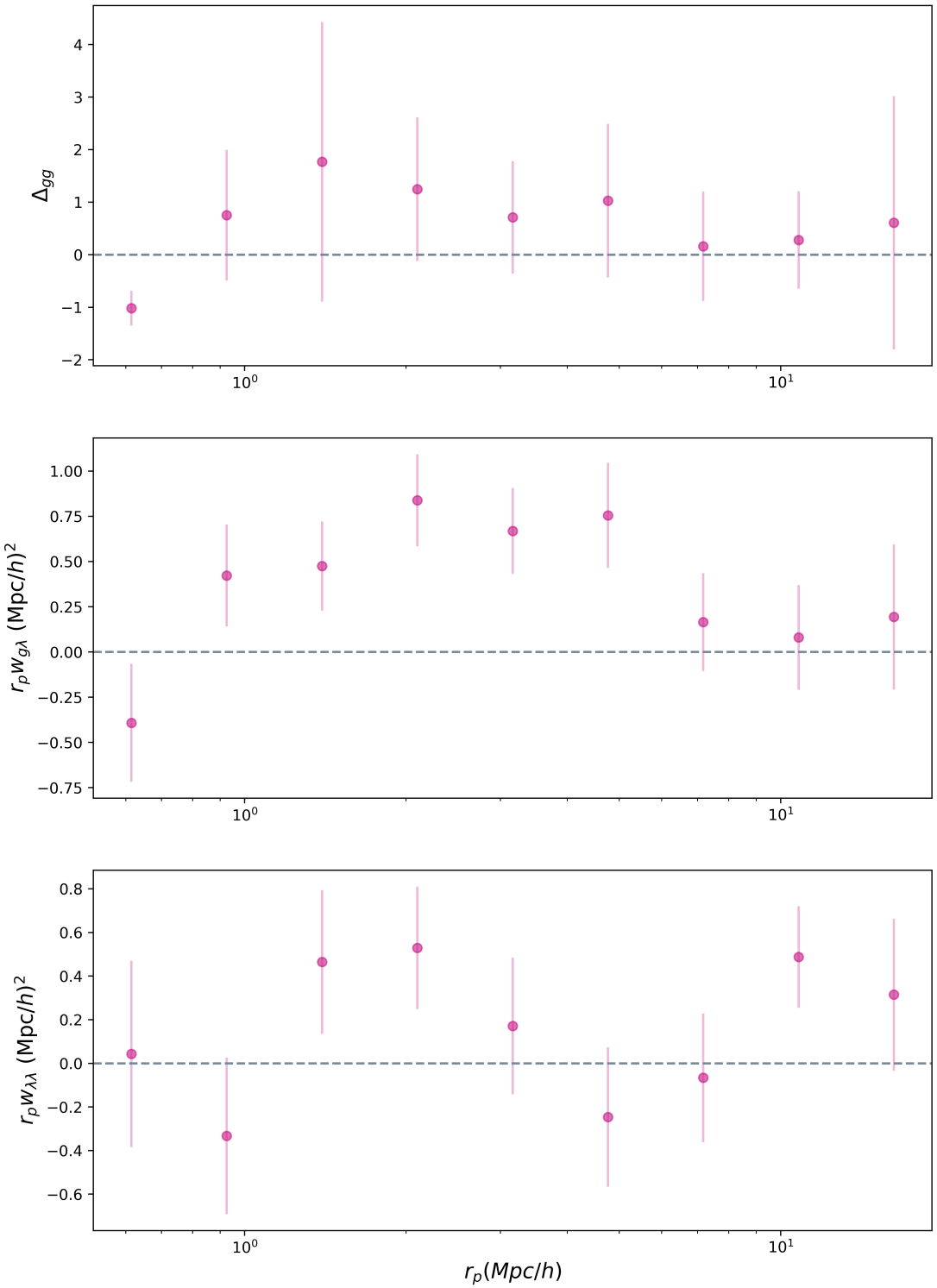

Figure 7: **Correlation estimators for spiral galaxies.** Here, Δ_{gg} is shown in the top panel, $w_{g\lambda}$ in the two center panels, and $w_{\lambda\lambda}$ in the bottom panel. On the horizontal axis, the transverse separation r_p is set out. Mind that the latter two signals are scaled by r_p on the vertical axis.

