3

Disentangling galaxy environment and host halo mass

Abstract

The properties of observed galaxies and dark matter haloes in simulations depend on their environment. The term "environment" has, however, been used to describe a wide variety of measures that may or may not correlate with each other. Useful measures of environment include, for example, the distance to the N^{th} nearest neighbour, the number density of objects within some distance, or, for the case of galaxies, the mass of the host dark matter halo. Here we use results from the Millennium simulation and a semi-analytic model for galaxy formation to quantify the relation between different measures of environment and halo mass. We show that most of the environmental parameters used in the observational literature are in effect measures of halo mass. The strongest correlation between environmental density and halo mass arises when the number of objects is counted out to a distance of 1.5 – 2 times the virial radius of the host halo and when the galaxies/haloes are required to be relatively bright/massive. For observational studies this virial radius is not easily determined, but the number of neighbours out to 1 – 2 h^{-1}Mpc gives a similarly strong correlation with halo mass. For the distance to the N^{th} nearest neighbour the (anti-)correlation with halo mass is nearly as strong provided $N \geq 2$. We demonstrate that this environmental parameter can be made insensitive to halo mass if it is constructed from dimensionless quantities. This can be achieved by scaling both the minimum luminosity/mass of neighbours as well as the distance to the nearest galaxy/halo to the properties of the object that the environment is determined for. We show how such a halo mass independent environmental parameter can be defined for both observational and numerical studies. The results presented here will help future studies to disentangle the effects of halo mass and external environment on the properties of galaxies and dark matter haloes.

3.1 Introduction

The formation and evolution of galaxies depends on both internal and external processes ('nature vs. nurture'). Among the internal processes are radiative cooling and the formation of a multi-phase medium, formation and feedback from stars and accretion of gas onto and feedback from super-massive black holes. It is generally assumed that halo mass is the fundamental parameter that drives the internal processes for isolated galaxies. External processes are important because galaxies do not live alone in the Universe. Galaxy interaction can induce gravitational torques that can significantly alter the angular momentum structure of the matter in galaxies. This can for example lead to a starburst or to more rapid accretion onto the central black hole, which may trigger a quasar phase. Smaller galaxies may accrete onto the halo of a more massive galaxy. As a galaxy moves through the gaseous halo of a more massive galaxy it may lose gas due to ram pressure forces. Winds and radiation from nearby neighbours may also affect the evolution of a galaxy. To what extent the properties of galaxies are determined by internal and external processes is still an open question.

Even if halo mass were the only driver of galaxy evolution, galaxy properties would still be correlated with environment. Because peaks in the initial Gaussian density field cluster together, more massive galaxies will live close to each other ('galaxy bias'). A correlation between surrounding galaxy density and internal galaxy properties therefore does not necessarily imply a causal relation between the two.

Early, analytic models predicted that the clustering of haloes depends only on their mass (Kaiser, 1984; Cole & Kaiser, 1989; Mo & White, 1996), while later papers have shown that clustering also depends on properties like formation time (Gao et al., 2005), concentration, substructure content, spin and shape, even for fixed mass (e.g. Harker et al., 2006; Wechsler et al., 2006; Bett et al., 2007; Gao & White, 2007; Jing et al., 2007; Macciò et al., 2007; Wetzel et al., 2007; Angulo et al., 2008; Faltenbacher & White, 2010). All dependencies other than the one with halo mass are, however, second-order effects. Lemson & Kauffmann (1999) already showed that the only property of a dark matter halo that correlates with the (projected) number density of surrounding galaxies is its host halo mass. Other properties like spin parameter, formation time and concentration do *not* depend on the surrounding dark matter density. The formation time and the halo merger rate are found to depend on environment (Gottlöber et al., 2001; Sheth & Tormen, 2004; Fakhouri & Ma, 2009; Hahn et al., 2009).

Both observations and simulations have difficulty disentangling halo mass from the external environment. The two are correlated (higher mass haloes live, on average, in denser environments) and finding an environmental parameter that does not correlate with halo mass is non-trivial. Of course, the mass of the dark matter halo hosting a galaxy is important for the evolution of that galaxy, so halo mass is as good an environmental parameter as any other. One would, however, like to be able to distinguish halo mass (the "internal environment") from the environment on large scales (the "external environment"). It is not a priori clear whether the environmental parameters used in literature measure halo mass, and if so, whether they measure *only* halo mass, or whether they are also, or predominantly, sensitive to the external environment.

Observationally, halo mass is hard to determine. Group catalogues, abundance (or stellar mass - halo mass) matching, and weak gravitational lensing all provide statistical measures of halo mass. Strong gravitational lensing is another way of measuring the total mass of a massive lens system. Nonetheless, most observational data sets will have to do without dark matter halo mass and define environmental parameters based on the distribution of visible matter (usually stellar luminosity) only.

Many observational studies have, nevertheless, investigated the effect of the environment on the physical properties of galaxies. In general, galaxies form their stars earlier and faster in higher density environments (e.g. Lewis et al., 2002; Baldry et al., 2004; Balogh et al., 2004a,b; Kauffmann et al., 2004; Thomas et al., 2005; Smith et al., 2006) and there galaxy morphologies become more (pressure support dominated) early type, as opposed to (rotation dominated) late type (e.g. Dressler, 1980; Dressler et al., 1997; Wilman et al., 2009). From observations alone it is very hard to judge whether these trends are driven mostly by halo mass or whether other halo properties and/or large-scale environment play an important role. Crain et al. (2009) find, using the *GIMIC* simulations that halo mass is the only driver of the star forming properties of galaxies. As in observations environment is usually contrasted with stellar mass (rather than halo mass), an observationally based distinction between mass and environment may tell us more about the stellar mass – halo mass relation than about the difference between external environment and halo mass.

In simulations, halo mass (and other halo parameters) are readily available. From simulations much 'cleaner' definitions of environment can be obtained, as the distance to other objects is very well known in three dimensions, contrary to observations which can only provide a precise distance perpendicular to the line of sight. Radial velocity differences give an indication of the distance along the line of sight, but peculiar velocities complicate a precise radial distance measure.

Many different measures of environment have been used in the literature. Some are closely related by construction, while the relation between others is more obscure. In this paper we compare several popular indicators of environments. The aim is to investigate which indicators correlate strongly with each other and with halo mass and which ones do not. We measure environmental parameters using a semi-analytic model for galaxy formation constructed on the merger tree of dark matter haloes formed in the Millennium Simulation (Springel et al., 2005), so that we also have halo masses available. We will present environmental parameters that measure halo mass, but are insensitive to external environment, along with environmental parameters that are insensitive to halo mass. These can be used for studies that aim to separate the effect of halo mass and external environment. We will show that most of the environmental indicators used in literature measure predominantly halo mass. In the remainder of the paper we will use the term 'environment' whenever we mean to quantify distances to nearby galaxies, surrounding galaxy densities etc., but never when referring to halo mass, in order to clearly distinguish the two.

This paper is organized as follows. Section 3.2 gives a short overview of the literature on environmental parameters, both from observations and simulations. In Section 3.3 we determine some of the often used environmental parameters and investigate their correlation with host halo mass. The strength of the correlation with halo mass depends on the distance scale used in the environmental parameters, as we will show in Section 3.4. In Section 3.5 we discuss how to construct an environmental parameter that is independent of halo mass. Finally, we conclude in Section 3.6.

3.2 Popular environmental parameters

The study of the effect of the environment on the evolution of galaxies has undergone considerable progress through large galaxy surveys, like the Sloan Digital Sky Survey (SDSS; Stoughton et al., 2002) and (z)COSMOS (Scoville et al., 2007; Lilly et al., 2007). Many different definitions of environmental density exist. Observationally, the density around galaxies must usually be based on the distribution of the galaxies themselves, as the full distribution of mass is very hard to measure reliably. In observational studies two slightly different flavours are very often used: one in which the number density of galaxies within a fixed distance are counted, and one in which the distance to the N^{th}e nearest neighbour is measured. Table 3.1 contains a short summary of the literature on the environmental dependence of galaxy properties, both from observations and from simulations. We will expand on these in this section and will study some of these in more detail using the galaxy catalogues in the Millennium database in the next section.

For the environmental parameters it is important, as we will show below, whether the masses of the other galaxies used to measure the environmental have a fixed physical lower limit (or luminosity), or whether the minimum mass is a fixed fraction of the mass of the galaxy one wants to know the environment of. It also matters whether the distance out to which the environment is measured is fixed in absolute terms or whether it is fixed relative to some length scale related to the galaxy in question (e.g. the virial radius of its host halo). In Table 3.1 we indicate for environmental parameter listed (described in the first column) out to what distance (or a distance equivalent parameter) the environment is measured (second column), and whether the minimum mass/luminosity of the galaxies used for the environmental estimate is fixed in absolute terms or whether it is a fixed fraction of the mass/luminosity of the galaxy in question (if applicable, third column). The final column lists references to papers employing the parameter. From Table 3.1 it is clear that only very few papers take minimum masses of neighbours and/or distances relative to properties of the galaxy's host halo.

Two main classes of observational parameters can be identified: those which measure the number of galaxies out to a given distance, and those which measure the distance out to a given N^{th} neighbour. Note that using the number of galaxies out to a given distance is equivalent to using the number density of that same sample of galaxies (and the same holds for the distance to N^{th} nearest neighbour and the density of galaxies in the volume out to the N^{th} nearest neighbour). These two broad classes of methods are not identical, but the difference is subtle. In high density regions the N^{th} neighbour is, on average, closer by and the scale on which the environment is measured is therefore smaller, while the other class of methods measures the density on a fixed scale.

The environmental parameters used in simulation studies are sometimes similar to the ones used for observations, but can also be very different. Using a similar definition allows one to directly compare models and observations. However, with the full (dark matter and baryonic) density field available, simulators can also determine parameters like the total amount of mass in spheres around the galaxy in question. Such quantities might influence the evolution of a galaxy, but are difficult or impossible to obtain observationally.

It is well known that high mass galaxies preferentially live in higher density environments. A correlation between halo mass and environmental density is therefore expected. For example, Kauffmann et al. (2004) use a semi-analytic model of galaxy formation to show how their measure of environmental density (number of galaxies within 2 h^{-1}Mpc projected, and a redshift difference less than 1000 km s^{-1}) correlates with halo mass. It is, however, unlikely that halo mass is the only characteristic of the environment that matters. With that in mind, Fakhouri & Ma (2009) have tried to construct an environmental parameters that does not scale with halo mass. They found that the mean over-density in a sphere of 7 Mpc, excluding the mass of the halo, gives the most mass-independent parameter of the three parameters they studied. They did not quantify the degree of correlation, but their

Table 3.1: Overview of environmental parameters that are frequently used in literature. They are grouped by the different ways of determining out to which distance the environment is measured either in observational or simulation studies. The first column specifies the environmental parameter, and the second and third column indicate out to what distance the environment is measured and whether the minimum mass/luminosity is fixed or scales with the galaxy in question. The fourth column specifies the references for the papers: 1: Dressler (1980), 2: Postman & Geller (1984), 3: Gómez et al. (2003), 4: Goto et al. (2003), 5: Whitmore & Gilmore (1991), 6: Whitmore et al. (1993), 7: Weinmann et al. (2006), 8: Cooper et al. (2005), 9: Cooper et al. (2006), 10: Cooper et al. (2008), 11: Balogh et al. (2004a), 12: Balogh et al. (2004b), 13: Baldry et al. (2006), 14: Bamford et al. (2009), 15: Cassata et al. (2007), 16: Pimbblet et al. (2002), 17: Lewis et al. (2002), 18: Blanton et al. (2003b), 19: Blanton et al. (2003a), 20: Hogg et al. (2003), 21: Hogg et al. (2004), 22: Blanton et al. (2005), 23: Kauffmann et al. (2004), 24: Blanton & Berlind (2007), 25: Kovač et al. (2010), 26: Fakhouri & Ma (2009), 27: Espino-Briones et al. (2007), 28: Ishiyama et al. (2008), 29: Lemson & Kauffmann (1999), 30: Harker et al. (2006), 31: Hahn et al. (2009), 32: Faltenbacher (2009), 33: Ellison et al. (2010), 34: Wilman et al. (2010), 35: Macciò et al. (2007), 36: Crain et al. (2009), 37: Hester & Tasitsiomi (2010), 38: Abbas & Sheth (2005)., 39: Maulbetsch et al. (2007), 40: Wang et al. (2007)

