Only marginal alignment of disc galaxies

René Andrae* and Knud Jahnke

Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany

Accepted 2011 August 11. Received 2011 August 10; in original form 2011 June 10

ABSTRACT

Testing theories of angular-momentum acquisition of rotationally supported disc galaxies is the key to understanding the formation of this type of galaxies. The tidal-torque theory aims to explain this acquisition process in a cosmological framework and predicts positive autocorrelations of angular-momentum orientation and spiral-arm handedness, i.e. alignment of disc galaxies, on short distance scales of 1 Mpc $h^{-1}$. This disc alignment can also cause systematic effects in weak-lensing measurements. Previous observations claimed discovering these correlations but are overly optimistic in the reported level of statistical significance of the detections. Errors in redshift, ellipticity and morphological classifications were not taken into account, although they have a significant impact. We explain how to rigorously propagate all the important errors through the estimation process. Analysing disc galaxies in the Sloan Digital Sky Survey (SDSS) data base, we find that positive autocorrelations of spiral-arm handedness and angular-momentum orientations on distance scales of 1 Mpc $h^{-1}$ are plausible but not statistically significant. Current data appear not good enough to constrain parameters of theory. This result agrees with a simple hypothesis test in the Local Group, where we also find no evidence for disc alignment. Moreover, we demonstrate that ellipticity estimates based on second moments are strongly biased by galactic bulges even for Scd galaxies, thereby corrupting correlation estimates and overestimating the impact of disc alignment on weak-lensing studies. Finally, we discuss the potential of future sky surveys. We argue that photometric redshifts have too large errors, i.e. PanSTARRS and LSST cannot be used. Conversely, the EUCLID project will not cover the relevant redshift regime. We also discuss the potentials and problems of front-edge classifications of galaxy discs in order to improve the autocorrelation estimates of angular-momentum orientation.

Key words: methods: data analysis – methods: statistical – galaxies: general.

1 INTRODUCTION

Disc galaxies constitute a substantial part of the galaxy population in the nearby Universe (Bamford et al. 2009). As these galaxies are rotationally supported, it is of vital importance to understand how disc galaxies acquire their angular momentum. The tidal-torque theory aims to explain this angular-momentum acquisition through tidal shearing from the dark matter host halo’s gravitational field and the moment of inertia of the forming protogalaxy (for a recent review see Schäfer 2009). This theory predicts alignment effects of disc galaxies, since angular-momentum acquisition is partially governed by environmental effects such that neighbouring disc galaxies residing in the same environment should exhibit similar angular momenta. Hence, testing intrinsic alignments of angular momenta of disc galaxies provides a fundamental test for our understanding of galaxy formation in the cosmological framework. Apart from enhancing our understanding of disc-galaxy formation, investigating these alignment effects is also important because they constitute a potentially significant systematic effect in weak-gravitational-lensing surveys (e.g. Catelan, Kamionkowski & Blandford 2001; Crittenden et al. 2001).

For this goal, we use autocorrelation estimates of spiral-arm handedness and galactic angular-momentum-orientation vectors, respectively. We revisit the works by Slosar et al. (2009) and Lee (2011) and explain that these estimates do not take into account all relevant error contributions and are therefore too optimistic in the reported statistical significance. In this article, we explain how to incorporate the relevant error sources and demonstrate their impact on the results. This methodological rigour is also in a general sense highly relevant, since at the frontier of astrophysical research data analysis can otherwise produce misleading results. Typically for methodological studies, the basic principle and the techniques presented here are also applicable to other astrophysical investigations which involve the estimation of spatial two-point correlation functions, for instances, investigations of baryonic acoustic

*E-mail: andrae@mpia-hd.mpg.de
oscillations (BAOs) (e.g. Blake et al. 2011). Although estimators used for investigations of BAOs are usually much more elaborate than the simple estimator we are going to use, our assessment of the impact of (star–galaxy) classification and redshift errors also applies to this setting. However, we also go beyond the purely methodological aspect and discuss the potential of improving the autocorrelation estimates with new surveys in order to eventually obtain astrophysical results. In particular, we discuss the potential of estimating the front edges of disc galaxies via dust extinction in order to improve the correlation estimates of angular-momentum orientation.

1.1 Strategy

We start in Section 2 by investigating the orientations of angular-momentum vectors in the Local Group. This is meant as an exercise, motivating the necessity of correlation functions. We then present in Section 3 the details, and the selection criteria, of the data samples we are using. In Section 4, we explain how to obtain correlation estimates and their corresponding error estimates for both handedness and angular-momentum-orientation vectors. The main body of this article is Section 5 in which we explain the difference between conditional and marginal errors, discuss the relevant error contributions, and explain how to propagate errors numerically by simple Monte Carlo sampling. In that section, we estimate marginal autocorrelations of handedness and angular-momentum orientations, respectively. This is also the section relevant to readers who are interested in marginal estimates of correlation functions in general, e.g. in the context of BAOs. In an attempt to improve the statistical significance of our results by replacing isophotal ellipticity estimates by less noisy estimators, we show in Section 6 that ellipticities based on second moments are strongly biased. We clearly demonstrate that this bias corrupts correlation estimates. We outline the possible improvements in and potential of future sky surveys in Section 7. We discuss our final results and conclude in Section 8.

2 ARE ANGULAR MOMENTA RANDOMLY ORIENTED IN THE LOCAL GROUP?

As we are investigating the alignment of angular momenta of disc galaxies, the Local Group is a natural first test bed. Apart from numerous dwarf galaxies, the Local Group consists of four disc galaxies, namely the Milky Way, Andromeda (M31), M33, and the Large Magellanic Cloud (LMC), all with pairwise distances of less than 1 Mpc.

2.1 Angular-momentum orientation of the Milky Way

We start by estimating the angular-momentum-orientation vector of the Milky Way in equatorial coordinates. In order to estimate the angular-momentum-orientation vector of the Milky Way, we need two ingredients:

(i) the unit vector \( \mathbf{r}_\odot \) pointing from the Galactic Centre to the position of the Sun;
(ii) the unit vector \( \mathbf{v}_\odot \) of the Sun’s velocity on its trajectory around the Galactic Centre.

1 The perfect BAO estimator would be a generative model that for every galaxy predicts the redshift and star–galaxy classification probability based on the observation conditions. This would enable us to directly take into account these error sources.

Only marginal alignment of disc galaxies

Given the valid assumption that the Sun lies inside and is corotating with the Galactic disc, we can then compute the Milky Way’s angular-momentum-orientation vector via

\[
\mathbf{L}_{\text{MW}} = \mathbf{r}_\odot \times \mathbf{v}_\odot. \tag{1}
\]

We can infer \( \mathbf{r}_\odot \) from the equatorial coordinates of the Galactic Centre, \( \alpha_{\text{GC}} \approx 266.42 \) and \( \delta_{\text{GC}} \approx -29.01 \). Here, we have to keep in mind the fact that these coordinates are pointing from the Sun towards the Galactic Centre, i.e. \( \mathbf{r}_\odot \) is the inverted direction,

\[
\mathbf{r}_\odot = -\begin{pmatrix}
\cos \alpha_{\text{GC}} \sin(90^\circ - \delta_{\text{GC}}) \\
\sin \alpha_{\text{GC}} \sin(90^\circ - \delta_{\text{GC}}) \\
\cos(90^\circ - \delta_{\text{GC}})
\end{pmatrix}. \tag{2}
\]

The unit vector \( \mathbf{v}_\odot \) has to be inferred from the rotation of the Galactic disc. By definition, \( \mathbf{v}_\odot \) points into the direction specified by Galactic longitude \( \ell = 90^\circ \) and Galactic latitude \( b = 0^\circ \) (e.g. Brunthaler et al. 2005). In equatorial coordinates, this direction is given by \( \alpha_s \approx 318.00 \) and \( \delta_s \approx 48.33 \), such that

\[
\mathbf{v}_\odot = \begin{pmatrix}
\cos \alpha_s \sin(90^\circ - \delta_s) \\
\sin \alpha_s \sin(90^\circ - \delta_s) \\
\cos(90^\circ - \delta_s)
\end{pmatrix}. \tag{3}
\]

Inserting these values into equation (1), we obtain the following estimate of the angular-momentum-orientation vector of the Milky Way:

\[
\mathbf{L}_{\text{MW}} \approx \begin{pmatrix}
0.86771 \\
0.19878 \\
-0.45560
\end{pmatrix}. \tag{4}
\]

We conduct two cross-checks: first, the two unit vectors \( \mathbf{r}_\odot \) and \( \mathbf{v}_\odot \) should be orthogonal and indeed their scalar product is \( \mathbf{r}_\odot \cdot \mathbf{v}_\odot \approx 0.00097 \ll 1 \). Secondly, \( \mathbf{L}_{\text{MW}} \) by construction should be normal to \( \mathbf{v}_\odot \approx -10.69 \) and \( \delta_{\text{NP}} \approx 41^\circ 27 \). Indeed, the scalar product is \( \mathbf{L}_{\text{MW}} \cdot \mathbf{u}_{\text{NP}} \approx -0.9999992 \), i.e. both vectors are almost perfectly antiparallel.

2.2 Angular-momentum orientations of Andromeda, M33 and the LMC

In order to estimate the angular-momentum orientations of Andromeda, M33 and the LMC, we use the formalism described in Lee (2011) which is based on ellipticity estimates and the assumption of intrinsically circular galactic discs.