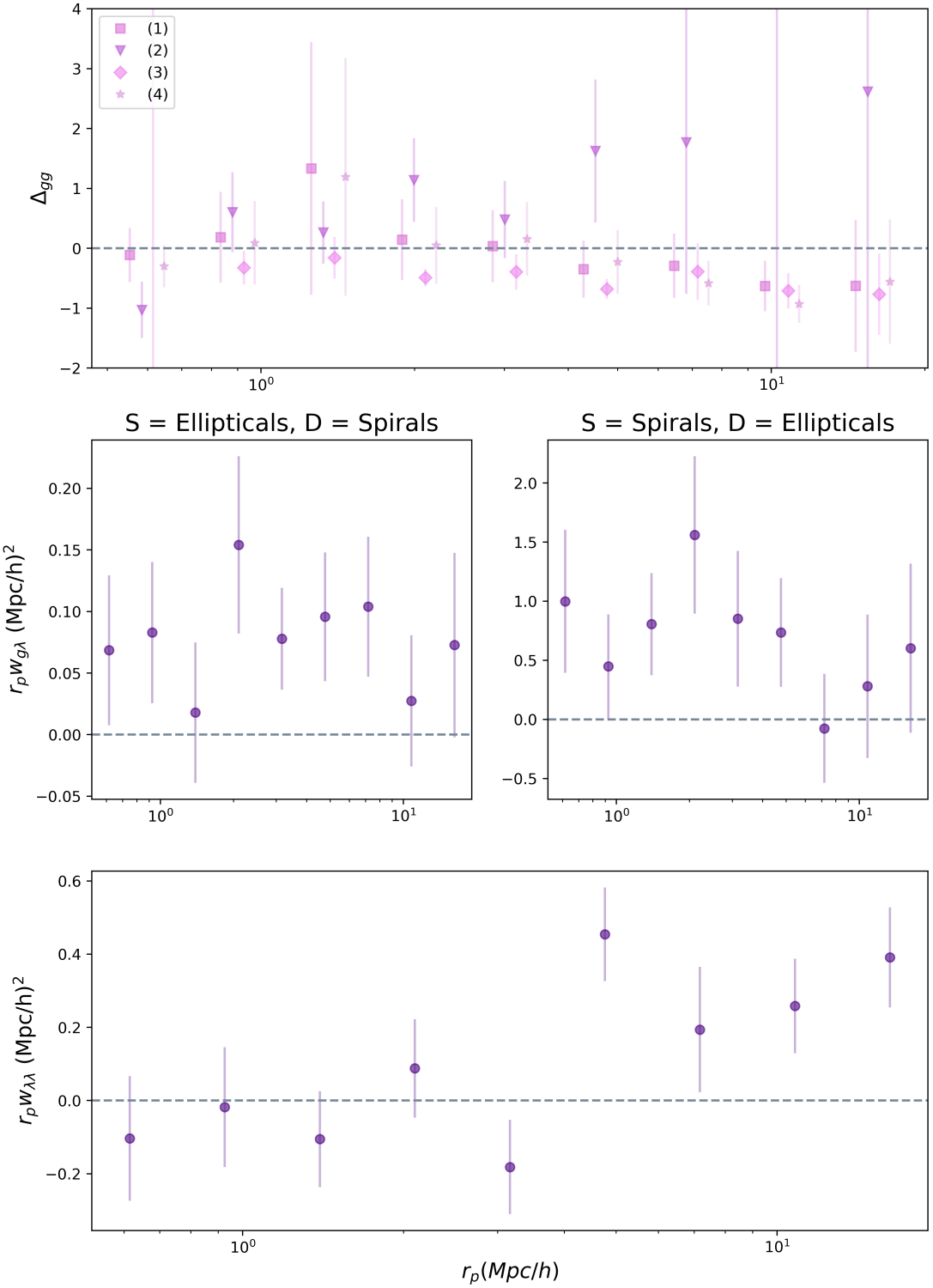

Figure 8: **Correlation estimators for elliptical versus spiral galaxies.** Here, four versions of Δ_{gg} are shown in the top panel, $w_{g\lambda}$ in the two center panels, and $w_{\lambda\lambda}$ in the bottom two panels. On the horizontal axis, the transverse separation r_p is set out. Mind that the latter two signals are scaled by r_p on the vertical axis.

Table 3: **Elliptical/spiral galaxy cross-correlation statistics.** Columns are the measured quantity, the first input sample of the correlation function S, the second input sample of the correlation function D, the reduced chi-squared of comparing the signal to a null-model, and the corresponding σ.

Quantity	S	D	$\chi^2_{red}(0)$	$\sigma(0)$
$\Delta^{(1)}_{gg}$	Ellipticals	Spirals	0.41	0.09
$\Delta^{(2)}_{gg}$	Ellipticals	Spirals	0.57	0.22
$\Delta^{(3)}_{gg}$	Ellipticals	Spirals	5.04	4.94
$\Delta^{(4)}_{gg}$	Ellipticals	Spirals	1.38	1.30
$w_{g\lambda}$	Ellipticals	Spirals	1.51	1.49
$w_{g\lambda}$	Spirals	Ellipticals	1.36	1.28
$w_{\lambda\lambda}$	Ellipticals	Spirals	3.05	3.24

the horizontal axis. The top panel displays $w_{\delta\delta}$ for elliptical galaxies versus spiral galaxies, the bottom left panel shows this quantity for ellipticals only, and the bottom right panel shows it for spirals only. For all galaxy sample combinations, we observe