Parameter	Distance related parameter value	Minimum mass/luminosity	References
From observations			
(Projected) galaxy number density	Average of nearest 10 galaxies	$m_V < 16.5$	1, 5, 6
		$M_V < -20.4$	6
	Group average	$M_B < -17.5$	2
Cluster/Group-centric radius	-	$M_r < -20.5$	3, 4
	-	$m_V < 16.5$	5
	-	$M_V < -20.4$	6
	Scaled to the virial radius	$r < 17.77$	7
Projected galaxy number density out to the $N^{\rm th}$ nearest neighbour with a maximum radial velocity difference Δv	$N = 3$, $\Delta v = 1000$ km s^{-1}	$R < 24.1$	8, 9, 10
	$N = 4{,}5$	$M_R < -20$	11 - 15, 33
	$N = 5$, $\Delta v = 1000$ km s^{-1}	$M_r < -20.6$	11
	$N = 5$, $\Delta v = 1000$ km s^{-1}	$M_r < -20$	12
	$N = 4{,}5$, $\Delta v = 1000$ km s^{-1}	$M_r < -20$	13, 14
	$N = 10$	$I < -24$	15
	$N = 4{,}5$, $\Delta v = 1000$ km s^{-1}	$M_r < -20.6$	33
	$N = 10$	$M_V < -20$	16
	$N = 10$, in clusters	$M_b < -19$	17
	$N = 5, 10, 20$, $\Delta v = 1000$ km s^{-1}	$I_{AB} < 25$	25
Galaxy number density in sphere of proper radius r	$r = 8\ h^{-1}$Mpc, $\Delta v \leq 800$ km s^{-1}	$r < 17.77$	18 - 20
	$r \simeq 1\ h^{-1}$Mpc	$r < 17.77$	22
Number of neighbours in cylinders with projected radius r	$r = 2\ h^{-1}$Mpc, $\Delta v = 1000$ km s^{-1}	$r < 17.77$	23
	$r = 1\ h^{-1}$Mpc, Δv corresponding to 8 Mpc	$r < 17.77$	21
	$r = 0.1$ - $10\ h^{-1}$Mpc, $\Delta v = 1000$ km s^{-1}	$M_{0.1r} - 5\mathrm{Log}_{10}h < -19$	24
	$r = 1$ - $10\ h^{-1}$Mpc, $\Delta v = 1000$ km s^{-1}	$I_{AB} < 25$	25
	$r = 0.5, 1, 2\ h^{-1}$Mpc, $\Delta v = 1000$ km s^{-1}	$M_r < -20$	34
Projected galaxy number density in annuli	$1 < R/(h^{-1}\mathrm{Mpc}) < 3$	$r < 17.77$	23
	$\{0.5,1,2\} < R/(h^{-1}\mathrm{Mpc}) < \{1,2,3\}$	$M_r < -20$	34
From simulations			
Halo mass	-	$M > 2.35 \times 10^{10} M_\odot$	26
Number of neighbours in spheres of radius R	$R = 2\ h^{-1}$Mpc	$V_{\rm max} > 120$ km s^{-1}	37
Mass or density in spheres of radius R	$R = 5\ h^{-1}$Mpc	-	27, 28
	$R = 5, 8\ h^{-1}$Mpc	-	38
	$R = 7\ h^{-1}$Mpc	-	26
	$R = 1, 2, 4, 8\ h^{-1}$Mpc	-	35, 39
	$R = 18, 25\ h^{-1}$Mpc	-	36
Matter density in spherical shells	$2 < R/(h^{-1}\mathrm{Mpc}) < 5$	-	29, 30, 31
	$2 < R/(h^{-1}\mathrm{Mpc}) < 7$	-	26
	$R_{\rm FOF} < R < 2\ h^{-1}$Mpc	-	26
	$R_{\rm vir} < R < 3R_{\rm vir}$	-	40
Average mass density of surrounding halos	$N = 7$	$200 < V_{\rm max}/\mathrm{km\ s^{-1}} < 300$	32
Distance to nearest halo with minimum mass	-	$M_2/M_1 > 3$	28

plots indicate a weak, but non-negligible correlation with host halo mass. Observationally, this quantity cannot be determined. As far as we are aware no study to date has found a measure of environment that is independent of halo mass.

3.3 Environmental parameters and their relation to halo mass

In this section we will investigate the relation between several environmental parameters and the host halo mass. First we will briefly summarize the main characteristics of the synthetic galaxy populations used. For the environmental parameters discussed, we will distinguish between the 'ideal case' in which the three dimensional locations and the masses of all galaxies are known (as in simulations), and the case in which only projected distances and velocity differences can be measured and only luminosities are available, as is the case for for observations.

3.3.1 Simulations

We will compare different environmental parameters using the galaxy catalogue constructed using the semi-analytic model of De Lucia & Blaizot (2007, see also Croton et al. 2006), run on the dark matter-only Millennium Simulation (Springel et al., 2005). The Millennium Simulation follows the evolution of the dark matter distribution using 2160^3 particles in a periodic volume of 500 comoving h^{-1}Mpc from very high redshift down to redshift 0. The model of De Lucia & Blaizot (2007) uses recipes for the evolution of the baryons inside dark matter haloes and is based on the halo merger trees constructed using the halo catalogues of the Millennium Simulation. The model predicts the galaxies' locations, physical properties such as their stellar masses and star formation histories and observables like colours and luminosities. The model is calibrated to reproduce the redshift zero luminosity function in the K- and b_J-bands. De Lucia & Blaizot (2007), De Lucia et al. (2007) and Kitzbichler & White (2007) showed that this model reproduces many other observed properties of the galaxy population in the local Universe (e.g. the luminosity function at higher redshift, the colour distributions, the stellar mass function and the clustering properties). We will only use the $z = 0$ results.

We take into account all galaxies with stellar masses in excess of $10^{10} M_\odot$. This is roughly the same lower mass limit as Fakhouri & Ma (2009) use (they use $1.2\times10^{12} M_\odot$ total mass). The reason for this choice is an estimate of the resolution limit of these simulations. Boylan-Kolchin et al. (2009) show that the subhalo abundance of haloes in the Millennium Simulation is converged for subhaloes more massive than about $10^{11} M_\odot$, roughly independent of parent halo mass (as long as the parent mass is larger than $10^{12} M_\odot$). Guo et al. (2010) also investigate the subhalo abundance convergence of the Millennium Simulation. They compare the dark matter halo mass functions for main- and subhaloes together and conclude that halo and subhalo abundance is converged for $M > 10^{12.1} M_\odot$. These halo masses were matched by Guo et al. (2010) to the stellar mass function from the seventh data release of SDSS from Li & White (2009), from which they conclude that the observed galaxies with stellar mass $M_* \gtrsim 10^{10.2} M_\odot$ reside in converged haloes. The exact number of neighbours counted in some volume depends on the lower stellar mass limit for galaxies in the sample (or, correspondingly, the flux limit of the survey), but as we will show, the scalings and correlations are usually not sensitive to this lower limit.

3.3.2 The ideal case: using 3-dimensional distances and masses

We will use the simplest version of both classes of observationally determined parameters: the number of galaxies, N_R, within some volume with radius R and the distance to the N^{th} nearest neighbour, R_N. Parameters derived from these numbers (such as the number density of galaxies within that volume, etc.) will obey the same qualitative conclusions.

In Fig. 3.1 we show the correlations between host (Friends-of-Friends) halo mass and three definitions of environment: the number of galaxies within 1.5 virial radii of the galaxies' host haloes, the number of galaxies within 1 h^{-1}Mpc, and the distance to the fourth nearest neighbour (left to right). While $N_{1\,\mathrm{Mpc}/h}$ and particularly $N_{1.5R\mathrm{vir}}$ are strongly correlated with halo mass over the full mass range, halo mass only varies with R_4 for $R_4 \lesssim 2h^{-1}$Mpc (corresponding to $M < 10^{13.5} M_\odot$).

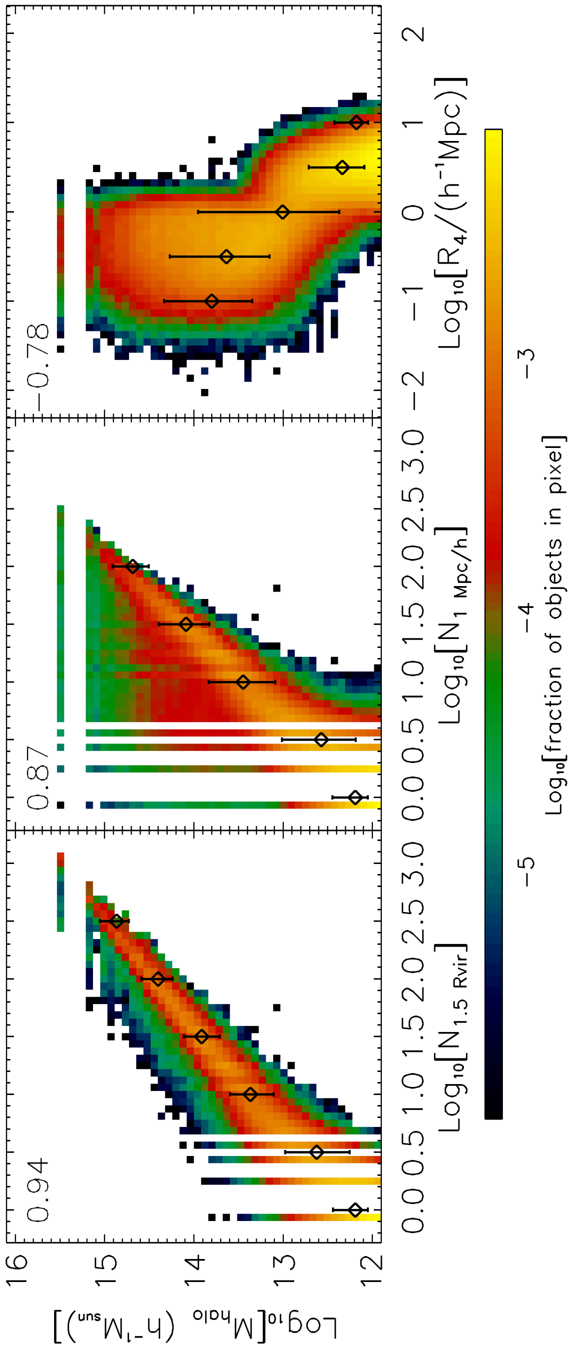

Figure 3.1: The correlations between halo mass and three environmental parameters: the number of galaxies within 1.5 $R_{\rm vir}$ (left panel), the number of galaxies within 1 h^{-1}Mpc (middle panel) and the distance to the fourth nearest neighbour (right panel) for all galaxies with $M_* > 10^{10} M_\odot$. The numbers printed in the top-left of each panel indicate the Spearman rank correlation coefficient and the diamond symbols and error bars show the binned medians and the 16th and 84th percentiles (for Gaussian distribution this would be $\pm 1\sigma$) of the distribution. All three parameters are a measure of halo mass. From left to right the correlations with halo mass decrease in strength.