2.2.1 Andromeda (M31)

For Andromeda, we adopt an inclination angle of \( 77^\circ \) (Walterbos & Kennicutt 1988) and an orientation angle of \( 38^\circ \) (Walterbos & Kennicutt 1987). Furthermore, dust lanes enable us to identify the front edge of Andromeda’s galactic disc, which is the north-western edge. Our front-edge estimate agrees with the result of Iye & Ozawa (1999) who investigated the reddening of globular clusters as a function of height above the major axis. Given its equatorial coordinates \( \alpha_{\text{M31}} \approx 10^\circ 69 \) and \( \delta_{\text{M31}} \approx 41^\circ 27 \), we can compute the angular-momentum-orientation vector of Andromeda up to its sign. Chemin, Carignan & Foster (2009) published spatially resolved H I spectra of M31, which enable us to infer the disc rotation directly.
Their map of radial velocities directly implies that the north-eastern part is receding from us, whereas the south-western part is rotating towards us. Consequently, the angular-momentum-orientation vector of Andromeda points south-east and away from our own position. Therefore, if we project \( \mathbf{L}_{\text{M31}} \) on to the unit direction vector pointing from the Milky Way towards Andromeda, this projection must be positive. This condition enables us to fully determine the angular-momentum-orientation vector of Andromeda:

\[
\mathbf{L}_{\text{M31}} \approx \begin{pmatrix} -0.08031 \\ -0.79651 \\ 0.59926 \end{pmatrix}.
\]  

\(2.2.2\) Triangulum galaxy (M33)

Concerning M33, we adopt an inclination angle of 49° and an orientation angle of 21° (Corbelli & Schneider 1997). M33 clearly is a right-handed (Z-wise) spiral. This rotational sense agrees with the results of Brunthaler et al. (2005) who observed the proper motion of two H\(_2\)O masers in M33. It also agrees with the results of Putman et al. (2009), who measured the radial-velocity field of Hi gas in M33. Again, this implies that the projections of both the possible front-edge configurations of \( \mathbf{L}_{\text{M33}} \) on to the unit direction vector pointing from the Milky Way towards M33 have to be positive. Unfortunately, M33 does not exhibit dust lanes, such that the front edge remains unknown. This is not surprising since M33 is not as highly inclined as Andromeda such that we are less likely to observe a dust lane. From dust reddening of C-rich AGB stars Cioni et al. (2008) concluded that there is weak evidence that the north-western side of M33 is the front edge. Given its equatorial coordinates \( \alpha_{\text{M33}} \approx 23\,^h 46\,^m \) and \( \delta_{\text{M33}} \approx 30\,^\circ 66\) the angular-momentum-orientation vector of M33 then reads

\[
\mathbf{L}_{\text{M33}} \approx \begin{pmatrix} 0.67170 \\ -0.47655 \\ 0.56721 \end{pmatrix}.
\]  

The front-edge estimate of Cioni et al. (2008) is still rather uncertain (see their fig. 9). However, it is sufficient for this exercise.

\(2.2.3\) Large Magellanic Cloud (LMC)

Concerning the LMC, we adopt an inclination angle of 35° and an orientation angle of 123° (van der Marel & Cioni 2001). Furthermore, van der Marel & Cioni (2001) find clear evidence that the north-eastern side of the disc is the front edge (their fig. 5). The rotational sense of the LMC is right-handed as is evident from the observed velocity fields (e.g. Olsen & Massey 2007). Given its equatorial coordinates \( \alpha_{\text{LMC}} \approx 80\,^h 89\,^m 38\,^s \) and \( \delta_{\text{LMC}} \approx -69\,^\circ 75\,^\prime\,61\) the angular-momentum-orientation vector of the LMC then reads

\[
\mathbf{L}_{\text{LMC}} \approx \begin{pmatrix} -0.29699 \\ -0.46945 \\ -0.83152 \end{pmatrix}.
\]  

\(2.3\) Random orientation

Are the angular-momentum-orientation vectors in the Local Group compatible with the null hypothesis of random orientation? In order to test this, we investigate the distribution of projection values. For the four disc galaxies, only three statistically independent projections can be derived. We choose the projections on to the Milky Way:

(i) \( \mathbf{L}_{\text{MW}} \cdot \mathbf{L}_{\text{M31}} \approx -0.5010 \),
(ii) \( \mathbf{L}_{\text{MW}} \cdot \mathbf{L}_{\text{M33}} \approx +0.2297 \),
(iii) \( \mathbf{L}_{\text{MW}} \cdot \mathbf{L}_{\text{LMC}} \approx +0.0278 \).

Adding further projection values, e.g. \( \mathbf{L}_{\text{M31}} \cdot \mathbf{L}_{\text{M33}} \), would introduce correlations compromising the KS test. Fig. 1 shows the resulting cumulative distribution of projection values for the Local Group. Furthermore, Fig. 1 shows the cumulative distribution for the null hypothesis where all projection values are equally likely. The KS distance is then \( D_{\text{max}} \approx 0.385 \) which yields a \( p \)-value of \( \approx 0.648 \) (Press et al. 2002). Consequently, the null hypothesis of randomly oriented angular-momentum vectors is capable of producing a maximum distance of 0.385 with a probability of 64.8 per cent.

We conclude from this simple hypothesis test that there is no evidence that disc alignment is at work in the Local Group. However, this hypothesis test is rather coarse given the small number of disc galaxies and the fact that it ignores galaxy separations. Hence, a more elaborate investigation naturally leads us to spatial autocorrelation functions estimated from large samples of disc galaxies as the key diagnostic tool for investigations of disc alignment.

\(3\) THE DATA

An autocorrelation analysis of angular momenta requires a survey covering a large area with homogeneous galaxy morphologies in order to (a) select disc galaxies and (b) estimate their three-dimensional angular-momentum-orientation vectors. The best data base for this purpose is the SDSS. We exploit visual morphological classifications from the Galaxy Zoo project and automated classifications from Huertas-Company et al. (2011), enhanced by additional information from the general SDSS data base.

\(3.1\) Galaxy Zoo

Galaxy Zoo (Lintott et al. 2008, 2011; Bamford et al. 2009) is a unique project where the morphology of nearly 900 000 galaxies from the Sloan Digital Sky Survey (SDSS) spectroscopic sample have been classified visually by the Internet community. Each galaxy has been classified multiple times by different Internet users, which provides a probabilistic object-to-class assignment.
Concerning galaxy morphologies, such a probabilistic assignment is more physical than a hard assignment, as has been discussed by Andrae, Melchior & Bartelmann (2010). In detail, the Galaxy Zoo data base provides probabilistic assignments to the following morphological classes:

(i) elliptical, \( p_{\text{ell}}^{GZ} \);
(ii) disc, \( p_{\text{disc}}^{GZ} \);
(iii) edge-on disc, \( p_{\text{edge}}^{GZ} \);
(iv) clock-wise/Z-wise spiral in projection, \( p_{Z}^{GZ} \);
(v) anticlock-wise/S-wise spiral in projection, \( p_{S}^{GZ} \);
(vi) merger, \( p_{\text{mg}}^{GZ} \).

All probabilities that are taken from Galaxy Zoo carry a ‘GZ’ superscript. The normalization is given by

\[
p_{\text{ell}}^{GZ} + p_{\text{disc}}^{GZ} + p_{\text{edge}}^{GZ} + p_{Z}^{GZ} + p_{S}^{GZ} + p_{\text{mg}}^{GZ} = 1.
\]

Land et al. (2008) reported a bias in the handedness classifications, \( p_{Z}^{GZ} \) and \( p_{S}^{GZ} \), where more spiral galaxies are classified as S-wise than as Z-wise.\(^2\) This bias is corrected in an asymmetric, additive fashion by Land et al. (2008) and Slosar et al. (2009) in order to enforce that the proportions of Z-wise and S-wise spirals are equal with regard to the whole sample. In contrast to this, we employ a symmetric, additive bias correction of the form

\[
p_{Z} = p_{Z}^{GZ} + b \quad \text{and} \quad p_{S} = p_{S}^{GZ} - b,
\]

where \( b \) is chosen such that the numbers of Z-wise and S-wise spirals are identical. There are two reasons for this as follows.

(i) The symmetric correction preserves the normalization of equation (8). This is important because in contrast to Slosar et al. (2009) we are handling the Galaxy Zoo results fully probabilistically in our analysis (cf. Section 5.2).

(ii) Demanding that the proportions of Z-wise and S-wise spirals are equal only provides a single condition, such that an asymmetric correction with two biases, \( b_{Z} \) and \( b_{S} \), is not fully constrained and therefore arbitrary.

Our value of \( b \) is 0.0105 and thus similar to Land et al. (2008). Slosar et al. (2009) argued that such a bias can only lead to a constant offset in the handedness autocorrelation function, but it cannot feign a distance-dependent autocorrelation, which is the predicted astrophysical signal.

3.2 Catalogue of Huertas-Company et al. (2011)

Similar to the Galaxy Zoo project, Huertas-Company et al. (2011) performed a morphological classification on the SDSS spectroscopic galaxy sample. There are two important differences with respect to Galaxy Zoo:

(i) The morphological classes are

- (a) elliptical, \( p_{\text{ell}}^{HC} \);
- (b) SO galaxy, \( p_{SO}^{HC} \);
- (c) Sab disc galaxy, \( p_{\text{Sab}}^{HC} \);
- (d) Scd disc galaxy, \( p_{\text{Scd}}^{HC} \).

All probabilities taken from the catalogue of Huertas-Company et al. (2011) carry a superscript ‘HC’.

\(^2\) Land et al. (2008) also used flipped galaxy images and still observed an excess of S-wise over Z-wise spirals in visual classifications. The exact origin of this bias is unknown, though one option considered by Land et al. (2008) is psychological effect.
3.4.1 Handedness sample

First, starting from the Galaxy Zoo sample, we select all galaxies with either $P_{GZ}^E \geq 0.778$ or $P_{GZ}^O \geq 0.8$, which results in 36,999 galaxies. These asymmetric probability thresholds are chosen this way in order to allow for some flexibility in the correction of the handedness bias of $b = 0.0105$.

Secondly, we obtained the $r$-band Petrosian radii from the SDSS Galaxy table, the spectroscopic redshift estimate and its error estimate from the SDSS SpecObjAll table. Actually, all objects in the Galaxy Zoo sample have been selected from the SDSS spectroscopic sample. For reasons unknown to us, we could not find 103 objects in the Galaxy table and another 5106 objects were untraceable in the SpecObjAll table. This leaves us with 31,790 objects with $r$-band Petrosian radius and estimates of spectroscopic redshift and its error.

Thirdly, we remove multiple objects from the sample, i.e. extended galaxies that have been shredded by the SDSS pipeline producing multiple entries of a single object. We automatically removed galaxy pairs whose angular separations were less than 1.5 times the maximum $r$-band Petrosian radius of both galaxies. Furthermore, Slosar et al. (2009) removed another 69 objects through visual inspection. This list has been kindly provided by Anže Slosar such that we are capable of removing these objects, too. This leaves us with 31,621 galaxies.

Finally, we apply the additive and symmetric bias correction of the handedness classifications given by equation (9). Naïvely interpreting any galaxy with $p_{GZ} \geq 0.8 - b$ as Z-wise spiral and any galaxy with $p_{GZ} \geq 0.8 + b$ as S-wise spiral, we end up with 15,083 Z-wise and 15,071 S-wise spirals for a bias correction of $b = 0.0105$. Therefore, our sample is slightly smaller than the one used by Slosar et al. (2009).

3.4.2 Angular-momentum-orientation sample

Based on the catalogue of morphological classifications by Huertas-Company et al. (2011), we select those galaxies with spectroscopic redshifts $0 < z \leq 0.02$ and probability $P_{pec} > 0.5$ to be a galaxy of type Sc or Sd. This leaves us with 4,236 galaxies satisfying these criteria, the same number of objects as reported by Lee (2011). For 25 of these objects we could not find any information in the SDSS data base, i.e. estimates of $r$-band Petrosian radii, Stokes parameters, their errors, and error estimates of spectroscopic redshift are missing. For these objects, we set the spectroscopic redshift error to $10^{-4}$, which is a typical value for this sample. The Petrosian radii are set to zero. Using the automated method described above, we find 20 rogue pairs in this sample. For each pair, we randomly discard one of the two galaxies, such that we are left with a sample of 4,216 Scd galaxies.