that the resulting value $w_{\delta\delta}$ is the sum of parts that lie far apart in amplitude. This gives reason to believe that size-weighting suppresses the clustering amplitude. What jumps out most is the clustering of small galaxies, $w_{\delta-\delta-}$, in all panels. We also observe that the signals are quite flat.

In addition, we looked at the different parts of $w_{g\lambda}$ ($w_{g\lambda+}$ and $w_{g\lambda-}$) and $w_{\lambda\lambda}$ ($w_{\lambda+\lambda+}$, $w_{\lambda+\lambda-}$, and $w_{\lambda-\lambda-}$) in the clustering of galaxies. The results for ellipticals are shown in figure 10, the other galaxy sample combinations exhibit similar behaviour. For $w_{g\lambda}$, we see that both parts lie above zero, so here, the clustering is not suppressed. This is not the case for $w_{\lambda\lambda}$ however, where we observe that the different parts of the signal are each others opposites. Size-weighting the correlation function thus suppresses the signal. We see that the different signs of λ make a non-trivial contribution to the value of the estimators. More work is needed to understand what the different parts of $w_{g\lambda}$ and $w_{\lambda\lambda}$ are telling us about the clustering of galaxies and intrinsic size correlations. Our conclusions are part of an ongoing investigation into the best estimators to use for this task.