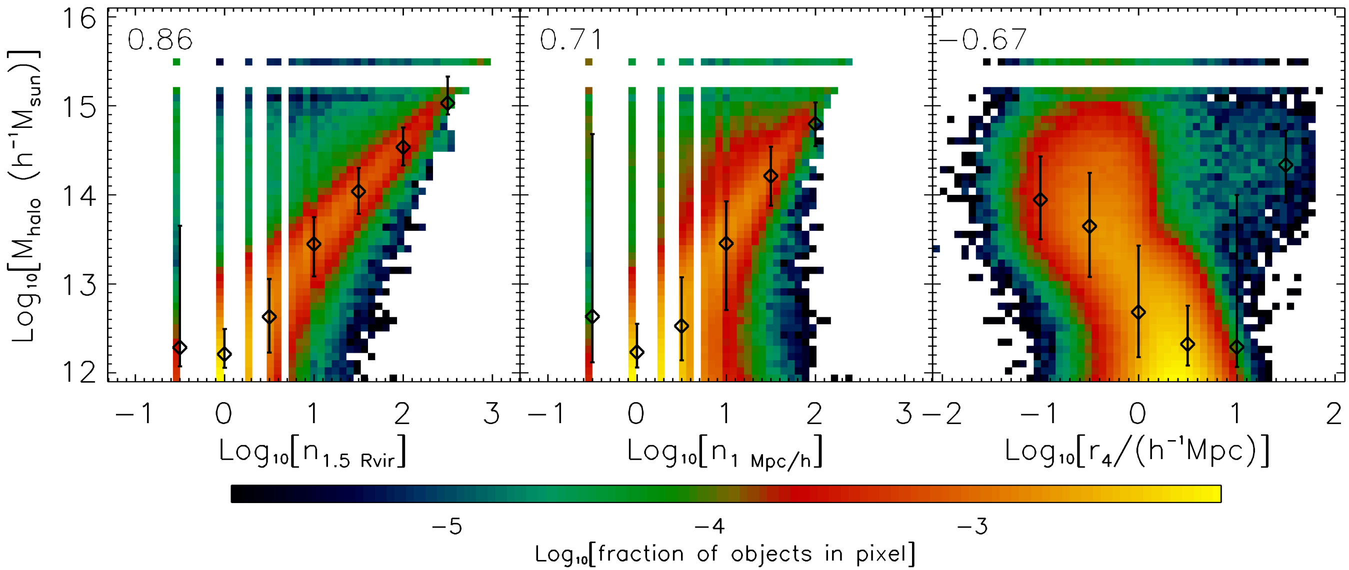

Figure 3.2: As Fig. 3.1, but now for observable versions of the environmental parameters: K-band luminosities rather than stellar masses, projected distances and a cut in redshift difference rather than 3-D separations for all galaxies with $K < -23$. For the left panel, the virial radii of the host haloes still need to be known. The galaxies with $n < 1$ are placed at $\mathrm{Log}_{10}n = -0.5$ for illustrative purpose. The correlations are slightly weaker than the ones found for the ideal case (Fig. 3.1), mainly due to projection effects, which make galaxies populate the regions in the plots which were unoccupied in Fig. 3.1. The correlations are, however, still strong.

Number of galaxies within a given distance

If the distance out to which galaxies are counted is scaled to the virial radius of the halo that the galaxy resides in, then the correlation between halo mass and environment is very strong, as is shown in the left panel of Fig. 3.1. Because the region within which galaxies are counted grows with halo mass, a more or less constant fraction of the satellites is counted. A fixed fraction of all satellites is a number of satellites that grows roughly linearly with halo mass, resulting in a very tight correlation. This can be understood in terms of the results found by Gao et al. (2004): the fraction of the mass in subhaloes, the distribution of subhaloes and the shape of the subhalo mass function are independent of host halo mass, while the normalization (so the total number of and total mass in subhaloes) scales (to first order) linearly with halo mass. The number of subhaloes (and thus satellite galaxies) within a radius that is fixed relative to the virial radius therefore grows roughly linearly with halo mass. This makes the parameter $N_{1\ \mathrm{Rvir}}$ a very strong measure of halo mass.

A slightly weaker correlation exists between halo mass and the number of galaxies within a fixed physical distance, as shown in the middle panel of Fig. 3.1 (for a distance of 1 h^{-1}Mpc). The upper envelope is populated by the central galaxies in the sample, while the satellites form the less tightly correlated cloud below the relation of the centrals. At the high mass end there are more galaxies with $M_* > 10^{10}\ M_\odot$ per halo, causing the correlation between $N_{1\mathrm{Mpc}/h}$ and M_{halo} to weaken.

Distance to the N^{th} nearest neighbour

In the right panel of Fig. 3.1 we show the correlation between the host halo mass and the distance to the fourth nearest neighbour, R_4 (which is very often used observationally, see Table 3.1). The distance R_4 decreases with halo mass, because more massive haloes are on average found in denser environments.

For halo masses $M > 10^{13.5}\ M_\odot$ the correlation between R_4 and M becomes much weaker. This behaviour arises from the fact that for low halo masses the 4$^{\mathrm{th}}$ nearest neighbour (with $M_* > 10^{10}\ M_\odot$) resides in another halo, whereas at high masses we are counting galaxies within the same halo. The transition between the two regimes depends on the rank n: for higher ranks, the jump occurs at higher halo mass.

The three parameters displayed in Fig. 3.1 all depend on three-dimensional distances. We will now proceed to investigate parameters that are observationally more feasible.

3.3.3 The realistic case: using projected distances and luminosities

Observationally we have no access to the three-dimensional separations between galaxies. Instead, one measures distances projected on the sky and differences in redshift. Moreover, while luminosities are readily available, stellar mass determinations depend on SED modelling, which comes with considerable uncertainty. We will now investigate to what extent the use of observables weakens the correlations compared with the 'ideal cases' discussed in Section 3.3.2. As is done in many observational studies (see Table 3.1) we will only make use of galaxies with redshifts that are within 1000 km s^{-1} of the redshift of the galaxy for which the environment is determined. We include both the Hubble flow and peculiar velocities in our calculation of the redshifts. For reference, a velocity difference of 1000 km s^{-1} corresponds to a distance of 10 h^{-1}Mpc if the peculiar velocity difference is zero. We will denote the parameters using the same symbols as we used for the 3-D distance variants, but with lower case letters. For example, r_4 denotes the projected distance to the fourth nearest neighbour (using only galaxies within the redshift difference cut). We only include galaxies with an absolute K-band magnitude smaller than -23, which corresponds to $M_* \approx 10^{10.2} M_\odot$. This results in a slightly smaller sample than the one used before. For the sample of galaxies with $M_* > 10^{10} M_\odot$, the luminosity function shows signs of incompleteness at magnitudes fainter than $K = -23$.

In Fig. 3.2 we show the dependence of the parameters similar to those used in Fig. 3.1, but using projected distances and luminosities rather than 3-D distances and stellar masses. Note that the left panel still requires knowledge of the virial radius of the host halo of the galaxy and is therefore hard to determine observationally (we left it in for completeness). The virial radius can be estimated if one has a group catalogue available, like the one by Yang et al. (2007) who grouped galaxies using a a friends-of-friends like algorithm. The total luminosities of the groups are then ranked and matched to a ranked list of halo masses, drawn from a halo mass function sampled in a volume equal to that of the survey. This procedure results in the assignment of a host halo mass to all galaxies in the sample. However, if such a catalogue is available, then the halo mass is of course just as well known as the virial radius, so using this environmental indicator as a measure of halo mass is not very useful.

In the middle panel of Fig. 3.2 we show the halo mass as a function of the number of galaxies with a projected distance less than 1 h^{-1}Mpc, with a redshift difference less than $\pm$1000 km s^{-1} and with $K < -23$. Compared with the 3-D version, there are now more low mass galaxies with a high number of neighbours. This is due to projection effects. We note that the correlation coefficient is still very high ($\approx$0.71), so we can conclude that this environmental indicator is a strong indicator of host halo mass. The horizontal scatter (in environmental parameter for fixed halo mass) at low halo masses (roughly 0.3 dex upwards and downwards in number of neighbours) is dominated by the projection effects, while at high masses the scatter (0.2 dex upwards, 0.5 dex downwards in number of neighbours within the projected distance) is mainly caused by satellites in the outskirts of the halo. The scatter in the environmental indicator is smallest for halo masses of about 10^{14} $M_{\odot}$, where it is roughly 0.2 dex both upwards and downwards. For a given $n_{1\ \mathrm{Mpc}/h}$ the spread in halo masses is small for low and high values of the environmental indicator (roughly 0.3 dex) and highest for values of about 10 neighbours within this distance ($\gtrsim 0.5$ dex in halo mass) and is roughly symmetrical.

In the right panel of Fig. 3.2 we show the projected distance to the fourth nearest neighbour with $K < -23$. Because of projection effects the bi-modal behaviour visible in the right panel of Fig. 3.1 has been smeared out. The correlation with host halo mass is therefore slightly weaker. Because of the discontinuity in the distribution, the correlation coefficient is a function of the masses (both galaxy stellar mass and host halo mass) of the objects that are taken into account.

3.3.4 A multi-scale approach

Wilman et al. (2010) recently measured the number density of galaxies in concentric rings in order to investigate trends in the $u - r$ colour distribution of galaxies with environment at several distance scales (for given small-scale density, if desired). They included all galaxies from the fifth data release of SDSS with magnitude brighter than 17.77 in the r-band and with a mean surface brightness within the half-light radius of $\mu_r \leq 23.0$ mag arcsec^{-2}. The number density of galaxies was determined in rings with radii fixed in physical coordinates. In this approach neither the mass nor the distance out to which the environment is determined scales with the properties of the galaxy in question. We therefore expect that these measures of environment vary strongly with halo mass.

The correlation coefficient for the density in annuli with halo mass is roughly 0.5, and depends on both the width and the radius of the annulus, such that smaller radii (within $\sim$ 0.5 Mpc) have larger correlation coefficients and wider annuli mostly show weaker correlations. The power of the method of Wilman et al. (2010) lies in the ability to measure residual trends of galaxy properties with large-scale (annular) environment, while controlling for the environment on some smaller scale (i.e. the projected number density in the inner circle, using the same definitions as our n parameter above). The samples are constructed by taking all galaxies for which the number density of galaxies within the inner radius of the annulus fall within some bin, and are therefore comparable to horizontal slices through the middle panel of Fig. 3.2. From this figure we can see that in such a slice, a very large range of halo masses (comparable to the full range of halo masses in the catalogue) is still present.

As an example, we show in Fig. 3.3 the correlation between halo mass and the number of galaxies in annuli with an inner and outer radius of 1 and 2 Mpc, respectively, for three narrow bins of the number of galaxies within 1 Mpc (projected distance, within a redshift difference of 1000 km s^{-1}). Each bin contains 1/8 of all the galaxies, where the lowest bin shown (the second panel from the left) corresponds to the lowest 1/8 of the total galaxy population, the middle panel shows the middle 1/8 and the right-hand panel shows the 1/8 galaxies with highest numbers of galaxies within 1 Mpc. From the colour scale it can clearly be seen that the different bins in central number density of galaxies favour different halo masses, as expected from Fig. 3.2.

The correlation coefficients are low, for the second and third panel from the left, which seems to make these parameters nearly halo mass independent. Looking more closely at the Figure, we see, however, a positive correlation between median $n_{\text{1-2 Mpc/h}}$ and M_h, especially at high mass. The relation with halo mass of this measure of large-scale environment, at fixed small-scale environment depends strongly on the (fixed) scales at which the environment is measured. This, together with varying flux limits in observational surveys makes it a fuzzy measure of halo mass, which is hard to interpret physically.

The trends seen in Fig. 3.3 are a typical example of the 'multi-scale' approach of Wilman et al. (2010). Changing the radii of the inner and outer edges of the annuli and/or the width of the bins in central galaxy number density does not affect the qualitative conclusions drawn from Fig. 3.3. The correlation of the number of galaxies in annuli with halo mass becomes weaker if very large distances from the galaxy in question are taken (5-10 Mpc), but it seems likely that that is merely a result of the fact that galaxies at such distances do not have much to do with the galaxy in question anyway.