3.5 From axial ratio to angular-momentum orientation

The orientation of the angular-momentum-orientation vector has to be inferred from the observed galactic disc by invoking several assumptions. We follow the formalism described, e.g. in Lee (2011),

in order to estimate the angular-momentum-orientation vector from the observed axial ratios, elliptical orientation angles and equatorial coordinates. In fact, we already used this formalism in Section 2.2. If not specified otherwise, we adopt the same correction for disc thickness like Lee (2011), who assumed an intrinsic axial ratio of $p = 0.1$ for Scd galaxies based on Haynes & Giovanelli (1984). For purposes that would come up later in the paper, we note that Heidmann, Heidmann & de Vaucouleurs (1972) compared different estimates of the intrinsic axial ratios and find values between $p = 0.083$ and $0.145$ for Scd galaxies.

4 CORRELATION ESTIMATORS

In this section, we discuss the correlation estimators for angular-momentum orientations and handedness. We also explain how to estimate errors. We start by explaining the general formalism and then specialize in both angular-momentum orientations and handedness.

4.1 Simple correlation estimator

Given two random variates $X$ and $Y$, we want to estimate their correlation $\xi_{XY}$ and its error. If $N$ samples $x_1, x_2, \ldots, x_N$ and $y_1, y_2, \ldots, y_N$ have been drawn from $X$ and $Y$ and are independent and identically distributed, a simple correlation estimator is given by

$$\hat{\xi}_{XY} = \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle = \langle XY \rangle - \langle X \rangle \langle Y \rangle,$$

(14)

where the hat on $\hat{\xi}_{XY}$ indicates an estimator and

$$\langle XY \rangle = \frac{1}{N} \sum_{n=1}^{N} x_n y_n,$$

(15)

$$\langle X \rangle = \frac{1}{N} \sum_{n=1}^{N} x_n \quad \text{and} \quad \langle Y \rangle = \frac{1}{N} \sum_{n=1}^{N} y_n.$$  

(16)

Merely obtaining a value of $\hat{\xi}_{XY}$ via equation (14) alone is not informative in any way. We also need an error estimate for $\hat{\xi}_{XY}$ in order to get a meaningful result. As $\hat{\xi}_{XY}$ is the mean of $(X - \langle X \rangle)(Y - \langle Y \rangle)$, the variance of $\hat{\xi}_{XY}$ is given by the variance of $(X - \langle X \rangle)(Y - \langle Y \rangle)$ divided by $N$. Consequently, we obtain the error estimate

$$\hat{\sigma}(\hat{\xi}_{XY}) = \frac{\sigma((X - \langle X \rangle)(Y - \langle Y \rangle))}{\sqrt{N}}.$$  

(17)

Here we assume that $N$ is large enough such that the likelihood function of the mean $(X - \langle X \rangle)(Y - \langle Y \rangle)$ is approximately Gaussian and we are allowed to take the square-root of the variance in order to obtain a standard deviation $\sigma$.

---

3 The Galaxy Zoo data base provides the SDSS OnID, which was used to identify objects in the Galaxy table. Cross-matching with the SpecObjAll table was done by retrieving the SpecObjID from the Galaxy table or – if this label was unavailable – by matching the given OnID with the BestOnID from the SpecObjAll table.

4 More elaborate estimators can be defined. Equation (14) is the maximum-likelihood estimate, if and only if $(X, Y)$ are drawn from a bivariate Gaussian, i.e. if there are no higher-order correlations.

5 Mark the following important difference: if we are interested in estimating some random variate $Z$, we employ its mean $(Z)$ and its variance $(Z^2) - (Z)^2$. However, in this case we are not interested in estimating $Z$ but in estimating the mean of $Z$ and the variance of $(Z)$ equals the variance of $Z$ divided by the number of samples drawn from $Z$. Loosely speaking, if we draw more samples from $Z$, the distribution of $Z$ does not change, in particular its width (variance) stays constant. However, drawing more samples from $Z$ enables us to estimate the mean of the distribution more accurately.
4.2 Angular-momentum orientation

Our aim is to estimate the scalar two-point autocorrelation function of angular-momentum orientations, \( \hat{\xi}_{LL}(r) \). Here, we assume spherical symmetry such that \( \hat{\xi}(r) = \hat{\xi}(r) \). This is a first-order approximation because the spatial distribution of galaxies in the Universe is not isotropic on short scales (‘Cosmic Web’). Usually, the following estimator is employed (e.g. Pen, Lee & Seljak 2000; Lee 2011):

\[
\hat{\xi}_{LL}(r) = \left\langle p_a p'_b | L_a \cdot L'_b |^2 \right\rangle + \left\langle p_a p'_b | L_b \cdot L'_a |^2 \right\rangle + \left\langle p_b p'_a | L_b \cdot L'_a |^2 \right\rangle - \frac{1}{3},
\]

(18)

where the primes indicate the second galaxy in the pair and the subscripts \( a, b \) denote the two possible orientations of the disc’s front edge with probabilities \( p_a \) and \( p_b \). If the front edge is not estimated, the default values are \( p_a = p_b = \frac{1}{2} \). Introducing the abbreviation \( Z = p_a p'_a (|L_a| |L'_a|^2 + |L_a| |L'_b|^2 + |L_b| |L'_a|^2 + |L_b| |L'_b|^2) \), an error estimate of \( \hat{\xi}_{LL}(r) \) is

\[
\hat{\sigma}^2(\hat{\xi}_{LL}) = \frac{\hat{\sigma}^2(Z)}{N},
\]

(19)

where \( N \) denotes the number of galaxy pairs in the relevant distance bin.

4.3 Handedness

We also want to estimate the two-point autocorrelation function of handedness \( \hat{\xi}_{HH}(r) \). Again assuming spherical symmetry, a general estimator is given by

\[
\hat{\xi}_{HH}(r) = \langle h h' \rangle,
\]

(20)

where we define handedness as

\[
h = p_Z - p_S.
\]

(21)

As explained in Section 3.1, the mean handedness is zero in the whole sample, i.e. \( \langle h \rangle = \langle h' \rangle = 0 \). Handedness alignments cannot change this in individual distance bins if the number of galaxy pairs is large enough. In every distance bin, let \( n_s \) denote the number of galaxy pairs with \( h h' = +1 \) and \( n_\circ \) the number of galaxy pairs with \( h h' = -1 \). We can then rewrite equation (20) to read

\[
\hat{\xi}_{HH}(r) = \frac{n_+ - n_\circ}{n_+ + n_\circ} = f_+ - f_\circ = 2f_+ - 1,
\]

(22)

where \( f_\pm = n_{\pm} / (n_+ + n_\circ) \) denotes the fraction of galaxy pairs with positive or negative handedness products, respectively. An error estimate of \( \hat{\xi}_{HH}(r) \) is obtained from the fact that counting positive handedness products is a Bernoulli trial, i.e. \( n_\pm \) are subject to the binomial distribution while \( f_\pm \) are subject to the beta distribution (e.g. Cameron 2011).

5 THE IMPACT OF ERRORS

This section is dedicated to a detailed investigation of the impact of various error sources on autocorrelation estimates of handedness and angular-momentum-orientation vectors, respectively. As key results, we finally provide marginal estimates of these autocorrelation functions which take into account all relevant error sources. Like Lee (2011), we employ isophotal ellipticity estimates as far as angular-momentum-orientation vectors are concerned. However, our methodological discussion of error propagation is also relevant in a wider context, e.g. concerning correlation functions for investigations of baryonic acoustic oscillations.

5.1 Conditional versus marginal errors

Previous estimates (e.g. Slosar et al. 2009; Lee 2011) employ certain input parameters such as redshift estimates using only their maximum-likelihood values, without propagating the errors of these values. Hence, these estimates are conditional instead of marginal estimates. Consequently, we now need to explain the conceptual difference between conditional and marginal errors.

For the sake of simplicity, let us consider fitting data \( D \) with Gaussian noise using a model with two linear parameters \( \theta_1 \) and \( \theta_2 \). In this case, the likelihood function \( L(D|\theta_1, \theta_2) \) is a bivariate Gaussian also in the linear parameters and its covariance matrix

\[
\Sigma = \begin{pmatrix}
\sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\
\rho_{12}\sigma_1\sigma_2 & \sigma_2^2
\end{pmatrix}
\]

(23)

can be found by a Fisher analysis (e.g. Heavens 2009). Here, \( \sigma_1^2 \) and \( \sigma_2^2 \) denote the variances of \( \theta_1 \) and \( \theta_2 \), whereas \( -1 \leq \rho_{12} \leq 1 \) is the correlation coefficient. \( \sigma_1 \) is the standard deviation of this Gaussian if sliced at the mean value of \( \theta_2 \). Therefore, \( \sigma_1 \) is the conditional error of \( \theta_1 \), ‘conditional’ because it depends on where the Gaussian has been sliced, i.e. the mean value of \( \theta_2 \). Conversely, the marginal error of \( \theta_1 \) is independent of the value of \( \theta_2 \). This marginal error is obtained by projecting the bivariate Gaussian on to the \( \theta_1 \)-axis, instead of slicing it. Marginal errors are never smaller than conditional errors. Consequently, the conditional error \( \sigma_1 \) underestimates the true error on \( \theta_1 \), such that, e.g. statistical significance is overestimated.

5.2 Uncertainties in classifications

The morphological classifications of Galaxy Zoo and Huertas-Company et al. (2011) are probabilistic, i.e. every object is assigned a probability to belong to either of the possible morphological types. This is in contrast to non-probabilistic – ‘hard’ – assignments, where every object is clearly assigned to a certain type. Hard assignments are easier to carry out and interpret, wherefore many astronomers have a natural affinity to this approach. Unfortunately, galaxy morphologies cannot be clearly assigned to morphological types in general – apart from singular prototypical examples of very obvious morphology. The bulk of galaxies has uncertain morphologies, i.e. the morphological types are overlapping such that hard classification schemes are biased (Andrae et al. 2010). For instance, a galaxy with \( p_{ZS} = 0.8 \) still has a 20 per cent chance not to be a Z-spiral – or a disc galaxy at all.7 Discarding the classification uncertainty by introducing a hard cut pretends that the data are more accurate than they appear to be. This inevitably leads us to understate the errors, thereby compromising the estimates of statistical significance.

6 Slosar et al. (2009) derived pseudo-marginal estimates of the handedness autocorrelation function. Although they marginalized their likelihoods, they used conditional input data.