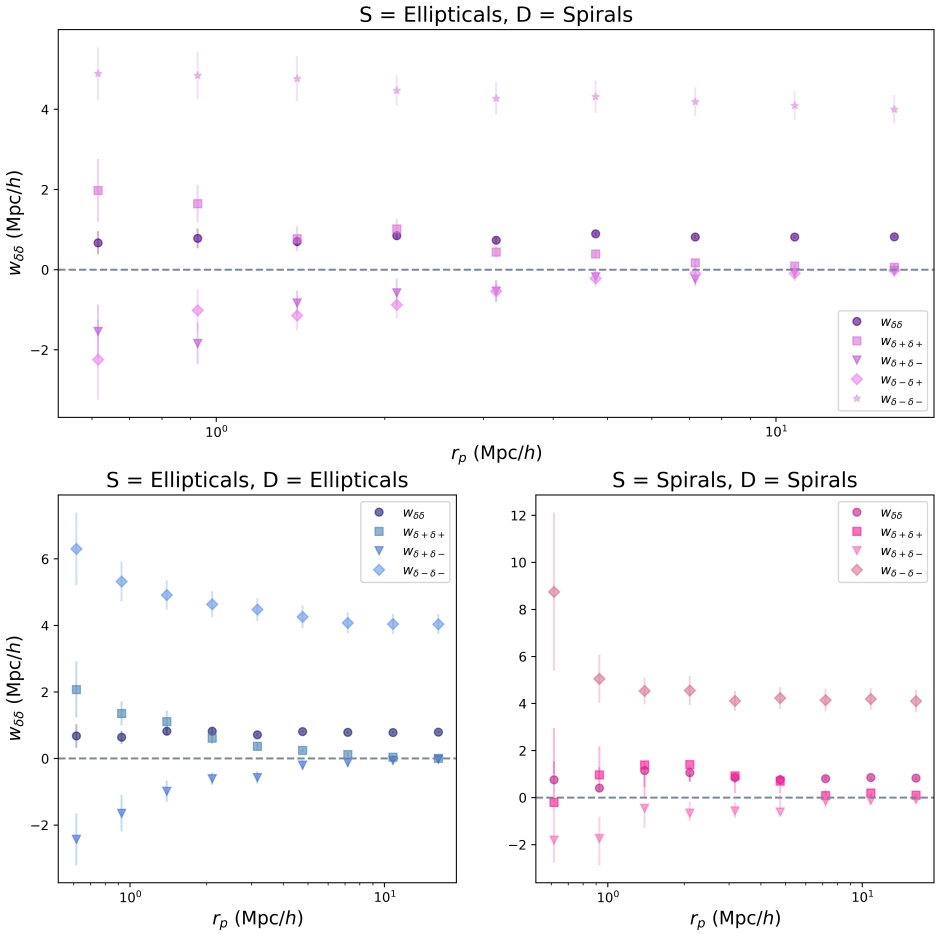

Figure 9: **λ-weighted clustering of galaxies.** The quantity $w_{\delta\delta}$ for elliptical versus spiral galaxies is shown in the top panel, for pure ellipticals in the bottom left panel, and for pure spirals in the bottom right panel. On the horizontal axis, the transverse separation r_p is set out.