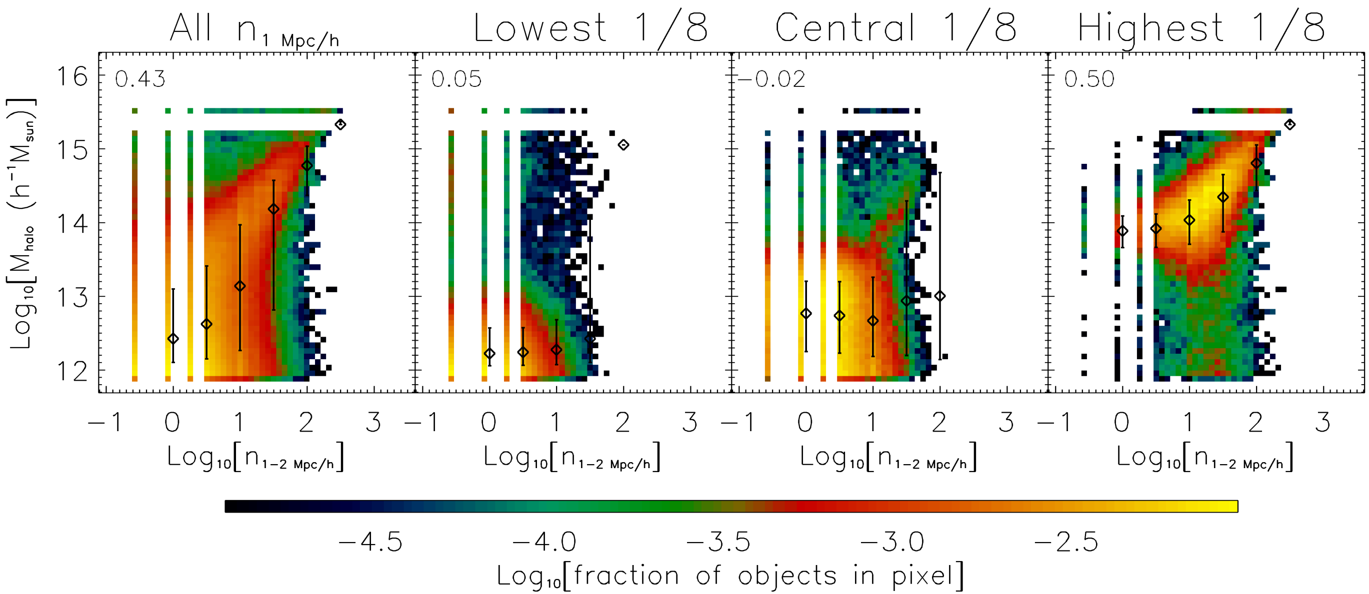

Figure 3.3: Halo mass as a function of the number of galaxies in annuli with an inner and outer radius of 1 and 2 Mpc, respectively. The first panel shows the full sample, while the other three panels correspond to three bins in number of galaxies within 1 Mpc, showing the lower, middle and upper one eighth of the total distribution from left to right. The galaxies with $n < 1$ are placed at $\mathrm{Log}_{10}n = -0.5$ for illustrative purpose. Although the range of halo masses in each bin is large, the typical halo mass is higher for the bins corresponding to higher values of $n_{1\ \mathrm{Mpc/h}}$. The median $\log(M/M_{\odot})$ are 12.3, 12.7 and 14.2, for the three bins respectively. Within each narrow bin of 'small-scale environment' there is, at the high halo mass end, still a correlation left between $n_{1\text{-}2\ \mathrm{Mpc/h}}$ and halo mass. The low Spearman rank correlation coefficients are dominated by the low halo masses which are much more abundant in every bin of small scale environment.

3.4 Environment as a measure of halo mass

In this section we will study the strength of the correlation between several environmental indicators and halo mass.

We expect the correlation between the number of neighbours and halo mass to be strongest at some given distance. Taking the distance very small will bias against massive galaxies (and results in strong discreteness effects if the number of neighbours is very small, as they can only be integer). Taking the distance too large, on the other hand, will result in a sample of galaxies that does not have much to do with the halo the galaxy resides in.

In Fig. 3.4 we show, for two different environmental parameters, the value of the Spearman rank correlation coefficient with halo mass, as a function of the distance related parameter used to measure the environmental density. In the left panel we show the correlation coefficient between halo mass and the environmental density indicator n_r (the number of galaxies within a fixed physical distance r projected on the sky and within $\Delta v = \pm 1000$ km s^{-1}) as a function of r. One example of this type of parameter was shown in the middle panel of Fig. 3.2. Fig. 3.4 shows that the correlation first strengthens with distance, reaches a maximum at a scale of roughly 1 h^{-1}Mpc, and declines slowly thereafter. The vertical arrows indicate the median virial radii for the haloes of all galaxies in the sample of the corresponding, and show that the peak of the correlation strength occurs at distances roughly corresponding to the median virial radius.

In the right panel of Fig. 3.4 we plot the Spearman rank correlation coefficient between halo mass and environment, now parametrized by r_N, the distance towards the N^{th} neighbour (as in the right panel of Fig. 3.2), as a function of the rank N. The correlation coefficients are now mostly negative, as a higher density (corresponding to a higher halo mass) will result in a smaller distance towards the N^{th} neighbour. However, for very massive haloes the distance to the first neighbour is an increasing function of mass, as the neighbour needs to be outside the galaxy itself, and more massive galaxies are larger. Taking more neighbours gives an anti-correlation that becomes stronger for larger numbers of neighbours for high mass galaxies. Lower mass galaxies show the strongest correlation when the distance to the N^{th} nearest neighbour is taken, with $N \gtrsim 3$, but the correlation does not weaken much for larger values. For a sample consisting of very high luminosity galaxies, slightly more neighbours need to be included to get the best measure of halo mass. The median number of neighbours within the virial radius, above the same luminosity cut is indicated with the arrows.

The vertical arrows in Fig. 3.4 indicate the median virial radius of the samples in the corresponding colour (left panel) and the median number of neighbours above the same luminosity limit within the virial radius (right panel). We conclude

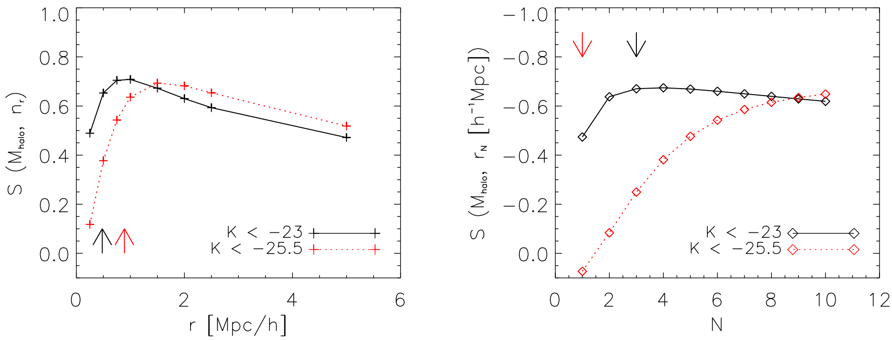

Figure 3.4: The strength of the correlations between halo mass and two of the environmental indicators used straightforwardly in observations, for two samples, with lower luminosity limits as indicated. In the left panel we plot the Spearman rank correlation coefficient between halo mass and the number of galaxies within a given projected physical distance r (and with a cut in redshift difference, as described in the text) as a function of r. The arrows show the value of the median virial radius of the haloes of all galaxies in the sample with the corresponding colour. The right panel shows the Spearman rank correlation coefficient between halo mass and the projected distance to the $N^{\rm th}$ nearest neighbour as a function of the rank N. The correlation coefficient is negative, because more massive galaxies have their $N^{\rm th}$ nearest neighbour closer by. The arrows indicate the median number of neighbours within the virial radius of the haloes above the indicated flux limit. If the environmental parameter is supposed to be a measure of halo mass, galaxies out to a distance of ~1 Mpc is a good choice, or the distance to the $N^{\rm th}$ neighbour, with $N = 1$ or 2. This second parameter is a worse measure of halo mass than the first, though the difference is small.

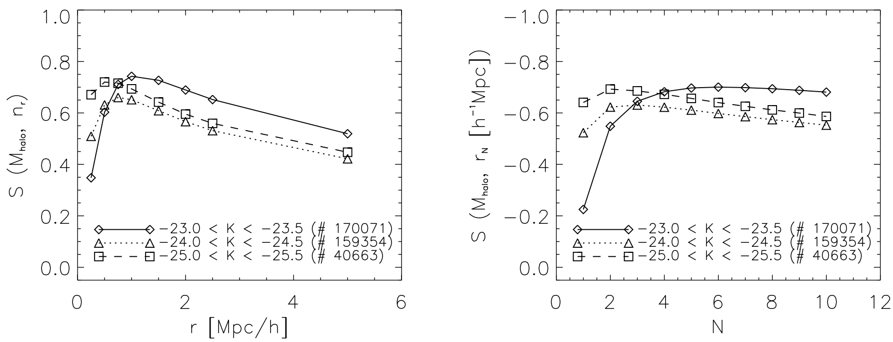

Figure 3.5: The same as Fig. 3.4, but now for three bins in absolute magnitude. We show the correlation coefficients between halo mass and the observationally feasible environmental parameters. For the neighbour search all galaxies with $K < -23$ are taken into account. The shape of the relation between correlation coefficient and the distance related parameters are relatively insensitive of mass, but the correlations are stringer for samples with higher luminosity galaxies. The numbers in between the brackets indicate the number of galaxies in the sample.

that n_r and r_N are both good measures of host halo mass, provided that n_r is measured at $r \geq r_{\rm vir}$ and/or that the rank of neighbours taken into account is small. If the host halo mass, and thus the virial radius, are not known a priori, it is better to take r larger, as the correlation rapidly weakens towards smaller distances and declines only slowly with increasing distance.

In Fig. 3.5 we break up the samples of Fig. 3.4 in bins of K-band magnitude. In the neighbour search we include all galaxies with $K < -23$, but we plot the Spearman rank correlation coefficient between the environmental parameters and host halo mass for bins of $\Delta K = 0.5$. The correlations are in general weaker than for the whole sample, although the maxima are very comparable. $K-$band luminosity correlates with stellar mass (although at low masses the mass to light ratios vary stronger), so together with the correlation between stellar and halo mass (which is very strong for central galaxies, which make up roughly half the sample averaged over all stellar masses, and a larger fraction for higher stellar mass or $K-$ band luminosity) one expects to weaken the correlation with halo mass if a narrow range of $K-$band luminosities is taken. Brighter samples of galaxies are more dominated by central galaxies, for which the correlations between halo mass and environmental indicator are stronger.

As we will show below, using K-band luminosity as a proxy for (virial) mass works well. Guided by the left panel of Fig. 3.2 one might expect that we can improve on n_r as a measure of halo mass if r scales with $L_K^{1/3}$. We have tried this, but the correlation between halo mass and environment does not get stronger (or it gets slightly weaker, with correlation coefficients of 0.65 – 0.7). In the range of halo masses for which we could test it (any range between 10^{12} and $10^{15.5} M_\odot$) the correlation is stronger if a projected distance of 1 Mpc is used than if $r \propto L_K^{1/3}$ is used. Specifically, we tried $r = 1h^{-1}\mathrm{Mpc}\cdot(L_K/L_0)^{1/3}$, with $L_0 = 10^{\{10.5,11.0,11.5,12.0\}} L_\odot$. We therefore conclude that using a fixed physical projected distance is safe, and easier in practice than a distance scaling with luminosity. We thus advise to use n_r with r of the order of $r \gtrsim R_{\mathrm{vir}}$, if a measure of halo mass is desired. For most observed samples of galaxies $r \sim 1$ Mpc will do, but by iteration better values can be obtained: use $r = 1\ h^{-1}$Mpc, calculate the halo virial radii from the environmental indicator (using the parametrization given in Appendix 3.6) and then iterate if the virial radii strongly deviate from 1 Mpc.

In Appendix 3.6 we provide polynomial fits for the halo mass as a function of several environmental parameters for several lower flux limits, which can be used to obtain halo masses from observed samples of galaxies with measured environmental indicators.

3.5 Environment independent of halo mass

3.5.1 Mass independent parameters for simulations

All the parameters we have looked at so far correlate with halo mass. The lower mass/luminosity limit of galaxies included as possible neighbours was set equal to the resolution limit of simulations, or the flux limit of a survey. As we saw in the left panels of Figs. 3.1 and 3.2, the correlation is strongest and almost linear with halo mass, if the scale out to which galaxies are counted scales with the virial radius of the host halo of the galaxy in question. Per unit halo mass, this galaxy number density (either projected or in a spherical region) is therefore roughly constant. This also holds for dark matter subhaloes in high resolution simulations, as shown by Gao et al. (2004).