7 The Galaxy Zoo probabilities may exhibit minor biases due to people voting incorrectly out of confusion or malice. However,Lintott et al. (2008) weighted the users depending on how their votes agreed with the majority. Moreover, on average, every galaxy has received 39 votes (Land et al. 2008) such that the impact of deliberate misclassification should give rise to a minor bias only. Certainly, that effect is much smaller than the bias we would obtain if we cut the classification probabilities. In fact, it is very hard to do worse than a discontinuous hard cut. Any reasonable continuous transition between two classes is virtually guaranteed to be a better approximation to reality than a hard cut which corresponds to a discontinuous step in such a two-class transition.
In fact, Slosar et al. (2009) turned the probabilistic assignments of Galaxy Zoo into hard assignments by introducing a hard cut: for the clean sample, every galaxy with $p_S \geq 0.8$ is considered a Z-wise spiral and every galaxy with $p_S \geq 0.8$ is considered a S-wise spiral, while all other galaxies are discarded. Similarly, Lee (2011) considers every galaxy with $p^{HC}_{kcd} > 0.5$ as an Scd galaxy. We explain in Sections 5.7 and 5.9 how to account for these classification uncertainties in estimating the correlation functions of handedness and angular-momentum orientations. There is no reason that enforces such a hard cut.

5.3 Errors in spectroscopic redshift estimates

Both autocorrelation functions introduced in Sections 4.2 and 4.3 require estimates of distances of galaxies pairs, and these distances are uncertain due to errors in the redshift estimates. In order to assess the impact of redshift errors, we randomly select a single galaxy from our SDSS subsample and draw 10,000 Monte Carlo samples from its redshift-error distribution. For every sampled value of redshift, we compute the comoving distance and monitor its distribution. As is evident from Fig. 2, the errors in the comoving distances are of the same order of magnitude as of the typical distance scale of the correlations reported in the literature ($\approx 1\, h^{-1}$ Mpc). Consequently, these errors are important and have to be taken into account. We explain in Section 5.5 how to propagate these redshift errors by Monte Carlo sampling.

5.4 Errors in ellipticity estimates

Errors in ellipticity estimates used as proxies for disc inclination clearly have an impact on the estimation of the angular-momentum orientations and their correlation function. We now try to estimate these errors. We explain in Section 5.5 how to propagate ellipticity errors by Monte Carlo sampling.

First, considering the isophotal ellipticities used by Lee (2011), the SDSS data base unfortunately does not offer error estimates. In fact, the table GALAXY contains columns for the errors of the isophotal ellipticities. However, for the relevant objects these columns are only filled with invalid default values.

Footnote:

8 In fact, the table GALAXY contains columns for the errors of the isophotal ellipticities. However, for the relevant objects these columns are only filled with invalid default values.

9 Here, we assume that for multiple entries the whole galaxy is used for parameter estimation and not only a shredded part of the galaxy.

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{fig2}
\caption{Likelihood function of comoving distance for a galaxy with spectroscopic redshift of $z = (6.5993 \pm 0.0078) \times 10^{-2}$. The likelihood has been estimated by drawing 10,000 Monte Carlo samples from the error distribution of the spectroscopic redshift and is approximately Gaussian with mean $(183.16 \pm 0.20)\, h^{-1}$ Mpc.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{fig3}
\caption{Distributions of differences in isophotal axial ratios (top) and isophotal orientation angles (bottom) for the 1596 rogue pairs in the catalogue of Huertas-Company et al. (2011). The top panel is well approximated by a Gaussian with mean zero and standard deviation of $0.0795$ (dashed line), which yields an error estimate of $\Delta q_{\omega} = 0.0795/\sqrt{2} \approx 0.0562$. The distribution of differences in orientation angles (bottom panel) is not Gaussian, but described by the ad hoc model of equation (26) (dashed line) based on equation (25) with manually adjusted parameters $\hat{\alpha} \approx 0.73$, $\hat{\sigma} \approx 2.7$ and $\hat{\theta}^2 \approx 15.0$.

Consequently, employing isophotal ellipticities, the SDSS data base strictly does not enable us to estimate a marginal autocorrelation function. In order to get a rough estimate of the errors in isophotal ellipticities, we make use of the rogue pairs in the SDSS data base, i.e. multiple entries of identical galaxies. Starting out from 698,420 galaxies in the classification table provided by Huertas-Company et al. (2011), we identify rogue pairs as galaxy pairs whose angular separation is less than 0.4 arcsec, which roughly corresponds to one pixel size. We find 1596 such pairs. We then monitor the difference in axial ratios and orientation angles of every pair. The resulting distributions are shown in Fig. 3. As a rough error estimate for the isophotal axial ratio, we obtain a standard deviation of $\Delta q_{\omega} \approx 0.0562$, when fixing the mean to zero. The distribution of differences in orientation angles is not Gaussian but has more prominent wings. We therefore model the likelihood function of orientation angles with mean angle $\theta_0$ as a mixture of two Gaussians of different width,

$$L(\theta|\theta_0, \sigma_1, \sigma_2, \alpha) = \alpha N(\theta|\theta_0, \sigma_1) + (1-\alpha)N(\theta|\theta_0, \sigma_2).$$

The bottom panel of Fig. 3 displays the distribution of differences of two values drawn from equation (25), whose likelihood is obtained by convolving $L(\theta|\theta_0, \sigma_1, \sigma_2, \alpha)$ with itself. The resulting likelihood function then reads

$$L(\Delta \theta|\sigma_1, \sigma_2, \alpha) = \alpha^2 N(\Delta \theta|0, \sqrt{2}\sigma_1) + 2\alpha(1-\alpha)N\left(\Delta \theta|0, \sqrt{\sigma_1^2 + \sigma_2^2}\right) + (1-\alpha^2)N(\Delta \theta|0, \sqrt{2}\sigma_2).$$

$$L(\Delta \theta|\sigma_1, \sigma_2, \alpha) = \alpha^2 N(\Delta \theta|0, \sqrt{2}\sigma_1) + 2\alpha(1-\alpha)N\left(\Delta \theta|0, \sqrt{\sigma_1^2 + \sigma_2^2}\right) + (1-\alpha^2)N(\Delta \theta|0, \sqrt{2}\sigma_2).$$

© 2011 The Authors, MNRAS 418, 2014–2031

Monthly Notices of the Royal Astronomical Society © 2011 RAS

Downloaded from https://academic.oup.com/mnras/article-abstract/418/3/2014/1066169 on 27 July 2018
Manually adjusting the model parameters of equation (25), we obtain a rough error estimate for the isophotal orientation angle with parameters $\hat{\alpha} \approx 0.73$, $\hat{\sigma}_1 \approx 2.7$ and $\hat{\sigma}_2 \approx 15.0$. If we required an angular separation of 0, i.e. identical coordinates, we would still end up with 17 pairs exhibiting similar scatter in both parameters. Secondly, the correction for intrinsic axial ratios of Scd galaxies is subject to uncertainties, too. Wherever we ignore ellipticity errors, we also ignore errors in intrinsic axial ratios and simply adopt $p = 0.1$. Conversely, if we take into account ellipticity errors, we will automatically also take into account errors in the intrinsic axial ratio. In this case, we assume that $p$ is drawn from a uniform distribution over the interval [0.083, 0.145] (see Section 3.5).

5.5 Propagating errors numerically

We now explain how to incorporate errors in redshift estimates and ellipticity estimates. The crucial problem is that both errors cannot be propagated analytically.

We propagate the measurement errors of spectroscopic redshift and ellipticity by drawing 1000 Monte Carlo realizations from the error distributions of both parameters and averaging the results over all the Monte Carlo realizations.\(^\dagger\) A value for the intrinsic axial ratio is drawn from the uniform interval [0.083, 0.145] once for every Monte Carlo realization, i.e. in each realization all galaxies have the same correction for intrinsic axial ratio. This Monte Carlo sampling is in fact a marginalization over the errors of both observables, spectroscopic redshift and ellipticity. Typically, both the error sources are not taken into account (e.g. Slosar et al. 2009; Lee 2011), which yields correlation estimates with conditional errors – conditional because they assume, e.g. the observed redshifts were the true ones.

A final remark concerning the correlation estimate: we monitor the distribution of the correlation values $\hat{\xi}$ resulting from the 1000 Monte Carlo realizations. However, a fundamental difference from equation (17) is that now $\hat{\xi}$ itself is a random variate. Consequently, we are now interested in the variance of $\hat{\xi}$ but not in the variance of the mean of $\hat{\xi}$. The difference is a factor of 1000 in the variances.

It is obvious that this approach is correct, since otherwise we could make the resulting errors arbitrarily small by increasing the number of Monte Carlo realizations.

5.6 Negligible error sources

There are further sources of errors which could be taken into account but are not relevant in our case.

For instances, uncertainties in the cosmological parameters have an impact on the comoving distances. In our case, this is irrelevant because all galaxies are affected the same way. However, if in a different context the task is to use marginal autocorrelation functions in order to draw a cosmological inference, it may be mandatory to also incorporate uncertainties of cosmological parameters into the Monte Carlo sampling described in Section 5.5. We experienced that this increases the error in comoving distances by approximately a factor of 2.

Another negligible error source is the position estimate of a galaxy in equatorial coordinates, which is obviously much smaller than, e.g. any redshift error. Given the pixel size of $\approx 0.4$ arcsec of SDSS, at a redshift of $z = 0.066$ and a comoving distance of $d = 183 \, \text{Mpc} \, h^{-1}$ a misestimation of 1 pixel corresponds to a transversal error of $0.35 \, \text{kpc} \, h^{-1}$. This is several orders of magnitude below the theoretically expected correlation length of roughly $1 \, \text{Mpc} \, h^{-1}$ (Schäfer & Merkel 2011).

5.7 Impact on autocorrelation of handedness

In this section, we discern the impact of various error sources on the estimates of the handedness autocorrelation function, namely classification uncertainties, redshift errors and number statistics. Our ultimate goal is a marginal estimate of the handedness autocorrelation function, where all errors have been marginalized out.

First, we use the hard estimator from Slosar et al. (2009), which does not account for uncertainties in classification and redshift. Panel (a) of Fig. 4 shows our estimate of this autocorrelation function for the Galaxy Zoo sample. Qualitatively, our results agree

Figure 4. Impact of various errors on estimates of the handedness autocorrelation function. Panel (a): hard estimate ignoring classification uncertainties and redshift errors, taking into account only number statistics. Panel (b): soft estimate accounting for classification uncertainties and number statistics, ignoring redshift errors. Panel (c): estimate accounting for redshift errors and number statistics, ignoring classification uncertainties. Panel (d): marginal estimate taking into account classification uncertainties, redshift errors and number statistics. Furthermore, we show autocorrelation estimates parametrized as exponential and Gaussian according to Fig. 5.
with the results of Slosar et al. (2009). We observe positive correlations, i.e. an alignment of handedness, on short distances, too. Although there are minor differences which might arise from the slightly different data sets used, the general agreement validates our method.