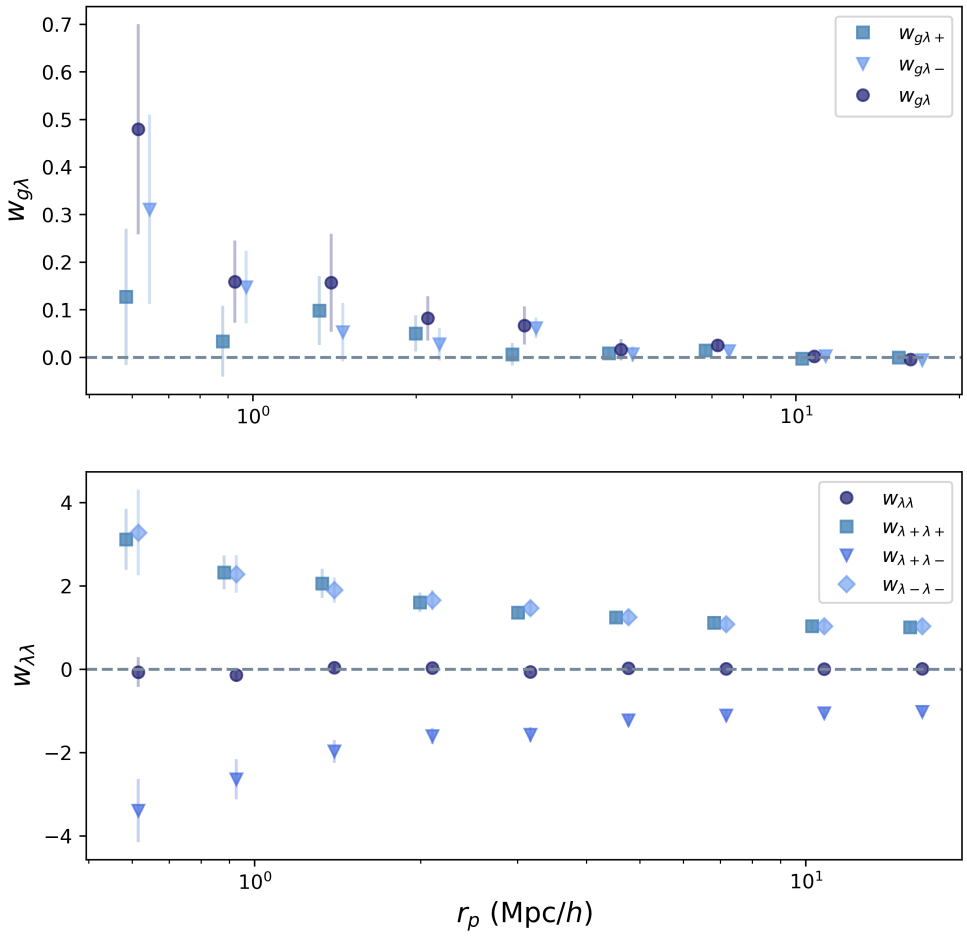

Figure 10: **Cuts on weighted cross-correlation between galaxy positions and λ, and auto-correlations of sizes.** Shown are the results for elliptical galaxies, the other samples look similar. On the horizontal axis, the transverse separation r_p is set out.

5 Conclusions

Summarising our results, we have showed that intrinsic size correlations of both elliptical and spiral galaxies are present and significant to varying degrees in the Horizon-AGN simulation. More specifically, spiral sizes correlate with spiral positions, and large ellipticals cluster with large spirals, while small ellipticals cluster with small spirals.

Needless to say, there is still much that can be improved on my research. The selection criteria we applied on our data greatly decreased the available number of galaxies. From 126361 galaxies, we went to 2432 ellipticals and 795 spirals. The majority of these galaxies were removed because they had less than 3000 stellar particles. This is a harsh cut, but neccessary for a reasonable fit to the fundamental plane for elliptical galaxies. As such, it could have had consequences for our results, like the unexpected small amplitude of w_{gg} in the clustering of galaxies. In follow-up research, one could look into relaxing this cut by using a different fundamental plane for the spirals. One option is to use a stellar mass based plane. This could be better for another reason as well, since the luminosity based fundamental plane depends on the waveband, so that measurements on different bands are harder to compare [Hyde and Bernardi, 2008]. There is no persuasive reason to hold on to the fundamental plane as derived by the virial theorem. The variables of our fundamental planes were to more or lesser extent degenerate. We already had to fit the parameters a, b, and c to correct for the deviation of galaxies from the virial theorem, not in the least because we assumed the mass to luminosity ratio to be constant. This assumption is not valid, since this ratio depends on the star formation history of the galaxy [Shen et al., 2001]. It would be ideal to have an estimate of the galaxy size that appears in both a fundamental plane for ellipticals and a plane for spirals. Every estimate knows its own errors and has its own bias, so using different estimates for the size and comparing the results afterwards could result in discrepancies.

It would be interesting to look at clustering of galaxies for other redshifts as well. The fundamental plane could again pose a problem here, since some planes can become biased at high redshift [Sheth and Bernardi, 2012], but a solution to correct for this has already been posed by Joachimi et al. [2015], by adding a term to the fundamental plane that scales with redshift. Measurements on other redshifts could also require the implementation of a brightness weighting V_{max} to make them more realistic. This increases the weight of fainter galaxies, and with that corrects for the fact that you miss those at high redshift [Ciarlariello et al., 2014].