In order to obtain an environmental indicator that is independent of halo mass we have to scale out both the mass/luminosity of the galaxy and the length scale in question. We define $D_{N,f}$ to be the three-dimensional distance to the N'th nearest neighbour with at least f times the virial mass of the halo under consideration, divided by the virial radius of the halo under consideration:

$$D_{N,f} = \frac{r_{N(M_{\mathrm{vir}} \geq f \cdot M_{\mathrm{halo}})}}{R_{\mathrm{vir,\ ngb}}} \tag{3.1}$$

where the subscripts 'ngb' and 'halo' indicate the neighbour of the halo under con-

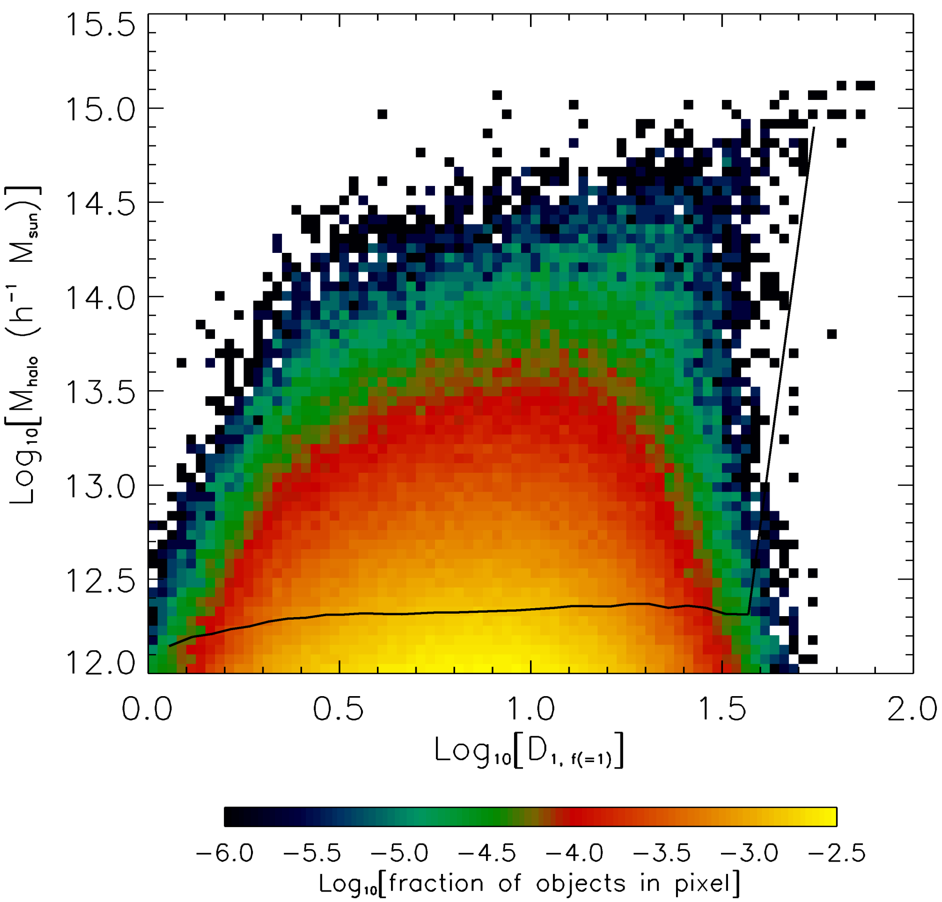

Figure 3.6: Halo mass as a function of the parameter $D_{1,1}$. The colour scale gives the distribution for all central galaxies in the sample, while the solid line is the median halo mass in bins of $D_{1,1}$. The median relation is very flat. The correlation coefficient of this parameter with halo mass is 0.07 (for correlation coefficients as a function of rank, see Fig. 3.7). We can therefore conclude that this measure of environment is highly insensitive to halo mass. At the high $D_{1,1}$ end, where the median halo mass is very high, there is a residual correlation visible because these haloes are on the exponential tail of the mass function.

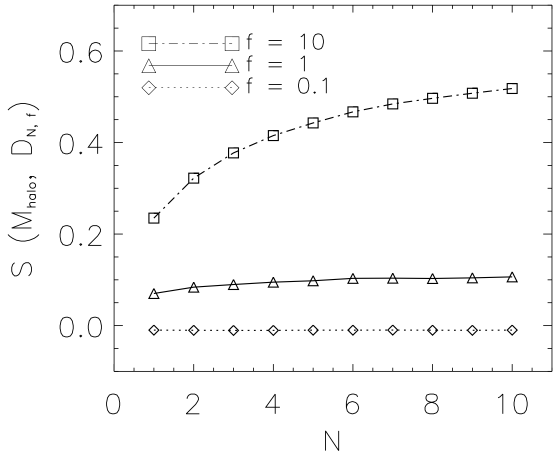

Figure 3.7: The Spearman rank correlation coefficient between halo mass and the environmental indicator $D_{N,f}$ (see Eq. 3.1) as a function of the rank N, for $f = \{1/10, 1, 10\}$. Higher values for f and N result in a stronger correlation in the range of ranks N and halo masses we tried. As $f = 1$ still gives a very small rank correlation coefficient, and because the environmental parameter can only be determined for the whole sample of galaxies for $f \gtrsim 1$, we conclude that using $f = 1$ and a low rank (e.g. $N = 1$) is a good choice if an environmental parameter that is insensitive to halo mass is desired. If haloes can be reliably identified for mass lower than the lowest mass one wants to know the environment for, then a value for f as low as possible should be used.

sideration and the halo itself, respectively. As we are dealing with halo properties, we only take central galaxies (i.e. only Friends-of-Friends haloes) into consideration. The use of the factor f to set the minimum mass of haloes taken into account in the neighbour search and the scaling to the virial radius are the two ingredients that we expect to make the environmental parameter insensitive to mass. $D_{N,f}$ only depends on the dimensionless parameters N and f for a given halo, and is also itself dimensionless.

Because the tidal field of the N'th nearest neighbour scales with the mass of and distance to this neighbour as M/R^3 and the mass scales with $R^3_{\rm vir}$, the parameter $D_{N,f}$ scales with the tidal field to the power $-1/3$. This makes $D_{N,f}$ a very natural environmental parameter for which the physical interpretation is clear.

The colour scale of Fig. 3.6 shows the distribution of haloes at $z = 0$ in the

$D_{1,1} - M_{\rm halo}$ plane. The curve shows the median $D_{1,1}$ in bins of halo mass. The median halo mass found is always the same for all D, irrespective of the factor f. The median $D_{1,f}$ in the sample is different for different f, though.

The weak correlation that starts to appear at very high values for $D_{1,f}$, especially for large f, is caused by the fact that these are probing the most massive haloes that are on the exponential tail of the Schechter-like halo mass function. Large scale structure is no longer self-similar in that regime, causing a slight positive correlation between $D_{N,f}$ and halo mass. We have verified (by inverting the axes) that for masses $M \ll M_*$ (where M_* is the mass at which the Schechter-like halo mass function transits from a power law into an exponential fall-off), where the mass function is a power law (and therefore scale free) the correlation is very weak. For higher masses, there is a mass scale imposed by the exponential cut-off of the Schechter-like halo mass function. For values roughly above $f^{-1}M_*$, the insensitivity to mass breaks down and a weak positive correlation between halo mass and $D_{N,f}$ appears.

In Fig. 3.7 we show the correlation coefficients between halo mass and $D_{N,f}$ as a function of the rank N for three different values of the mass ratios of galaxies counted as neighbours $f = \{1/10, 1, 10\}$. For all f the correlation between the rank N and host halo mass increases for with the rank. If an environmental indicator is desired that is insensitive to halo mass, $N = 1$ is therefore a good choice. The correlation is weaker for lower values of the ratio between host halo mass and the masses of possible galaxies that are included in the neighbour search. For a value lower than $f = 1$ the environmental indicator cannot be determined for the full resolved sample of haloes (as halo masses need to be at least $M > f^{-1}M_{\rm res}$, with $M_{\rm res}$ the resolution limit, in order to resolve all possible neighbours). We therefore advise to take $f = 1$, as then the parameter can be defined for all galaxies in the sample and it gives only a very weak correlation with halo mass. If in a sample of haloes some of the studied properties demand a much more stringent resolution limit (e.g. if detailed halo profiles need to be fitted), and if haloes of much lower mass are resolved in terms of their virial mass and position, then one should use values of $f < 1$, e.g. 0.1, as the correlation between halo mass and environment vanishes.

If in the definition of $D_{N,f}$ the virial radius of the neighbour would be replaced by the virial radius of the halo under consideration (thereby losing the connection to the tidal force of the neighbour), the correlation between halo mass and environment gets even slightly weaker (e.g. a Spearman rank correlation coefficient of 0.04 instead of 0.07 between halo mass and $D_{1,1}$). As using the virial radius of the neighbour gives a more intuitive external environmental parameter, we still advice to use the virial radius of the neighbour.

We can conclude that the parameter $D_{N,f}$, with $N = 1$ and $f \leq 1$ results in an intuitive environmental parameter that is very insensitive to halo mass. We do note, however, that in order to calculate this halo mass independent environmental indicator, one needs a measure of the virial mass of the host halo. From simulations these can be obtained trivially. For observed samples of galaxies this can be estimated using the environmental indicators that do correlate with halo mass strongly, as described in the previous section and detailed in Appendix 3.6. In the next section we will present an environmental indicator that can be obtained from observations that is also insensitive to halo mass.

3.5.2 Halo mass independent parameters for observed samples of galaxies

In some cases it is possible to obtain virial masses and radii for the host haloes of observed galaxies. Using techniques like halo-matching, in which the total luminosity of all galaxies in a group or cluster are added and the ranked luminosities matched to a ranked list of halo masses (from either an analytic halo mass function or a simulation), it is possible to get a reliable estimate for the host halo virial mass of the observed galaxies, see e.g. Yang et al. (2003); van den Bosch et al. (2003); Yang et al. (2007). This requires, however, that a group catalogue is available for the observed sample of galaxies. As such catalogues are only available for a limited number of observational samples, it is something which is often not easily done.

Hence, observationally neither the halo mass independent environmental indicator $D_{N,f}$ nor the virial mass or radius of a halo can be easily determined. We therefore set out here to formulate a variable that can be very easily determined by observers and that is as independent of halo mass as possible. We let the definition of $D_{N,f}$ guide us. We know that we have to scale the minimum masses/luminosities of the galaxies that are taken into consideration in the search for neighbours to be a fixed fraction of the mass/luminosity of the galaxy under consideration and that we have to scale the distance to the neighbours to some typical distance of the neighbour.

We use an observable, the K-band luminosity, instead of stellar mass. Luminosity is easier to measure and does not require the modelling of the spectral energy distribution of the galaxy. We use the K-band because in the very red optical bands and in the near-IR the correlation between luminosity and stellar mass is strongest (aside from the uncertainties arising from the treatment of thermally pulsing asymptotic giant branch, TP-AGB, stars, see e.g. Maraston, 2005; Tonini et al., 2010). We will also have to normalize in distance. As a reference we use typical values for central galaxies in a halo with a virial mass of $10^{13}\,M_{\odot}$, and therefore a virial radius equal to 0.58 h^{-1}Mpc.