Secondly, we take into account uncertainties in the handedness classifications, but still ignore redshift errors. In every distance bin, we compute the handedness products

\[ h' = (p_2 - p_3)(p_2' - p_3') = p_2 p_3' + p_3 p_2' - p_2 p_3 - p_3 p_2', \]  

which can now take any value in the interval \([-1, 1]\). The correlation estimator of equation (22) remains unchanged. However, \(n\) now are not the number of pairs where \(h' = \pm 1\), but are rather defined by

\[ n_+ = \sum_{\text{pairs}} (p_2 p_3' + p_3 p_2') \quad \text{and} \quad n_- = \sum_{\text{pairs}} (p_2 p_3' + p_3 p_2'). \]  

Note that if \(N\) denotes the number of galaxy pairs in a given distance bin, then \(n_+ + n_- \leq N\). Consequently, this reduces the ‘effective’ number of galaxy pairs in a given distance bin because the contribution of every galaxy pair is downweighted by the probability that either galaxy is not a spiral with handedness. Furthermore, reducing the effective number of galaxy pairs also increases the error of the correlation estimate through the beta distribution (e.g. Cameron 2011). Results of this estimator are shown in panel (b) of Fig. 4. As expected, the error bars are indeed slightly larger.

Thirdly, we account for redshift errors but ignore classification uncertainties. As described in Section 5.5, we draw 1000 Monte Carlo realizations from the error distributions of the spectroscopic redshift estimates and average over all realizations. Panel (c) of Fig. 4 shows the resulting estimate of the handedness autocorrelation. In comparison to panels (a) and (b), the autocorrelation function now looks remarkably smooth. Errors in redshift cause uncertainties in the distances, i.e. galaxy pairs end up in different distance bins in different realizations. Consequently, a likely explanation for all the substructures in panels (a) and (b) is that they are noise features that have been enhanced by binning.

Finally, panel (d) shows the marginal autocorrelation function, which takes into account all the important sources of uncertainty. The error bars are so large that apparently no statistically significant positive correlation of handedness can be detected. However, we have to refine this question in the next section, as we should not attempt to assess statistical significance from binned data.

### 5.8 Parameter estimation

Fig. 4 shows binned versions of the estimated correlation function. This is acceptable as long as we only study the dependence of the error bars on the different error sources. However, in order to assess the statistical significance of positive autocorrelations in the final marginal estimate, we should try to avoid the ambiguities introduced by binning. For this purpose, we employ the likelihood function of the data \(D\) introduced by Slosar et al. (2009):

\[ \mathcal{L}[D|\xi(r)] = \prod_{\text{pairs}} \left( 1 + d_p \frac{\xi(r_p)}{2} \right)^{n_+} \left( 1 - d_p \frac{\xi(r_p)}{2} \right)^{n_-}, \]  

where \(r_p\) is the distance between the two galaxies of the \(p\)th pair. The coefficient \(d_p\) is the handedness product of both galaxies. As Slosar et al. (2009) used hard cuts of the Galaxy Zoo classifications, \(d_p = \pm 1\) in their case. We modify this by equating \(d_p\) with equation (27)

\[ \text{such that now } -1 \leq d_p \leq +1 \text{ and galaxy pairs are weighted by the probability that both of them are spirals.} \]

In order to assess the statistical significance of potential positive autocorrelations in spiral-arm handedness, we follow Slosar et al. (2009) in using the Bayes factor,

\[ \text{prob}(D|M_\text{++}) = \frac{\int \text{prob}(D|\theta_\text{+}, M_\text{++}) \text{prob}(\theta_\text{+}|M_\text{++}) d\theta_\text{+}}{\int \text{prob}(D|\theta_\text{+}, M_\text{++}) \text{prob}(\theta_\text{+}|M_\text{++}) d\theta_\text{+}}. \]  

Here, \(\text{prob}(D|M_\text{++})\) denotes the likelihood of the data \(D\) given the model \(M_\text{++}\), irrespective of what the parameter values \(\theta_\text{+}\) of model \(M_\text{++}\) are. Conversely, \(\text{prob}(D|\theta_\text{+}, M_\text{++})\) denotes the likelihood of the data given the model \(M_\text{++}\) and certain parameter values \(\theta_\text{+}\), while \(\text{prob}(\theta_\text{+}|M_\text{++})\) denotes the prior probability of the parameter values \(\theta_\text{+}\) of model \(M_\text{++}\).

In our case, the model \(M_0\) describes the null hypothesis that no autocorrelation exists, i.e. \(\xi(r) = 0\). This model has no free parameters, such that we can directly evaluate \(\text{prob}(D|M_0)\) via equation (29). Conversely, the model \(M_{\text{++}}\) is supposed to describe positive autocorrelations. Here, we have to make a choice of how we parametrize such positive autocorrelations. Like Slosar et al. (2009), we then employ two parametrizations, an exponential and a Gaussian,

\[ \xi_{\text{exp}}(r) = a e^{-r/b} \quad \text{and} \quad \xi_{\text{Gauss}}(r) = a e^{-r^2/2b^2}, \]  

with model amplitudes \(a\) and model correlation lengths \(b\). For both models, we use flat and normalized priors within the intervals \(a \in (0, 1)\) and \(b \in (0, 3)\). In contrast to Slosar et al. (2009), we also exclude \(a = 0\) in order to ensure that \(M_{\text{++}}\) and \(M_0\) are indeed mutually exclusive. As both parametrizations introduced in equation (31) have two free parameters, \(a\) and \(b\), we cannot evaluate \(M_{\text{++}}\) directly. Rather, we compute the likelihood manifolds of \(a\) and \(b\) for both models using a brute-force grid. Fig. 5 shows the resulting likelihood manifolds averaged over the 1000 noise realizations drawn from the redshift-error distribution. Our results look very similar to those shown in fig. 4 of Slosar et al. (2009) and our most likely values agree nicely with their values. The best-fitting estimates for both models are also shown in panel (d) of Fig. 4. Given the brute-force likelihood grid \(\mathcal{L}_{ij} = \mathcal{L}(a_i, b_j)\), the marginalization integral

\[ 11 \]  

If we assume that both models, \(M_{\text{++}}\) and \(M_0\), are equally likely a priori, i.e. if we have no a priori preference, then the Bayes factor is identical to the ratio of model posteriors, which quantify the probability of the model given the data.
in equation (30) can be approximated by a Riemann sum,
\[
\int_{0}^{1} \int_{0}^{3} dh \, \text{prob}(D|a, b, M_{+}) \text{prob}(a, b|M_{+}) \\
\approx \sum_{i,j} L(a_{i}, b_{j}) \Delta a_{i} \Delta b_{j} \frac{1}{3} .
\]
(32)
where \(\text{prob}(a, b|M_{+}) = \frac{1}{\Delta} \) is the normalized flat prior on the interval \(a \in (0, 1] \) and \(b \in (0, 3] \), while \(\Delta a = \Delta a_{i} \) and \(\Delta b = \Delta b_{j} \) denote the equidistant stepsizes of the brute-force likelihood grids shown in Fig. 5. This results in Bayes factors of 27.9 for the exponential model\(^\text{12}\) and 13.1 for the Gaussian model, respectively. These values can be interpreted as strong but not yet decisive evidence in favour of positive autocorrelations. Decisive evidence requires Bayes factors larger than 100 (e.g. Kass & Raftery 1993).

5.9 Impact of angular-momentum orientation on autocorrelation

Now, we discern the impact of various error sources on estimates of the autocorrelation function of angular-momentum orientation vectors. Again, our ultimate goal is a marginal estimate of the handness autocorrelation function.

First, we try to reproduce the estimate of angular-momentum-orientation autocorrelation of Lee (2011). The only difference is that we have removed 20 objects from the galaxy sample in order to eliminate rogue pairs. Panel (a) of Fig. 6 shows our resulting estimate of the autocorrelation via equation (18). Our result is identical to the one of Lee (2011). This implies that, first, our method is working correctly, and, secondly, that a few rogue pairs have negligible impact on the results of Lee (2011).

Secondly, we study the impact of uncertainties of morphological classification. Formally, the estimator defined in equation (18) does not change, only the effective number of galaxy pairs in all redshift bins is reduced. Picking one of the four terms in equation (18), we change the definition:
\[
\langle p_{a} p_{b} | L_{a} \cdot L'_{b} \rangle = \frac{\sum_{\text{pairs}} p_{HC}^{\text{Scd}} p_{HC}^{\text{Scd}} p_{a} p_{b} [L_{a} \cdot L'_{b}]^2}{\sum_{\text{pairs}} p_{HC}^{\text{Scd}} p_{HC}^{\text{Scd}}} .
\]
(33)
This weights the contribution of each pair by the probability that both galaxies are Scd galaxies. Furthermore, the number \(N\) of pairs in the distance bin are replaced by the sum of weights \(\sum_{\text{pairs}} p_{HC}^{\text{Scd}} \leq N\). Obviously, this weighting also affects the error estimate of equation (19). Panel (b) of Fig. 6 shows the probabilistic correlation estimate. Evidently, the hard estimator used by Lee (2011) substantially underestimates the errors, thereby overestimating the actual statistical significance. As class probabilities were cut at \(p_{HC}^{\text{Scd}} > 0.5\), classification uncertainties have a larger impact than in the case of handedness where the cut of handedness probabilities was at 0.8.

Thirdly, we incorporate errors in spectroscopic redshift by drawing 1000 Monte Carlo realizations from the redshift’s error distribution. The resulting conditional estimate, now out to 10 Mpc \(h^{-1}\), is shown in panel (c) of Fig. 6. Qualitatively, the impact of redshift errors on the correlation estimate of angular-momentum-orientation vectors is not as severe as in the case of handedness (cf. marginal estimate of Fig. 4). Note that the binsize in Fig. 6 is much larger than in Fig. 4 because here we are studying a smaller sample with fewer galaxy pairs. None the less, the estimated errors have indeed increased, which is particularly obvious for the first distance bin. As the binning is logarithmic in distance, this is not surprising because the first distance bin has the smallest binsize and is thereby strongest affected by redshift errors ‘smearing out’ galaxy pairs along the horizontal axis.\(^\text{13}\)

Finally, we also take into account errors in ellipticity estimates. As mentioned in Section 5.4, the SDSS data base actually does not provide error estimates for the isophotal ellipticities. Hence, we need to proceed using the rough error estimates of equations (24) and (25) as well as the uniform error in intrinsic axial ratios. This means that it is 27.9 times more likely that the data have been drawn from an exponential whose amplitude is somewhere in the range of (0, 1] and the scale radius is somewhere in (0, 3] Mpc \(h^{-1}\) than that the data have been drawn from a zero correlation function.