Another thing that deserves attention, is the V/σ criterium for the classification of elliptical and spiral galaxies. We chose a threshold of $V/\sigma = 0.6$, but this is quite arbirtrary really. For example, Dubois et al. [2016] used everything below $V/\sigma = 1$ as an elliptical. Moving up the cut will provide a purer sample of spirals, however this was not feasible for us to do now, since it would leave us with too few spirals to yield any result. Another, perhaps wild, option is to make the criterium less rigid, and assign a galaxy a class according to which fundamental plane it fits best, but this would require more careful consideration.

Furthermore, we have shown that $w_{\lambda\lambda}$ is a non-optimal estimator for measuring auto correla-

tions of λ, since positive and negative parts cancel each other out. This means that weighting the paircounts by a size that has a symmetric distribution around zero suppresses the clustering amplitude. Measuring partial signals of $w_{\lambda\lambda}$ have however showed that these auto correlations are interesting, so a different type of estimator is definitely warranted. Another view on this would be to bin the fundamental plane residuals λ, instead of classifying them as large or small.

Something else that stood out is the deviant behaviour of the datapoints of the smallest scales in some cases. One could look into what is causing this, if for example is due to non-linear effects, and what would be observed on even smaller scales.

In addition, the discrepancy we found between results obtained from observational data and from simulated data deserves to be explored in depth.

Improvements can also be made on the modelling of the galaxy bias and size bias function. Joachimi et al. [2015] have tried this, and assumed a linear relation between the density field of matter and the amplitude of the correlations. This was only valid for larger separations, but these too could come into its own in a more complicated model such as the halo model, proposed by Ciarlariello et al. [2014], and Ciarlariello and Crittenden [2016].

Looking at our results at a whole other level, we used only the 2-dimensional projected positions of the galaxies in the correlation functions, while the third dimension was used only to define a maximum line of sight separation. Now, 2-dimensional data is more realistic, because it represents the images we find in observational data, where we too can only see projections of galaxies in the sky. In contrast, 3-dimensional data is more faithful to the truth, and could reveal properties of galaxies that we can use in modelling and forecasting. In addition, a larger simulation dataset is already available from the latest Horizon-AGN cuboid box, in which one of the dimensions of the simulation box is of length 1000 Mpc/h instead of 100 Mpc/h. This is also certainly worth investigating, as well as the matter power spectrum, which is the fourier transform of the correlation function.

With my thesis, I hope to inspire other researchers to look into the correlations of intrinsic sizes and galaxy densities, and especially the clustering of spirals. Spirals dominate weak lensing data, while not much has been done on intrinsic size correlations of them. This could give new perspective on the contamination of weak lensing data, and in that way enrich our knowledge of the Universe.

6 Thanking word

This project was the largest science project I ever did. I learned so much about cosmology, and on top of that developed skills in critical thinking, problem solving, coding, handling the command line and the English language. It overall increased my interest in physics. I also learned what it's like to be part of a research group.

Even though I enjoyed working on the project a lot, I could not have done it all by myself. My thanks go to the people who set up and maintain the Gemini cluster. The system of sending a script to another computer worked great for me, and running multiple scripts in parallel was the only way I could have gotten the matter density field measurements ready in time. I want to express my gratitude to my first supervisor Harry Johnston, with whom I spoke every week about anything that was bothering me about the project or that I wanted to double-check. He had absolute confidence in me, and provided me with all materials I needed to proceed and made sure I stayed on the right track. I also want to extend thanks to my second supervisor, Elisa Chisari, who came up with interesting additions for my thesis, and assigned me this project in the first place. The other members of the cosmology research group helped me as well, especially by mocking a presentation session for me to practice. And lastly, I am grateful for my mom and little sister at home, with their inexhaustible faith in me. I really appreciate them, for taking over some of my tasks when I was busy, and listening to my thoughts, even without exactly knowing what they were about. Of course I can't leave out my pets, for giving me a laugh, and for calmly sleeping in the room with me while I was working.