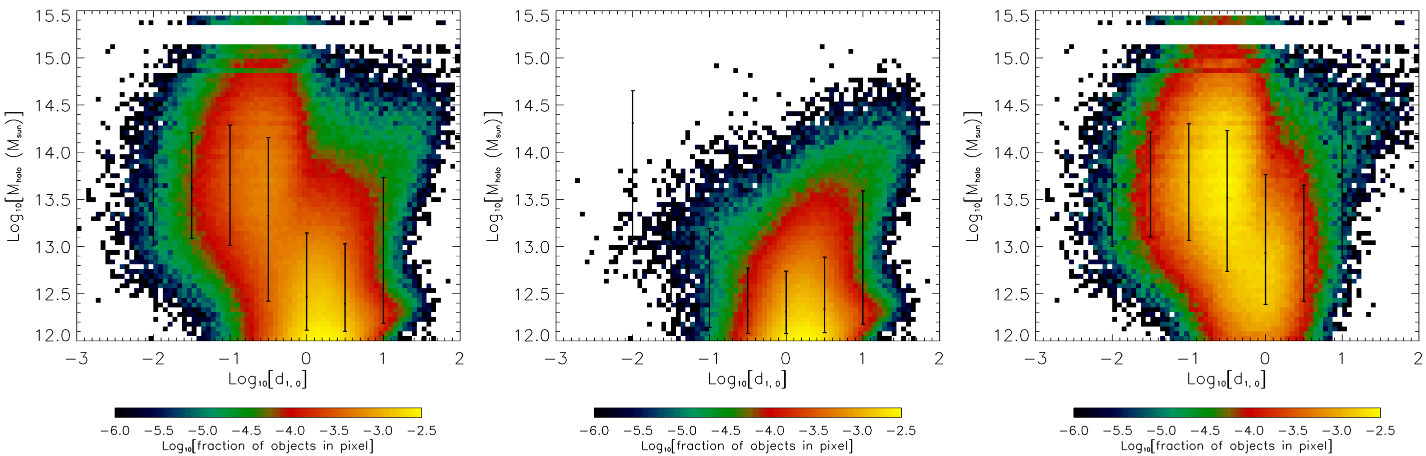

Figure 3.8: The host halo masses of galaxies versus the dimensionless environmental parameter $d_{1,0}$ (see Eq. 3.2), which is easy to measure for observers. In the left panel we show all galaxies in the sample, for which the Spearman rank correlation coefficient is -0.28. The middle and left panel show the relation for the central galaxies and the satellites, respectively. For these subsamples the Spearman rank correlation coefficients are 0.09 and -0.35, respectively, showing that the relation is very weak for central galaxies, as expected, but not so weak for satellite galaxies.

For all central galaxies in a bin of halo mass extending from $10^{12.9}$ to $10^{13.1} M_\odot$ we have determined the median K-band luminosity to be $1.4 \times 10^{11} L_\odot$. The virial radius, which is used in the definition of $D_{N,f}$, scales with halo mass as $R_{\rm vir} \propto M_{\rm halo}^{1/3}$, so we scale the distance used to normalize the environment as $r \propto L_K^{1/3}$ (see below for the neighbour search strategy). As projected distances are more easily measured than three dimensional distances, we use the projected distances (and test both with and without a cut in velocity difference). Our environmental indicator $d_{N,m}$ then becomes

$$d_{N,m} = \frac{r_{{\rm N}(K \leq K_{\rm gal} - m)}}{0.58 h^{-1}{\rm Mpc}} \cdot \left(\frac{L_{K,{\rm ngb}}}{1.4 \times 10^{11} L_\odot} \right)^{-1/3} \qquad (3.2)$$

where the subscript ‘ngb’ again denotes the neighbour of the galaxy in question, m is the difference in magnitudes (corresponding to a ratio in luminosity/mass, a positive m means that the neighbours must be brighter) between the galaxy in question and the galaxies counted as possible neighbours (we will show $m = 0$ below, and therefore look only for neighbours that are at least as bright as the galaxy under consideration), K is the absolute K-band magnitude and L_K the luminosity in the K-band. $R_{\rm vir,13} = 0.58 h^{-1}$Mpc is the virial radius of the ‘reference mass’ of $10^{13} M_\odot$.

If $R_{\rm vir,13}(L_K/1.4 \times 10^{11} L_\odot)^{1/3}$ would be the virial radius (i.e. if the halo mass to K-band light ratio would be constant), then the external environmental indicator $d_{N,m}$ could be described as distance to the $N^{\rm th}$ nearest neighbour which is at least m magnitudes brighter than the galaxy we are measuring the environment of, normalized to the galaxy’s virial radius.

The colour scale in the left panel of Fig. 3.8 shows the distribution of galaxies in the $M_{\rm halo} - d_{1,0}$ plane. We include all galaxies in the catalogue with $K < -23$. The sample of galaxies with $M_* > 10^{10} M_\odot$ shows signs of incompleteness at magnitudes fainter than $K = -23$. Fig. 3.8 shows that halo mass indeed is weakly sensitive to the parameter $d_{1,0}$.The Spearman rank correlation coefficient is -0.28, which indicates a weak anti-correlation.

The parameter shown in Fig. 3.8 includes only galaxies within a radial velocity difference of 1000 km s^{-1}. Without this cut in redshift difference the correlation becomes stronger. Taking into account only galaxies within a redshift window is important, but the width of the redshift window is less important as long as it is $\lesssim 10^3$ km s^{-1}.

The dependence of the correlation between host halo mass and $d_{N,m}$ on the rank N is shown in Fig. 3.9, for three different values of m. We have chosen to show $m = \{-2.5, 0, 2.5\}$ magnitudes, because a magnitude difference of 2.5 corresponds to a luminosity ratio of 10, similar to the mass ratio of 10 used above. Whenever

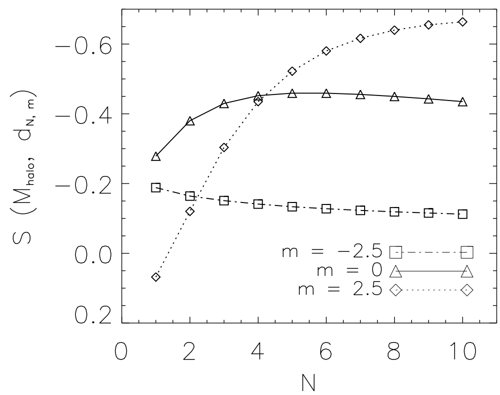

Figure 3.9: The Spearman rank correlation coefficient between halo mass and $d_{N,m}$ as a function of the rank N, for $m = \{-2.5, 0, 2.5\}$ magnitudes. 2.5 magnitudes corresponds to a factor 10 in luminosity. For the sample for $m = 2.5$ magnitudes there are fewer possible neighbours and the nearest neighbour will usually be found in another halo (often even a more massive halo), causing a weak correlation with halo mass. In the sample for $m = -2.5$ magnitudes, the parameter is only defined for a small sample, because neighbours, which have a luminosity 10 times lower than the galaxy in question, need to be resolved as well. If the neighbours are not required to be much more luminous ($m = 0$) they can be either in the same or in another halo, causing a correlation with halo mass that rises for low rank and decrease for higher ranks.

possible neighbours are supposed to be a factor 10 less luminous ($m = -2.5$), the sample for which this parameter can be determined is much smaller (because all possible, lower mass neighbours need to be resolved as well) and the typical haloes the galaxies are in are more masssive. This results in the very weak correlation with halo mass for all ranks N, as shown in Fig. 3.9. If neighbours are required to be more than a factor 10 brighter, the most likely neighbours will reside in other (more massive) haloes. If the minimum brightness of possible neighbours is the same as that of the galaxy in question, or higher, the correlation between host halo mass and $d_{N,0}$ first increases with the rank N and goes down after some maximum (because for large rank N the neighbours are more likely to reside in other haloes). This maximum and the rank at which the maximum occurs depend on the lower luminosity limit of the sample and on the difference in magnitudes m. The lowest possible rank $N = 1$ gives a very weak correlation and for the same reason as before we advice to use a luminosity ratio of 1 ($m = 0$) between the galaxy in question and its possible neighbours. Again, if neighbours within a redshift window can be identified below the flux limit used for the analysis, it is wise to use a value for m as low as possible.

3.5.3 Splitting the sample in centrals and satellites

The middle and right panel of Fig. 3.8 show the distribution of central galaxies and satellites, respectively, in the $M_{\rm halo} - d_{1,0}$ plane. For these subsamples the Spearman rank correlation coeffeicient between $d_{1,0}$ and halo mass are 0.09 and -0.35, respectively. The samples combined give the correlation as shown in the left panel. Central galaxies find brighter neighbours that are (often central) galaxies in neighbouring haloes, while for the satellites mostly their own central galaxy is found as neighbour. We expect that the correlation between halo mass and environment is predominantly caused by galaxies finding satellites in their own halo as possible neighbours. Excluding these satellites should result in a much weaker correlation. We postpone such an analysis for future work.

We have verified that for a sample in which the neighbours of galaxies are defined as the nearest brighter galaxy that itself has no brighter neighbour at smaller distance (so it is not itself a satellite of that other galaxy) results in a very low correlation coeffecient between halo mass and $d_{N,m}$ for the satellites too. In this case, a satellite galaxy usually finds its own central as a neighbour (unless there is another satellite that is brighter and closer to that the galaxy you are looking at than to its central) and central galaxies find the nearest brighter other central galaxy. A combined sample of all centrals and satellites then still shows a correlation coefficient of ~ -0.4, as the centrals and satellites show the same bimodal behaviour as shown in the middle and right panels of Fig. 3.8.

Splitting the sample first in a sample of satellites and centrals and excluding the central galaxy of the galaxy's own host halo would probably result in a weaker correlation for the sample as a whole. This could be done by defining a virial radius for each galaxy (based, for example, on its $K-$band luminosity) and identify satellites by searching for galaxies that fall within the virial radius of another, more luminous, galaxy. These can then be flagged as satellites. A neighbour search for the satellites should then exclude a region as large as the virial radius of their central, in order to be sure that the central galaxy in a neighbouring halo is selected as neighbour. This would significantly complicate the neighbour search and we will postpone this for future work.

3.6 Conclusions

The properties of observed galaxies and dark matter haloes in simulations depend on their environment. The term "environment" has, however, been used to describe a wide variety of measures that may or may not correlate with each other. Useful measures of environment include, for example, the distance to the N^{th} nearest neighbour, the number density of objects within some distance, or, for the case of galaxies, the mass of the host dark matter halo. In this paper we carried out a detailed investigation of several environmental parameters which are popular in the (observational) literature, focusing in particular on their relationship with halo mass.

We measured the environmental indicators from the synthetic galaxy catalogues produced using the semi-analytic models by De Lucia & Blaizot (2007), built on the Millennium Simulation (Springel et al., 2005). This model reproduces the number density and clustering properties of observed galaxies in the low-redshift Universe.

We showed that it is of crucial importance to realise that the degree to which environmental parameters measure host dark matter halo mass is determined by (1) whether the scale out to which the environment is measured scales with some typical scale (e.g. the virial radius) of the galaxy in question and (2) whether or not the minimum mass/luminosity that the neighbours are required to have is fixed in absolute terms or relative to the mass/luminosity of the galaxy in question. Specifically, we found that

1. All frequently used environmental indicators (i.e. some function of the distance to the N^{th} nearest neighbour or the number of galaxies within some given distance, either using three dimensional distances or using projected distances for all galaxies within some radial velocity difference) correlate strongly with halo mass.

2. For the number of galaxies within a given distance, n_r, the correlation with halo mass peaks for distances of 1.5–2 virial radii. The virial radius is for observers in general a difficult quantity to measure, but the correlation with halo mass is nearly as strong for galaxy counts within ~ 1 Mpc.

3. The strength of the anti-correlation between the distance to the N^{th} nearest neighbour, r_N, and halo mass is nearly constant for $N \geq 2$ and only slightly weaker for $N = 1$. The relation between r_N and halo mass is slightly weaker than for n_r if r is taken to be similar to the virial radius.

4. Both n_r and r_N correlate more strongly with halo mass if the neighbours are required to be more luminous or massive.

We have shown that it is possible to construct environmental parameters that are insensitive to halo mass by using only dimensionless quantities. For the case of dark matter haloes in numerical simulations this can for for example be achieved by scaling the distance out to which environment is measured to the viral radius of the halo for which the environment is determined and by scaling the minimum required mass to that of the halo in question. The correlation with halo mass becomes smaller if the minimum mass required for neighbours is lower. If the neighbours are more massive than the halo for which the environment is measured, then scaling the distance to the neighbour's virial radius gives more intuitive results and lead to only a slight increase in the strength of the correlation with halo mass. These environmental parameters are, however, only insensitive to halo mass for haloes that are not on the exponential tail of the mass function.

For observers, usually only a position on the sky, some rough indication of the distance along the line of sight and the flux or luminosity in some waveband are available. We showed that analogous environmental measures that are highly insensitive to halo mass can also be constructed using only the K-band luminosities, projected distances on the sky, and a maximum radial velocity difference for neighbours. Specifically, the parameter $d_{1,0}$, defined as the projected distance to the nearest brighter galaxy within a radial velocity difference of 1000 km s^{-1} (that itself does not have a brighter neighbour closer by and therefore probably is a central galaxy of a halo) divided by the K-band luminosity of the neighbour to the power one third, correlates only very weakly with host halo mass.