\(^{12}\) This means that it is 27.9 times more likely that the data have been drawn from an exponential whose amplitude is somewhere in the range of (0, 1] and the scale radius is somewhere in (0, 3] Mpc \(h^{-1}\) than that the data have been drawn from a zero correlation function.

\(^{13}\) We do not expect distance errors of the order of 0.2 Mpc \(h^{-1}\) to have a large impact on a distance bin of 1 Mpc \(h^{-1}\) binsize.
enables us to estimate a marginal autocorrelation function which is shown in panel (d) of Fig. 6. In comparison to panel (c), there is only a minor increase in the error bars. However, we would not rely too much on the marginal estimate because the error estimate of ellipticities is rather coarse. Nevertheless, compared to panel (a), the marginal estimate differs substantially from a conditional estimate and there are no statistically significant autocorrelations.

5.10 Constraining theoretical parameters

The autocorrelation of angular-momentum orientations can be used to estimate free parameters in the tidal-torque theory (e.g. Lee & Pen 2008). Let $\xi(r, R)$ denote the two-point correlation function of the density field, smoothed over the scale $R$. In this case, one can derive a model prediction for the linear regime (e.g. Pen et al. 2000):

$$\xi_{\text{LL}}(r) \approx \frac{a^2}{6} \frac{\xi^2(r, R)}{\xi^2(r, 0)} + \frac{\xi(r, R)}{\xi(r, 0)},$$

(34)

where $a$ is a free model parameter. For the non-linear regime, Lee & Pen (2008) derived the following model prediction:

$$\xi_{\text{LL}}(r) \approx \frac{a^2}{6} \frac{\xi^2(r, R)}{\xi^2(r, 0)} + \frac{\xi(r, R)}{\xi(r, 0)} + \varepsilon_{\text{NL}},$$

(35)

where $a$ and $\varepsilon_{\text{NL}}$ are free model parameters describing the linear and non-linear contributions. Estimating values for these model parameters is important in order to constrain the tidal-torque theory. The impact of the additional error sources on this parameter estimation is devastating. First, the marginal estimate of $\xi_{\text{LL}}(r)$ has large errors. Secondly, errors in redshift estimates and morphological classification also affect the estimation of the two-point correlation function $\xi(r, R)$. Given these considerations and the SDSS sample, we have to conclude that it is currently impossible to place decisive constraints on the theoretical parameters.

The same argument applies to the generic autocorrelation model proposed by Schäfer & Merkel (2011),

$$\xi_{\text{LL}}(r) = A \exp \left[ - \left( \frac{r}{R} \right)^C \right],$$

(36)

which contains a linear amplitude $A$ and two non-linear model parameters $R$ and $C$ which cannot be constrained properly. Fig. 7 demonstrates this by showing the marginal likelihoods of fitting equation (36) to the binned data of panel (d) of Fig. 6. Evidently, the (marginal) uncertainties in all the model parameters are extremely large. Nevertheless, let us note that the correlation length of 1 Mpc $h^{-1}$ predicted by Schäfer & Merkel (2011) is in agreement with our estimate. Furthermore, for later purposes, we identify the best-fitting model:

$$\xi_{\text{LL}}(r) \approx 0.026 \exp \left[ - \left( \frac{r}{0.34 \text{ Mpc} h^{-1}} \right)^{0.46} \right].$$

(37)

14 In fact, this is the reason why Lee (2011) restricts the sample to galaxies with $z \leq 0.02$ in order to obtain a volume-limited sample. Otherwise, the density field of galaxies cannot be meaningfully defined and $\xi(r, R)$ cannot be estimated.

15 As $\xi_{\text{LL}}(r)$ has been estimated for Scd galaxies, also $\xi(r, R)$ has to be estimated for this type of galaxies.

16 Actually, we should estimate the correlation function from unbinned data like in Section 5.8. However, a meaningful likelihood function is not easily defined in this case such that we have to resort to fitting binned data. We are fully aware that binning may compromise our assessment of statistical significance.

17 Note that the maximum of the joint likelihood does not coincide with the maxima of the marginalized likelihoods in Fig. 7.

6 BIASED ELLIPTICITY ESTIMATES FROM SECOND MOMENTS

Isophotal ellipticity estimates have the disadvantage that they strongly depend on the choice of a particular isophote and therefore may suffer strongly from pixel noise. Ellipticity estimates based on the moments of the galaxy’s light distribution at first glance seem to be more promising, since no isophote is required and the complete data enter the estimate. Consequently, we would expect that ellipticity estimates based on light moments are more robust against pixel noise than isophotal ellipticities which might improve autocorrelation estimates of angular-momentum-orientation vectors. However, in this section, we demonstrate that ellipticity estimates based on second moments of the light distribution are so strongly biased that they cannot be used for investigations of disc alignment. In particular, this bias would cause us to overestimate the correlation due to alignment such that, e.g. we would overestimate its impact on weak-lensing studies.

6.1 Revealing the bias

We also assess the usage of ellipticity estimates based on unweighted second moments of the galaxies’ light distributions.
Fig. 8. Pseudo-marginal estimate of the angular-momentum-orientation autocorrelation for the sample of Scd galaxies taking into account uncertainties in classification, number statistics, errors in redshift and ellipticity estimates. Results have been averaged over 1000 Monte Carlo samples drawn from the error distribution of spectroscopic redshifts. The dots indicate mean values and the error bars correspond to 1 Gaussian standard deviation. The horizontal dashed line indicates the background correlation level estimated from 100 random shufflings of the galaxy positions, a method used by Lee (2011).

Fig. 9. Comparing ellipticities based on isophotes and unweighted second moments. Top panels: axial ratios (left) and orientation angles (right) for Scd galaxies. Bottom panels show the same for Sab galaxies. Axial ratios estimated from second moments are systematically larger than those estimated from isophotal contours, i.e. second moments find the disc galaxies to be rounder. Orientation angles are unbiased. The distributions of axial ratios for Scd and Sab galaxies agrees with the results of Huertas-Company et al. (2011) (their fig. 2).

Fig. 10. Impact of circular Gaussian PSF with Petrosian radius of 1.3 pixel on to convolved axial ratios \(q\) and orientation angles \(\theta\) of exponential-disc profiles with Petrosian radii of 15.8 pixels and intrinsic axial ratios \(0.1 \leq q \leq 1\) and orientation angles \(\theta = 30^\circ\). All the profiles have been truncated at five scale radii. There was no noise in this simulation. The PSF leads to an overestimation of the axial ratios by at most 1.2 per cent for highly elongated objects. As the PSF was circular in this test, orientation angles are not affected.

6.2 Point spread function

Is this bias an effect of the point spread function (PSF) which makes galaxies look rounder than they actually are? This is unlikely because all our objects are large compared to the size of the PSF. The median \(r\)-band Petrosian radius of the 4211 Scd galaxies with SDSS data is 15.8 pixel, whereas the \(r\)-band Petrosian radius of the SDSS PSF is approximately 1.3 pixel.\(^{18}\) Consequently, the impact of the PSF should be small. This expectation is supported by Fig. 10, where we simulate the impact of a Gaussian PSF with the Petrosian radius of 1.3 pixel on to exponential-disc profiles with Petrosian radii of 15.8 pixel and different intrinsic axial ratios. We find a maximum overestimation of axial ratios of only 1.2 per cent, which is not enough to explain the strong bias in Fig. 8 or the discrepancy in Fig. 9.

6.3 The origin of the bias: Galactic bulges

We are now going to argue that the heavily biased correlation estimate of Fig. 8 stems from the galactic bulges biasing the second moments and thereby the ellipticity estimates. At first glance, this may seem to be a rather unlikely explanation, since we explicitly selected only Scd galaxies in order to minimize the impact of galactic bulges. However, this hypothesis can explain the substantial discrepancy between isophotal axial ratios and axial ratios based on second moments revealed by Fig. 9. If bulges were an issue, they would affect the second moments and would lead us to overestimate the axial ratios, since bulges are in any case ‘roundish’. On the

\(^{18}\) The \(r\)-band Petrosian radius of the SDSS PSF has been estimated as the median \(r\)-band Petrosian radius of 100,000 stars downloaded from the SDSS data base.
Figure 11. Bulge-disc decomposition of an example Scd galaxy (g band). The bulge is a circular deVaucouleur profile while the disc component is an exponential profile with ellipticity. The bulge is pinned to the pixel of the peak-of-light whereas the centroid of the disc component is free. Panel (a) shows the original galaxy. Panel (b) is the disc component, while panel (c) is the bulge component. Panel (d) displays the fit residuals. The fit was performed by $\chi^2$-minimization using a Simplex algorithm (Nelder & Mead 1965) and reached a minimum value of 3.18 per pixel.

On the other hand, isophotal ellipticity estimates should remain unaffected by the presence of bulges as long as the isophote used is in the disc component. In fact, Bernstein (2010) discussed this issue in the context of shear measurements in weak lensing. We demonstrate that the presence of a bulge can bias the estimate of axial ratio based on second moments. For this purpose, we perform a bulge-disc decomposition of a prototypical Scd galaxy from our data sample, which is shown in Fig. 11. Indeed, the axial ratio estimated from the second moments of the complete model (including bulge) is $q_{bd} \approx 0.48$, whereas the axial ratio used by the disc model is only $q_{disc} \approx 0.38$. We conclude that the bulge is well capable of biasing the ellipticity estimate substantially, even in the case of Scd galaxies.

As another test for our hypothesis to pass, we compare the axial ratios based on isophotes and second moments for Sab galaxies from the sample of Huertas-Company et al. (2011). As Sab galaxies have more prominent bulges than Scd galaxies, we would expect a stronger bias than in Fig. 9. We select all galaxies with $p_{Sab} \geq 0.8$ and download the $r$-band Stokes parameters from the SDSS data base, if available. For the resulting 8496 Sab galaxies, Fig. 9 also shows the comparison of ellipticities estimated from isophotes and second moments. Evidently, the second moments are biased too, and the bias is also more pronounced than for Scd galaxies. This answers our expectation.

From our hypothesis of bulges biasing second moments, we can deduce the following prediction: if galactic bulges indeed bias second moments such that estimated angular-momentum-orientation vectors are bent into the line of sight, the angular correlation function should exhibit a bias of the form

$$b(\theta) = A + B \cos^2 \theta,$$

(38)

where $\theta$ now denotes the angular separation of two galaxies. The reason is that due to the bending of orientation vectors, the scalar product $L \cdot L'$ is on average equal to the cosine of the two galaxies’ separation angle. This prediction is confirmed by Fig. 12 which strongly suggests that $\hat{\xi}_{LL}(\theta)$ is dominated by this bias. This suspect behaviour is also exhibited by the autocorrelation function in real space, as shown in the top panel of Fig. 13. Fig. 12 also shows that when using isophotal ellipticity estimates, $\hat{\xi}_{LL}(\theta)$ does not exhibit such a bias.