References

Lauren Anderson, Eric Aubourg, Stephen Bailey, Dmitry Bizyaev, Michael Blanton, Adam S. Bolton, J. Brinkmann, Joel R. Brownstein, Angela Burden, Antonio J. Cuesta, Luiz N. A. da Costa, Kyle S. Dawson, Roland de Putter, Daniel J. Eisenstein, James E. Gunn, Hong Guo, Jean-Christophe Hamilton, Paul Harding, Shirley Ho, Klaus Honscheid, Eyal Kazin, D. Kirkby, Jean-Paul Kneib, Antione Labatie, Craig Loomis, Robert H. Lupton, Elena Malanushenko, Viktor Malanushenko, Rachel Mandelbaum, Marc Manera, Claudia Maraston, Cameron K. McBride, Kushal T. Mehta, Olga Mena, Francesco Montesano, Demetri Muna, Robert C. Nichol, Sebastian E. Nuza, Matthew D. Olmstead, Daniel Oravetz, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Kaike Pan, John Parejko, Isabelle Paris, Will J. Percival, Patrick Petitjean, Francisco Prada, Beth Reid, Natalie A. Roe, Ashley J. Ross, Nicholas P. Ross, Lado Samushia, Ariel G. Sanchez, David J. Schlegel Donald P. Schneider, Claudia G. Scoccola, Hee-Jong Seo, Erin S. Sheldon, Audrey Simmons, Ramin A. Skibba, Michael A. Strauss, Molly E. C. Swanson, Daniel Thomas, Jeremy L. Tinker, Rita Tojeiro, Mariana Vargas Magana, Licia Verde, Christian Wagner, David A. Wake, Benjamin A. Weaver, David H. Weinberg, Martin White, Xiaoying Xu, Christophe Yeche, Idit Zehavi, and Gong-Bo Zhao. The clustering of galaxies in the sdss-iii baryon oscillation spectroscopic survey: Baryon acoustic oscillations in the data release 9 spectroscopic galaxy sample. 3 2012. doi: 10.1111/j.1365-2966.2012.22066.x. URL `http://arxiv.org/abs/1203.6594http://dx.doi.org/10.1111/j.1365-2966.2012.22066.x`.

Sandro Ciarlariello and Robert Crittenden. Modelling the impact of intrinsic size and luminosity correlations on magnification estimation. 7 2016. doi: 10.1093/mnras/stw2052. URL `http://arxiv.org/abs/1607.08784http://dx.doi.org/10.1093/mnras/stw2052`.

Sandro Ciarlariello, Robert Crittenden, and Francesco Pace. Intrinsic size correlations in weak lensing. 12 2014. doi: 10.1093/mnras/stv447. URL `http://arxiv.org/abs/1412.4606http://dx.doi.org/10.1093/mnras/stv447`.

Yohan Dubois, Julien Devriendt, Adrianne Slyz, and Romain Teyssier. Self-regulated growth of supermassive black holes by a dual jet-heating active galactic nucleus feedback mechanism: Methods, tests and implications for cosmological simulations. *Monthly Notices of the Royal Astronomical Society*, 420:2662–2683, 3 2012. ISSN 00358711. doi: 10.1111/j.1365-2966.2011.20236.x.