In summary, when measuring environments for (virtual) observations, we advise to make use of both a halo mass independent measure and a measure that is highly sensitive to halo mass. For purely theoretical studies the halo mass is already known and we therefore advise to use an environmental parameter that is insensitive of halo mass. The following parameters are good choices:

- *Insensitive to halo mass; for simulations:* The distance to the nearest (main) halo that is at least f times more massive than the halo in question, divided by the virial radius of that neighbour. The choice $f = 1$ works well, but if resolution permits it, smaller values yield even weaker correlations with halo mass. Dividing instead by the virial radius of the halo itself gives a slightly weaker correlation with halo mass, at the expense of losing the intuitive definition in which the environment relates to the tidal field due to the neighbour.

- *Sensitive to halo mass; for observations:* The number of brighter galaxies within a projected distance of $\sim 1\ h^{-1}$Mpc, within a redshift window corresponding to $\Delta v \lesssim 1000$ km s^{-1}($n_{1\ \mathrm{Mpc/h}}$). Even better would be to subsequently iterate the following two steps until the procedure converges: (i) check what the corresponding halo masses are using the relations between n_r and halo mass given in Appendix 3.6; (ii) adapt the maximum projected distance to 1.5 times the typical virial radius of the haloes in the sample.

- *Insensitive to halo mass; for observations:* The parameter $d_{1,0}$, as given by Eq. 3.2. The correlation with halo mass is weaker if satellites of the galaxy in question are excluded. This may be possible by requiring neighbours to be further away than some minimum distance. It may even be possible to vary this distance with the virial radius of the neighbour, which can be determined using the measure that is very sensitive to halo mass. This is work in progress.

Many studies have measured galaxy properties as a function of both stellar mass and environment. The environmental indicators used by most authors are effectively measures of halo mass. While halo mass is a perfectly valid measure of environment, and may be particularly relevant for satellites, we note that because stellar mass is also expected to correlate strongly with halo mass, these studies may not have separated "internal" and "external" influences as well as one might naively think. The work presented here will enable future observational and theoretical studies to disentangle the effects of halo mass (internal environment) from those of the external environment. This may eventually tell us whether halo mass is the only important driver of the physics governing galaxy evolution.

Appendix A. Obtaining the halo mass from environmental parameters

In this Appendix we provide fitting functions in order to obtain the halo mass from different environmental indicators, for several lower limits on the galaxy luminosity. This luminosity limit holds for both the galaxies the environment is determined for and for the galaxies included in the neighbour search. We will use the projected quantities, as described in Section. 3.3.3, with a maximum radial velocity difference of 1000 km s^{-1} (the fits are not sensitive to this choice) at redshift 0. We show figures corresponding to Fig. 3.2, but without the colour scale and including a polynomial fit that can facilitate future studies that will use the environmental indicators to measure halo mass.

Environmental indicators that are directly obtained from observations

Here we will use environmental parameters that can be obtained directly from observations. In the next section we will describe how a better estimate of halo mass

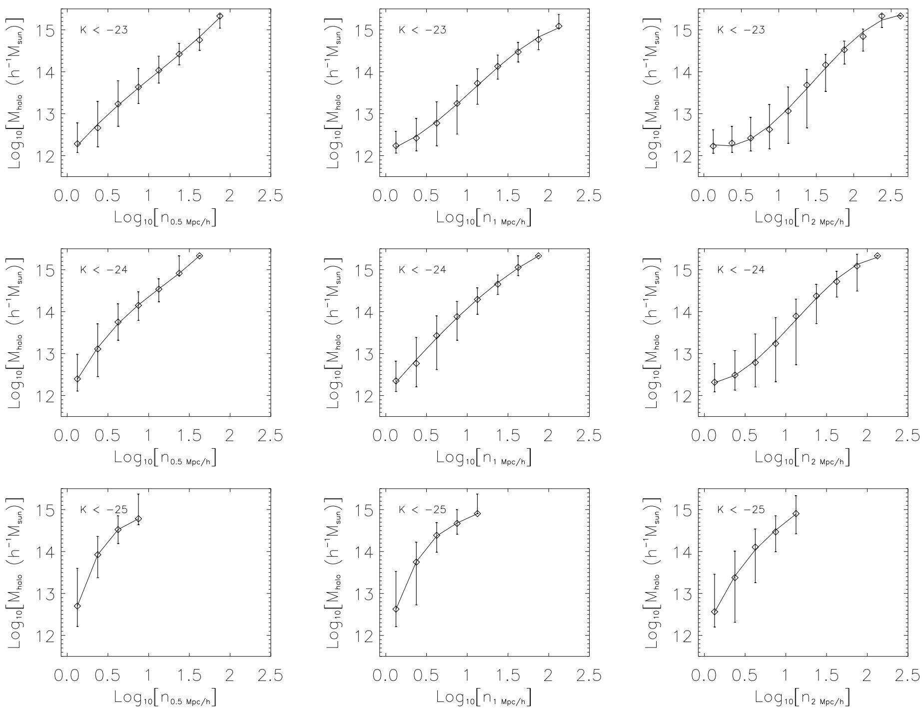

Figure 3.10: Halo mass as a function of three different environmental indicators (corresponding to the columns, $n_{0.5\ \mathrm{Mpc/h}}$ $n_{1\ \mathrm{Mpc/h}}$ and $n_{2Mpc/h}$), for three different lower luminosity limits (corresponding to the rows, $K < \{-23, -24, -25\}$). The symbols are the medians of the data, while the errors represent the 1σ spread (as defined in the text). The solid line is the best fit third order polynomial with coefficients given in Table 3.2.

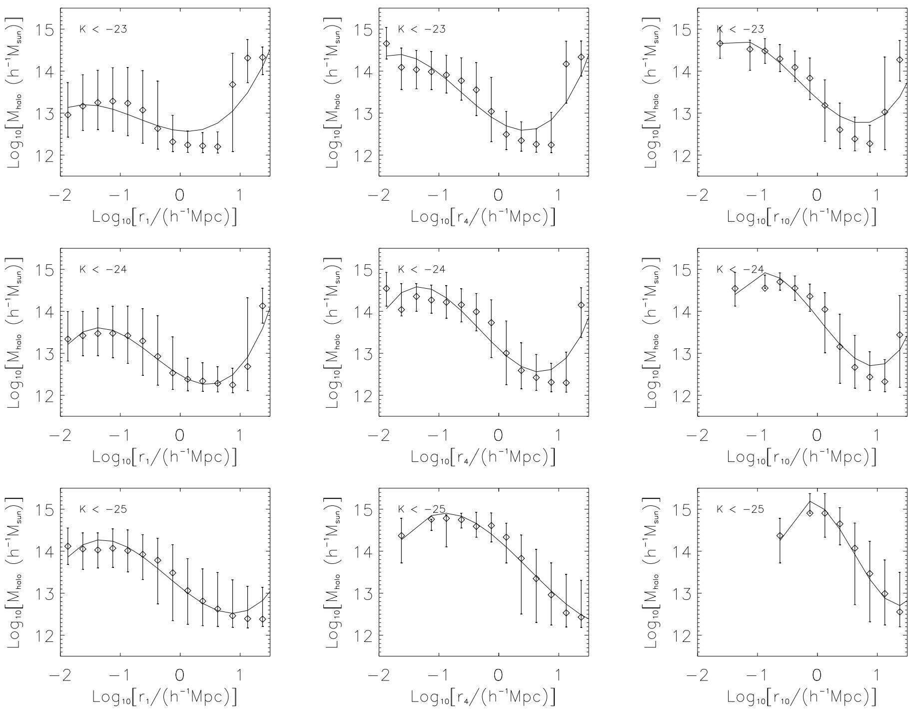

Figure 3.11: Halo mass as a function of three different environmental indicators (corresponding to the columns, r_1 r_4 and r_{10}), for three different lower luminosity limits (corresponding to the rows, $K < \{-23, -24, -25\}$). The symbols are the medians of the data, while the errors represent the 1σ spread (as defined in the text). The solid line is the best fit third order polynomial with coefficients given in Table 3.2.

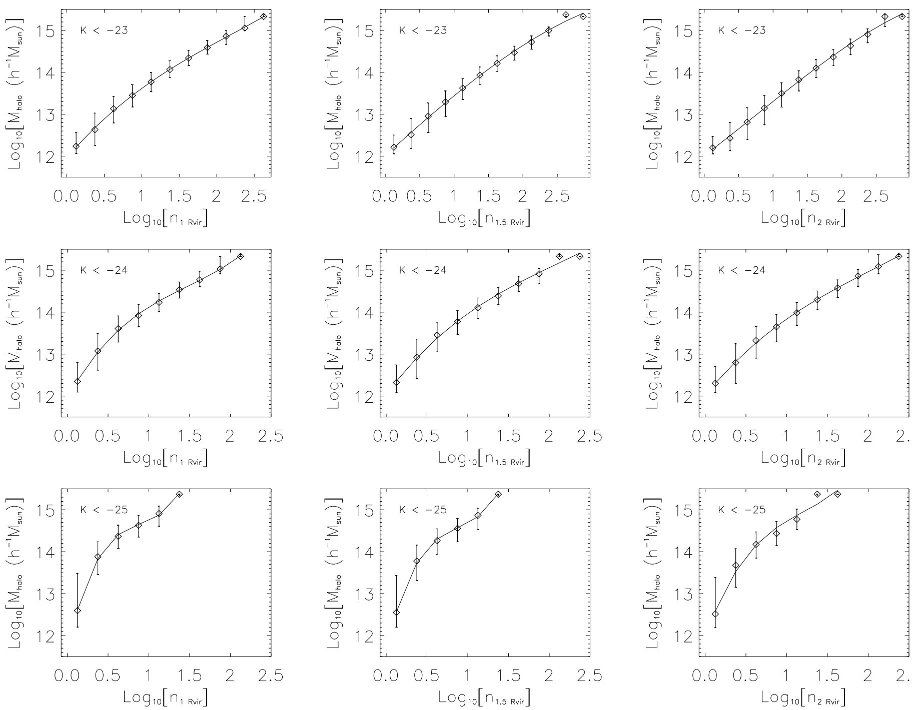

Figure 3.12: Halo mass as a function of three different environmental indicators (corresponding to the columns, $n_{1\ \mathrm{Rvir}}$ $n_{1.5\ \mathrm{Rvir}}$ and n_{2Rvir}), for three different lower luminosity limits (corresponding to the rows, $K < \{-23, -24, -25\}$). The symbols are the medians of the data, while the errors represent the 1σ spread (as defined in the text). The solid line is the best fit third order polynomial with coefficients given in Table 3.2.

can be obtained iteratively. We provide the parameters corresponding to third order polynomial fits for the halo mass as function of the environmental indicators. We fit a function of the form

$$\log M_{\mathrm{halo}} = (\log M_{\mathrm{halo}})_0 + AP + BP^2 + CP^3 \tag{3.3}$$

Where P indicates the logarithm of the environmental parameter in question. We fit on the medians in bins separated by $\Delta P = 0.25$ for all indicators.

The fitted values for the normalization $\log(M_{\mathrm{halo}})_0$ and the three other polynomial coefficients are (A, B, C) are given in Table 3.2 for six different environmental parameters ($n_{0.5\ \mathrm{Mpc}/h}$, $n_{1\ \mathrm{Mpc}/h}$, $n_{2\ \mathrm{Mpc}/h}$, r_1, r_4 and r_{10}) and for six different upper magnitude limits ($K = \{-23, -23.5, -24, -24.5, -25, -25.5\}$). Similarly, we fit the

(1σ) spread in halo mass at fixed environment:

$$\sigma(\log M_{\rm halo}) = \sigma(\log M_{\rm halo})_0 + \alpha P + \beta P^2 + \gamma P^3 \tag{3.4}$$

Note that the distribution is not perfectly Gaussian, nor symmetric, so as a 1σ error we use $\sigma = (p_{84} - p_{16})/2$, where $p_{84,16}$ are the 84'th and 16'th percentile of the distribution. The fit parameters are also given in Table 3.2. The halo mass for a given environment can then be estimated from observational data sets using Eq. 3.3, with the uncertainty given by Eq. 3.4. For completeness, the final column of Table 3.2 indicates the Spearman rank correlation coefficient between the halo mass and the environmental indicator in question for the sample in question.