Is it possible to debias the autocorrelation function by subtracting equation (38) from all pairwise projections of angular-momentum-orientation vectors. Top panel: the biased autocorrelation function based on ellipticity estimates from second moments. Middle panel: ‘debiased’ correlation function where equation (38) has been subtracted from all pairwise projections. The solid orange line is the fit given by equation (39). Bottom panel: autocorrelation function based on isophotal ellipticities.

Figure 12.

Figure 13.

Debiasing the autocorrelation function of angular-momentum-orientation vectors. Top panel: the biased autocorrelation function based on ellipticity estimates from second moments (top) and isophotes (bottom). The bias model of equation (38) with $1\sigma$ errors is shown in the top panel.

Figure 13.

Debiasing the autocorrelation function of angular-momentum-orientation vectors. Top panel: the biased autocorrelation function based on ellipticity estimates from second moments (top) and isophotes (bottom). The bias model of equation (38) with $1\sigma$ errors is shown in the top panel.

The $g$-band axial ratio noted in the SDSS data base for this example galaxy is $q_{iso} \approx 0.41$ estimated from isophotes and $q_{mom} \approx 0.63$ estimated from second moments (Stokes parameters). The discrepancy of axial ratios from the SDSS data base and the bulge-disc decomposition is the consequence of a non-optimal model.

The parameter values $A$ and $B$ depend on the details of the bias caused by the galactic bulges and are not generally predictable.

Note that the angular correlation estimate in Fig. 12 looks worse than the spatial correlation estimate of Fig. 6d. This is due to the fact that the angular correlation function does not use distance information.
orientation vectors? We investigate this question in Fig. 13, where we show the biased and debiased autocorrelation function. Indeed, the debiased autocorrelation function looks very promising. For later modelling purposes, we parametrize the debiased autocorrelation function by

\[ \xi_{LL}(r) \approx (0.013 + 0.002r - 0.00036r^2) \exp \left( -\frac{r}{6.1 \text{ Mpc} h^{-1}} \right), \]

(39)

where no error estimate is required because we only use this fit as an input in simulations. Is the debiased autocorrelation function trustworthy? For comparison, Fig. 13 also shows the unbiased autocorrelation function based on isophotal ellipticities. Evidently, the debiased and isophotal autocorrelation functions do not agree. However, this does not necessarily rule out the debiased autocorrelation function because we actually expect that ellipticity estimates based on second moments are less noisy than isophotal ellipticity estimates since they use the whole light distribution instead of a single isophote. Hence, it is not a priori implausible that the debiased autocorrelation function exhibits more information than the isophotal autocorrelation function.

In order to assess the trustworthiness of the debiased autocorrelation estimate, we conduct the following self-consistency test: we take the original galaxies as in Fig. 13, maintaining their true spatial positions, but when estimating the autocorrelation function, we replace the actual angular-momentum-orientation vectors by simulated vectors which exhibit the correlation function given by equation (39). This simulation is described in Appendix A2. Panel (a) of Fig. 14 validates our simulation method. We then simulate the bias of second moments. For every galaxy, we take the simulated angular-momentum-orientation vector and infer the actual axial ratio \( q_{\text{true}} \) from it. Motivated by the top-left panel of Fig. 9, we then replace the true axial ratio by an ‘overestimate’ drawn from the uniform distribution over the interval \( [q_{\text{true}}, 1] \). Using this biased axial ratio, we recompute the angular-momentum-orientation vector and estimate the correlations. As shown in panel (b) of Fig. 14, the resulting biased autocorrelation function closely resembles the observation from Fig. 13. For debiasing, we then also estimate the autocorrelation in angular space, as shown in panel (c) of Fig. 14. Indeed, the estimate is dominated by a bias of the form of equation (39), i.e. our bias simulation is realistic. We then estimate the debiased autocorrelation function, which is shown in panel (d). Evidently, the debiased result exhibits systematic and significant deviations from the input autocorrelation function. We emphasize that the debiased result is not an obscured version of the input correlation function. Neither their difference nor their ratio is a constant, i.e. the debiasing was not successful. Consequently, the debiasing is not self-consistent and the debiased autocorrelation estimate shown in Fig. 13 is not trustworthy.

7 IMPROVEMENTS IN AND POTENTIAL OF FUTURE SURVEYS

We showed in Figs 4(d) and 6(d) that with current data there are no statistically significant autocorrelations. What can be done to improve these results? In this section, we briefly elaborate on the improvements in ellipticity estimates and the potential of future sky surveys, namely PanSTARRS, LSST and EUCLID, to enhance the estimates of handedness and angular-momentum-orientation autocorrelations. We discuss the impact of number statistics and improvements in redshift estimates. We also discuss morphological classification and estimation of front-edges of disc galaxies.

Figure 14. Self-consistency test of debiasing the autocorrelation function. Panel (a): the input autocorrelation function as given by equation (39), validating our simulation technique. Panel (b): the biased autocorrelation function. Panel (c): the debiasing of the autocorrelation function in angular space. Panel (d): the debiased autocorrelation function, which exhibits significant deviations from the input.

7.1 Improving ellipticity estimates

We demonstrated in Section 6 that ellipticity estimates based on second moments are strongly biased by galactic bulges even for Scd galaxies. In fact, Fig. 12 suggests that correlation estimates based on second moments are completely dominated by this bias which overwrites the desired astrophysical signal. Therefore, we conclude that ellipticity estimates based on second moments overestimate the axial ratios and thereby corrupt the estimates of angular-momentum-orientation autocorrelation. This bias also corrupts similar correlation estimates, such as ellipticity autocorrelations (e.g. Blazek, McQuinn & Seljak 2011), leading us to overestimate the impact of disc alignment on weak-lensing studies. What are the alternative ellipticity estimators? This same bias also applies to adaptive moments (Bernstein & Jarvis 2002; Hirata & Seljak 2003) in this context. Furthermore, model-based ellipticity estimates are problematic, since nearby disc galaxies usually exhibit rich azimuthal structures, which are virtually impossible to model faithfully. The only kind of model designed to describe such rich azimuthal structures are basis-function expansions (e.g. Massey & Réfrégier 2005; Ngan et al. 2009), which unfortunately suffer from other severe conceptual problems (Melchior et al. 2010; Andrae, Melchior & Jahnke 2011). We have to conclude that isophotal ellipticities—though relying on a somewhat arbitrarily chosen isophote—are
the only useful ellipticity estimates for investigations of angular-momentum-orientation autocorrelation, since they are closest to the disc ellipticity.

There is yet another serious conceptual issue we have to face. In the weak-lensing context galaxies are usually rather small with radii of a few pixels only. In our case, however, we are considering large extended disc galaxies. Disc galaxies usually exhibit substructures such as galactic bars, rings or star-forming regions. In particular, the Scd galaxies considered by Lee (2011) and in this work typically exhibit very open spiral-arm patterns. For such objects there are considerable ellipticity gradients (Bernstein 2010) and ‘disc ellipticity’ is not a well-defined concept anymore. Therefore, it may be helpful to estimate ellipticities in the near-infrared regime, where, e.g. star-forming regions are not as prominent as in the optical regime such that disc galaxies look smoother.

7.2 Improving number statistics

An obvious strategy to improve estimates of handedness or angular-momentum-orientation autocorrelations is to increase the number of galaxies in the data sample. For instances, SDSS and thereby Galaxy Zoo cover approximately one quarter of the full sky. How would an extension to an (extragalactic) all-sky survey improve the autocorrelation estimates? If we assume an identical depth, this areal extension leaves the galaxy density unchanged, it only increases the number of galaxy pairs in all distance bins.

In order to study the improvement in an enlarged survey area, we randomly draw subsamples from the Galaxy Zoo data base (a larger data base is not available, so we use smaller data bases) and estimate their handedness autocorrelations. In fact, we do not draw the subsamples from the data base itself, which would correspond to reducing the galaxy density. Instead, we randomly draw the subsamples from the list of galaxy pairs. Fig. 15 clearly shows that the errors in the handedness autocorrelation function are indeed dominated by number statistics, since the errors depend on sample size with a power law of exponent \(-\frac{1}{2}\). Consequently, an extension leaves the galaxy density unchanged, it only increases the number of galaxy pairs in all distance bins.

![Figure 15](https://example.com/figure15.png)

Figure 15. Impact of number statistics on the errors of the innermost three distance bins in the marginal handedness autocorrelation function. The x-axis shows the fraction of galaxy pairs selected from all pairs, which is equivalent to a survey covering the same fraction of the total survey area. Both axes are in logarithmic scale, i.e. the dependence of the errors is approximately a power law for all the three bins. The dashed line indicates a power law of \(N^{-1/2}\), where \(N\) is the number of pairs in every bin.

7.3 Improving redshift estimates

Reducing the errors in spectroscopic redshift estimates would clearly help in reducing the errors in the autocorrelation functions. For instances, the redshift error of \(\sigma_z = 7.8 \times 10^{-5}\) at \(z = 6.5993 \times 10^{-2}\) quoted in Fig. 2 corresponds to an error in the radial-velocity estimate of \(\sigma_v = c \sigma_z (1 + z)^2 \approx 20.6 \text{ km s}^{-1}\). However, given the typical velocity dispersion of galaxies in small groups of \((202 \pm 10) \text{ km s}^{-1}\) and in large clusters of \((854 \pm 102) \text{ km s}^{-1}\) (Becker et al. 2007), the spectroscopic redshift estimates of SDSS are already picking up peculiar motions of individual galaxies instead of cosmological expansion. Consequently, further improving the accuracy of spectroscopic redshifts cannot improve estimates of, e.g. the handedness autocorrelation function.

Given the impact of uncertainties in spectroscopic redshift estimates on, e.g. the handedness autocorrelation function, it is obvious that photometric redshift estimates cannot help us to improve the situation. Typically, uncertainties in photometric redshift estimates are 2 orders of magnitude larger than uncertainties in spectroscopic redshift estimates. Considering Fig. 2, this would lead us to an error in the comoving distance of several tens of Mpc \(h^{-1}\). Moreover, though there are many more galaxies with photometric redshift estimates than galaxies with spectroscopic redshift estimates (typically at least 1 order of magnitude), these additional objects are typically also much fainter because selection for spectroscopic observations is usually triggered by the galaxy’s brightness. The faintness of these additional objects would therefore also complicate the morphological classification. For a disc galaxy, the fainter the object, the more difficult it is to identify the disc. Consequently, surveys that offer only photometric but no spectroscopic redshift estimates are of no use to estimate these autocorrelation functions. This essentially rules out PanSTARRS and LSST. Conversely, the EUCLID survey will gather of the order of 100 million spectroscopic redshifts of galaxies. Unfortunately, the galaxy sample observed by EUCLID will have redshifts between 0.5 and 2. As was shown by Crittenden et al. (2001), estimates of handedness and angular-momentum-orientation correlations are compromised by weak-lensing signals for \(z > 0.3\).