Yohan Dubois, Christophe Pichon, Charlotte Welker, Damien Le Borgne, Julien Devriendt, Clotilde Laigle, Sandrine Codis, Dmitry Pogosyan, Stéphane Arnouts, Karim Benabed, Emmanuel Bertin, Jeremy Blaizot, François Bouchet, Jean-François Cardoso, Stéphane Colombi, Valérie de Lapparent, Vincent Desjacques, Raphaël Gavazzi, Susan Kassin, Taysun Kimm, Henry McCracken, Bruno Milliard, Sébastien Peirani, Simon Prunet, Stéphane Rouberol, Joseph Silk, Adrianne Slyz, Thierry Sousbie, Romain Teyssier, Laurence Tresse, Marie Treyer, Didier Vibert, and Marta Volonteri. Dancing in the dark: galactic properties trace spin swings along the cosmic web. 2 2014. doi: 10.1093/mnras/stu1227. URL `http://arxiv.org/abs/1402.1165http://dx.doi.org/10.1093/mnras/stu1227`.

Yohan Dubois, Sebastien Peirani, Christophe Pichon, Julien Devriendt, Raphael Gavazzi, Charlotte Welker, and Marta Volonteri. The horizon-agn simulation: morphological diversity of galaxies promoted by agn feedback. 6 2016. doi: 10.1093/mnras/stw2265. URL http://arxiv.org/abs/1606.03086http://dx.doi.org/10.1093/mnras/stw2265.

Mauro D'Onofrio, Stefano Cariddi, Cesare Chiosi, Emanuela Chiosi, and Paola Marziani. On the origin of the fundamental plane and faber–jackson relations: Implications for the star formation problem. *The Astrophysical Journal*, 838:163, 4 2017. ISSN 0004-637X. doi: 10.3847/1538-4357/aa6540.

J. Hartlap, P. Simon, and P. Schneider. Why your model parameter confidences might be too optimistic. unbiased estimation of the inverse covariance matrix. *Astronomy and Astrophysics*, 464:399–404, 3 2007. ISSN 00046361. doi: 10.1051/0004-6361:20066170.

J. B. Hyde and M. Bernardi. The luminosity and stellar mass fundamental plane of early-type galaxies. 10 2008. doi: 10.1111/j.1365-2966.2009.14783.x. URL http://arxiv.org/abs/0810.4924http://dx.doi.org/10.1111/j.1365-2966.2009.14783.x.

Benjamin Joachimi, Sukhdeep Singh, and Rachel Mandelbaum. Detection of spatial correlations of fundamental plane residuals, and cosmological implications. 4 2015. doi: 10.1093/mnras/stv1962. URL http://arxiv.org/abs/1504.02662http://dx.doi.org/10.1093/mnras/stv1962.

Harry Johnston. Kids+gama: Intrinsic alignment model constraints for current and future weak lensing cosmology, 4 2020. ISSN 14320746.

Henry C King. *The History of the Telescope*. Courier Corporation, 2003.

J Knapp. Astronomical terms and constants, 2011.

S.D. Landy and S. Szalay. Bias and variance of angular correlation functions. 1993.

Avery McIntosh. The jackknife estimation method. 6 2016. URL http://arxiv.org/abs/1606.00497.

Barbara Ryden. *Introduction to Cosmology*. Cambridge University Press, second edition edition, 2017.

Shiyin Shen, H J Mo, and Chenggang Shu. The fundamental plane of spiral galaxies: Theoretical expectations. *Mon. Not. R. Astron. Soc*, 000:0–000, 2001.

Ravi K. Sheth and Mariangela Bernardi. Plain fundamentals of fundamental planes: Analytics and algorithms. 2 2012. doi: 10.1111/j.1365-2966.2011.19757.x. URL http://arxiv.org/abs/1202.3438http://dx.doi.org/10.1111/j.1365-2966.2011.19757.x.

Sukhdeep Singh, Byeonghee Yu, and Uroš Seljak. Fundamental plane of boss galaxies: Correlations with galaxy properties, density field and impact on rsd measurements. 1 2020. doi: 10.1093/mnras/staa3263. URL http://arxiv.org/abs/2001.07700http://dx.doi.org/10.1093/mnras/staa3263.

Michael C. Whitlock and Dolph Schluter. *The Analysis of Biological Data*. MacMillan Education, second edition edition, 2015.