Similar fits can be requested at the author for different filters used for the selection, different redshifts, different environmental parameters and/or different flux limits.

In Fig. 3.10 we show some of the relations between environment, parametrized by n_r, and halo mass for three different values of r and for three different samples with different lower luminosity limits. The symbols are the medians used in the fits, and the error bars are the 1σ spreads of the data. The solid line is the best fit third order polynomial for which the coefficients are given in Table 3.2.

Fig. 3.11 shows the same, but now for the environment parametrized by r_N for three values of the rank N. Note that these distributions are bimodal as shown in Fig. 3.2, so the correlation with halo mass is in general slightly weaker.

For the samples with a very high flux limit the fits are based on a limited number of galaxies and bins, and are therefore more uncertain. We do not expect that the brightest flux limits quoted here are used for low redshift studies.

A better halo mass estimator

As we have shown in Section 3.3.2 the strongest correlation between halo mass and environment is obtained whenever galaxies are counted within a distance that scales with the virial radius of the halo. In order to to do so, an estimate of the halo mass is necessary. Using the relations described earlier in this Appendix, from the observable environmental indicators an estimate of the halo mass can be made. Using

$$R_{\rm vir} = 0.27\, h^{-1}{\rm Mpc} \left(\frac{M_{\rm halo}}{10^{12} M_\odot}\right)^{1/3} \frac{1}{1+z}, \tag{3.5}$$

which is the same relation as used in the rest of the paper to obtain virial radii, an estimate for the virial radius can be obtained. z is the redshift, which is zero throughout this paper.

Table 3.2: The coefficients of third order polynomial fits to the halo mass as a function of six different environmental indicators which can be obtained directly from observations above a flux limit (indicated in the first column, fit coefficients in columns 2-5), as well as coefficients of third order polynomial fits to the spread of the data (columns 6-9) and the Spearman rank correlation coefficient between halo mass and the environmental indicator (final column). The symbols are as defined in Eqs. 3.3 and 3.4.

$P = \mathrm{Log}_{10}[n_{0.5\ \mathrm{Mpc}}]$

Maximum K	$(\log M_{\mathrm{halo}})_0$	A	B	C	$\sigma(\log M_{\mathrm{halo}})_0$	α	β	γ	$S(M_{\mathrm{halo}}, P)$
-23	12.0	2.34	-0.70	0.21	0.52	2.17	-2.66	0.78	0.65
-23.5	12.0	2.83	-1.00	0.26	0.52	2.70	-3.84	1.33	0.61
-24	11.9	3.97	-2.15	0.61	0.86	0.98	-1.57	0.41	0.54
-24.5	11.9	5.60	-3.77	0.97	0.92	2.31	-5.53	2.79	0.44
-25	11.8	8.48	-8.40	3.02	1.51	-0.63	-3.33	3.47	0.32
-25.5	12.0	128.00	128.00	0.00	2.38	-6.00	4.00	8.00	0.20

$P = \mathrm{Log}_{10}[n_{1\ \mathrm{Mpc}}]$

Maximum K	$(\log M_{\mathrm{halo}})_0$	A	B	C	$\sigma(\log M_{\mathrm{halo}})_0$	α	β	γ	$S(M_{\mathrm{halo}}, P)$
-23	12.1	0.53	1.15	-0.35	0.16	2.84	-2.71	0.67	0.71
-23.5	12.1	1.30	0.56	-0.21	0.32	2.60	-2.52	0.59	0.65
-24	12.0	2.25	-0.14	-0.06	0.50	2.63	-3.05	0.82	0.58
-24.5	11.8	5.06	-3.99	1.56	0.61	4.12	-6.91	2.76	0.49
-25	11.8	7.45	-6.78	2.32	1.02	3.78	-9.57	5.16	0.38
-25.5	12.0	64.00	128.00	0.00	2.62	-4.00	0.00	12.00	0.25

$P = \mathrm{Log}_{10}[n_{2\ \mathrm{Mpc}}]$

Maximum K	$(\log M_{\mathrm{halo}})_0$	A	B	C	$\sigma(\log M_{\mathrm{halo}})_0$	α	β	γ	$S(M_{\mathrm{halo}}, P)$
-23	12.4	-1.04	2.06	-0.47	0.20	1.88	-1.05	0.11	0.63
-23.5	12.3	-0.77	2.11	-0.53	0.16	2.45	-1.53	0.20	0.58
-24	12.3	-0.20	2.05	-0.61	0.30	2.69	-1.86	0.26	0.52
-24.5	12.2	1.47	0.78	-0.31	0.41	4.09	-4.13	0.91	0.45
-25	12.0	4.71	-2.86	0.84	0.56	7.28	-14.15	7.07	0.38
-25.5	12.7	6.72	-9.81	5.63	2.00	-2.41	-2.15	4.63	0.28

$P = \mathrm{Log}_{10}[r_1\ (h^{-1}\mathrm{Mpc})]$

Maximum K	$(\log M_{\mathrm{halo}})_0$	A	B	C	$\sigma(\log M_{\mathrm{halo}})_0$	α	β	γ	$S(M_{\mathrm{halo}}, P)$
-23	12.6	-0.15	0.57	0.26	1.17	-0.35	-0.02	0.06	-0.47
-23.5	12.5	-0.50	0.63	0.37	1.13	-0.41	-0.02	0.10	-0.56
-24	12.5	-0.85	0.59	0.45	1.00	-0.49	0.25	0.29	-0.56
-24.5	12.7	-1.07	0.51	0.46	1.12	-0.32	0.10	0.19	-0.50
-25	13.1	-1.16	0.21	0.33	1.38	0.08	-0.18	-0.01	-0.43
-25,5	14.1	-0.70	-0.30	0.08	1.23	1.04	-0.06	-0.37	-0.29

$P = \mathrm{Log}_{10}[r_4\ (h^{-1}\mathrm{Mpc})]$

Maximum K	$(\log M_{\mathrm{halo}})_0$	A	B	C	$\sigma(\log M_{\mathrm{halo}})_0$	α	β	γ	$S(M_{\mathrm{halo}}, P)$
-23	12.8	-0.84	0.70	0.37	1.04	-0.07	-0.07	0.02	-0.67
-23.5	12.9	-1.03	0.64	0.41	1.10	0.17	-0.12	-0.06	-0.61
-24	13.1	-1.35	0.48	0.49	1.05	0.18	-0.12	-0.07	-0.52
-24.5	13.5	-1.48	0.20	0.43	1.05	0.54	-0.05	-0.13	-0.42
-25	14.3	-1.24	-0.37	0.24	1.01	0.95	-0.01	-0.37	-0.35
-25.5	14.9	0.26	-2.42	1.02	0.58	0.26	2.36	-1.35	-0.20

$P = \mathrm{Log}_{10}[r_{10}\ (h^{-1}\mathrm{Mpc})]$

Maximum K	$(\log M_{\mathrm{halo}})_0$	A	B	C	$\sigma(\log M_{\mathrm{halo}})_0$	α	β	γ	$S(M_{\mathrm{halo}}, P)$
-23	13.3	-1.30	0.39	0.42	0.99	0.47	-0.06	-0.16	-0.62
-23.5	13.6	-1.63	0.21	0.56	0.93	0.48	-0.10	-0.14	-0.53
-24	13.9	-1.80	-0.06	0.70	0.89	0.71	0.06	-0.25	-0.43
-24.5	14.6	-1.89	-1.27	1.22	1.01	0.52	0.09	-0.19	-0.34
-25	15.1	-0.83	-2.67	1.45	0.97	0.42	0.31	-0.15	-0.28
-25.5	10.6	10.40	-9.15	2.34	2.79	-4.01	4.05	-1.20	-0.13

Table 3.3: The coefficients of third order polynomial fits to the halo mass as a function of three different environmental indicators for which a good estimate of the virial radius is needed, above a flux limit (indicated in the first column, fit coefficients in columns 2-5), as well as coefficients of third order polynomial fits to the spread of the data (columns 6-9) and the Spearman rank correlation coefficient between halo mass and the environmental indicator (final column). The symbols are as defined in Eqs. 3.3 and 3.4.

$P = \mathrm{Log}_{10}[n_{1\ \mathrm{Rvir}}]$									
Maximum K	$(\log M_{\mathrm{halo}})_0$	A	B	C	$\sigma(\log M_{\mathrm{halo}})_0$	α	β	γ	$S(M_{\mathrm{halo}}, P)$
-23	12.0	2.07	-0.50	0.08	0.61	0.04	-0.14	0.02	0.85
-23.5	12.0	2.68	-1.00	0.20	0.65	0.15	-0.32	0.07	0.81
-24	12.0	3.52	-1.75	0.40	0.85	-0.43	0.17	-0.06	0.74
-24.5	12.0	4.38	-2.42	0.56	1.15	-1.32	1.03	-0.34	0.63
-25	11.7	8.66	-8.91	3.33	1.76	-4.44	5.51	-2.31	0.49
-25.5	12.5	8.96	-14.37	8.92	3.06	-14.13	26.25	-15.34	0.31
$P = \mathrm{Log}_{10}[n_{1.5\ \mathrm{Rvir}}]$									
Maximum K	$(\log M_{\mathrm{halo}})_0$	A	B	C	$\sigma(\log M_{\mathrm{halo}})_0$	α	β	γ	$S(M_{\mathrm{halo}}, P)$
-23	12.0	1.50	-0.05	-0.02	0.50	0.42	-0.40	0.07	0.86
-23.5	12.0	1.94	-0.34	0.03	0.58	0.35	-0.41	0.07	0.82
-24	12.0	2.59	-0.79	0.13	0.75	0.10	-0.31	0.06	0.75
-24.5	11.9	4.46	-2.74	0.72	1.13	-1.30	1.15	-0.39	0.66
-25	11.7	8.20	-8.34	3.16	1.64	-3.72	4.65	-2.02	0.53
-25.5	12.9	4.59	-3.71	1.40	2.84	-11.68	19.69	-10.26	0.34
$P = \mathrm{Log}_{10}[n_{2\ \mathrm{Rvir}}]$									
Maximum K	$(\log M_{\mathrm{halo}})_0$	A	B	C	$\sigma(\log M_{\mathrm{halo}})_0$	α	β	γ	$S(M_{\mathrm{halo}}, P)$
-23	12.0	1.24	0.09	-0.04	0.44	0.59	-0.48	0.08	0.86
-23.5	12.0	1.71	-0.20	0.01	0.53	0.56	-0.53	0.10	0.81
-24	12.0	2.39	-0.68	0.11	0.73	0.15	-0.27	0.05	0.75
-24.5	12.1	3.15	-1.18	0.21	0.94	-0.11	-0.25	0.04	0.67
-25	12.0	5.38	-3.62	1.01	1.36	-1.51	0.93	-0.33	0.56
-25.5	12.7	6.42	-7.77	3.74	2.79	-11.36	19.20	-10.01	0.37

A better estimate of the halo mass can then be found by measuring the projected number of neighbours within a given multiple of the virial radius (with the same cut in radial velocity difference), as shown in Section 3.3.3. In Table 3.3 we provide the same third order polynomial fits as in Table 3.2, but for the relation between halo mass and $n_{1\ \mathrm{Rvir}}$, $n_{1.5\ \mathrm{Rvir}}$ an $n_{2\ \mathrm{Rvir}}$, as well as the corresponding (higher) Spearman rank correlation coefficients. Fig. 3.12 shows the relations for a selection of the fits.

This procedure of obtaining a better estimate for the halo mass can then be used to iterate towards a reliable estimate for the halo mass, including the spread in halo masses at fixed environment (note that this spread is very small for high mass haloes if the neighbours are counted within a multiple of the virial radius of order one.)

We note that these halo masses are measured in the Millennium Simulation, which uses the WMAP first year results for the cosmology, which has (among other differences) a larger amplitude of fluctuations (σ_8). This means that for a given galaxy luminosity, the haloes will be slightly too massive. How this affects the relations between environment and halo mass is not clear.