7.4 Morphological classification in future surveys

Evidently, autocorrelation estimates of handedness and angular-momentum orientation require morphological classification in future surveys. As we cannot probe high-redshift galaxies for this purpose, the morphological classes used by Galaxy Zoo or Huertas-Company et al. (2011) are sufficient and no further diversification is necessary. In particular, this implies that we can build on these two morphological catalogues to classify galaxies in future surveys: first, we match for the galaxies of known morphological types in the new survey; secondly, we use the new survey’s imaging or spectroscopic data to estimate those galaxy’s parameters. Finally, using these parameters and the galaxies of known morphological types as a training sample, we can set up a probabilistic classification algorithm to extend this classification scheme to the new survey catalogue. In fact, this is precisely the same exercise as Huertas-Company et al. (2011) did, but on a much larger scale. In particular, the Galaxy Zoo sample with approximately 900 000...
We analysed the galaxy population as a whole and did not specialize on particular dust lanes. We visually inspect $g$-band images, of all the five SDSS bands this band is most strongly affected by dust extinction while still being of a decent depth. The outcome of such a visual inspection is as follows.

1. Equal weights $p_a = p_b = \frac{1}{2}$ if we are uncertain.
2. Weight of 0.6 to indicate a somewhat uncertain trend.
3. Weight of 0.9 if we believe it to be certain.

We do not assign a weight of 1 in the last case, since there is always some uncertainty. By construction, this method works best for edge-on discs, since face-on discs do not display dust lanes. Unfortunately, knowing the front-edge would have a larger impact for nearly face-on discs than for edge-on discs (see definitions in Lee 2011). We visually inspected $g$-band images of the 500 largest galaxies, sorted by their Petrosian radii. For smaller galaxies, the resolution is not good enough to identify dust lanes. Unfortunately, we find only very few decisive front-edge classifications, namely 40 Scd galaxies with certain front-edge classifications and 39 with somewhat uncertain front-edge. Consequently, we find no substantial improvement in the marginal correlation estimate. Nevertheless, future sky surveys may have an improved imaging quality, such that a visual front-edge classification is possible for more objects.

8 DISCUSSION AND CONCLUSIONS

We have shown that when all the relevant error sources are taken into account, there are no statistically significant autocorrelations, neither of spiral-arm handedness nor of angular-momentum-orientation vectors of Scd galaxies. Previous estimates (Slosar et al. 2009; Lee 2011) did not account for these error sources and therefore are conditional estimates that underestimated the errors and overestimated statistical significance. Using a KS test to analyse the angular-momentum-orientation vectors in the Local Group, the null hypothesis of random orientation yields a $p$-value of 64.8 per cent, i.e. it cannot be rejected. Therefore, there is no evidence that disc alignment is at work in the Local Group. Nevertheless, this does not yet falsify the tidal-torque theory for two reasons: First, we indeed see indications for potential autocorrelations, though they are not statistically significant. These indications are consistent with the theoretically predicted correlation length of 1 Mpc$^{-1}$. Improving the data might help us to test these indications. Secondly, the tidal-torque theory predicts the alignment for angular momenta of dark matter haloes and not for the disc galaxies residing inside these haloes. For instances, van den Bosch et al. (2002) find a median misalignment of angular momenta of disc galaxies and their host haloes of $\approx 30^\circ$. Furthermore, even minor mergers can significantly disturb the angular momenta of disc galaxies by transferring orbital angular momentum (e.g. Moster et al. 2010). Conversely, we could speculate whether there is some relaxation process compensating, e.g. for perturbations by mergers. However, we do not want to stretch this discussion too far because we are wary of turning the tidal-torque theory from an empirical into a ‘vampirical’ hypothesis where virtually any observational result can be explained such an empirical falsification becomes impossible (Gelman & Weakliem 2009).

We must conclude that with currently available SDSS data it is not possible to place decisive constraints on the free parameters of theoretical models. We discussed that a full-sky survey of SDSS quality might improve the situation such that these autocorrelations could become statistically significant. Furthermore, we argued that photometric redshift estimates of SDSS quality have too large errors to be useful for this task, instead spectroscopic redshift estimates are necessary. Finally, we discussed that a front-edge classification of disc galaxies might improve the autocorrelation estimate of angular-momentum orientation, since it breaks the geometric degeneracy of the galaxy’s disc inclination. However, we find that imaging data allow visual front-edge classification only for a minute fraction of the galaxy’s disc inclination, where the impact of dust extinction is larger than in $r$, $i$, $z$ whereas the $g$ band is not as shallow as the $u$ band. This approach would compare the fluxes above and below the major axis, thereby estimating the front edge. In contrast to colour-based methods, this approach does not rely on accurate photometric positions. However, like any automated method for front-edge classification, it suffers from several other effects such as star-forming regions in the galaxy or foreground stars which compromise colour gradients and flux differences. These effects are the major obstacles which have to be overcome in order to set up a reliable front-edge classification algorithm.

23 We analysed the galaxy population as a whole and did not specialize on any environment such as filaments (Jones, van de Weygaert & Aragón-Calvo 2010).
of objects in the catalogue, whereas automated front-edge classification is severely hampered by foreground stars and star-forming regions. Unfortunately, there are no upcoming surveys that fulfill all these requirements. Consequently, the search for autocorrelations of angular momenta of disc galaxies may remain an open issue for the unforeseeable future.

We demonstrated that ellipticity estimates based on second moments of the galaxies’ light distributions are strongly biased by the presence of galactic bulges even for Scd galaxies. This bias corrupts autocorrelation estimates of angular-momentum orientation because it dominates over the expected astrophysical signal. For instances, this leads to an overestimation of the impact of disc alignment in weak-lensing studies (Blazek et al. 2011).

ACKNOWLEDGMENTS

First and foremost, RA thanks Ellen Andrae for extensive discussions on the contents of this article. RA furthermore thanks Andrea Maccio for discussing the interpretation of the results and Nicolas Martin for valuable information about the Local Group. RA would also like to thank Anže Slosar who kindly provided his list of Galaxy Zoo object pairs that were visually classified as rogue pairs. Furthermore, Jounghun Lee was very kind in helping us to reproduce her results. Both, Anže Slosar and Jounghun Lee, also provided detailed comments that were of great help to improve this manuscript. RA also thanks Marc Huertas-Company, who kindly provided additional morphological information on his catalogue of classifications. The authors also want to thank Matthias Bartelmann and Coryn Bailer-Jones, who helped them to improve this article. RA is funded by the Klaus Tschira Foundation via the Heidelberg Graduate School of Fundamental Physics (HGSFP). KJ is supported by the Emmy-Noether-programme of the DFG.

REFERENCES

Andrae R., Melchior P., Bartelmann M., 2010, A&A, 522, A21
Andrae R., Melchior P., Jahnke K., 2011, preprint (arXiv:1106.6045)
Bamford S. P. et al., 2009, MNRAS, 393, 1324
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Becker M. R. et al., 2007, ApJ, 669, 905
Bernstein G. M., 2010, MNRAS, 406, 2793
Bernstein G. M., Jarvis M., 2002, AJ, 123, 583
Blake C. et al., 2011, MNRAS, 415, 2892
Blazek J., McQuinn M., Seljak U., 2011, J. Cosmol. Astropart. Phys., 5, 10
Brunthaler A., Reid M. J., Falcke H., Greenhill L. J., Henkel C., 2005, Sci, 307, 1440
Cameron E., 2011, Proc. Astron. Soc. Australia, 28, 128
Catelan P., Kamionkowski M., Blandford R. D., 2001, MNRAS, 320, L7
Chemin L., Carignan C., Foster T., 2009, ApJ, 705, 1395
Crittenden R. G., Natarajan P., Pen U., Theuns T., 2001, ApJ, 559, 552
Gauci A., Zarb Adami K., Abela J., 2010, preprint (arXiv:1005.0390)
Gelman A., Weakliem D., 2009, American Sci.
Haynes M. P., Giovanelli R., 1984, AJ, 89, 758
Heavens A., 2009, preprint (arXiv:0906.0664)
Heavens A., 2009, preprint (arXiv:0906.0664)
Heidmann J., Heidmann N., de Vaucouleurs G., 1972, Mem. R. Astron. Soc., 75, 85
Huertas-Company M., Aguerri J. A. L., Bernardi M., Mei S., Sánchez Almeida J., 2011, A&A, 525, A157
Iye M., Ozata T., 1999, in Merritt D. R., Valluri M., Sellwood J. A., eds, ASP Conf. Ser. Vol. 182, Galaxy Dynamics – A Rutgers Symposium. Astron. Soc. Pac., San Francisco, p. 255
Jones B. J. T., van de Weygaert R., Aragón-Calvo M. A., 2010, MNRAS, 408, 897
Kass R. E., Raftery A. E., 1993, Bayes Factors and Model Uncertainty. Tech. Rep., Dep. Statistics, Univ. Washington
Land K. et al., 2008, MNRAS, 388, 1686
Larson D. et al., 2011, ApJS, 192, 16
Lee J., 2011, ApJ, 732, 99
Lee J., Pen U., 2008, ApJ, 681, 798
Lintott C. J. et al., 2008, MNRAS, 389, 1179
Lintott C. et al., 2011, MNRAS, 410, 166
Massey R., Réfrégier A., 2005, MNRAS, 363, 197
Melchior P., Böhnet A., Lombardi M., Bartelmann M., 2010, A&A, 510, A75
Moster B. P., Macciò A. V., Somerville R. S., Johansson P. H., Naab T., 2010, MNRAS, 403, 1009
Nelder J. A., Mead R., 1965, Comput. J., 7, 308
Ngo N., van Waerbeke L., Mahdavi A., Heymans C., Hoekstra H., 2009, MNRAS, 396, 1211
Olsen K. A. G., Massey P., 2007, ApJ, 656, L61
Pen U., Lee J., Seljak U., 2000, ApJ, 543, L107
Press W., Teukolsky S., Vetterling W., Flanner B., 2002, Numerical Recipes in C++. Cambridge Univ. Press, Cambridge
Putman M. E. et al., 2009, ApJ, 703, 1486
Schäfer B. M., 2009, Int. J. Modern Phys. D, 18, 173
Schäfer B. M., Merkel P., 2011, preprint (arXiv:1101.4584)
Slosar A. et al., 2009, MNRAS, 392, 1225
van den Bosch F. C., Abel T., Croft R. A. C., Hernquist L., White S. D. M., 2002, ApJ, 576, 21
van der Marel R. P., Cioni M.-R. L., 2001, AJ, 122, 1807
Walterbos R. A. M., Kennicut R. C., Jr, 1987, A&AS, 69, 311
Walterbos R. A. M., Kennicut R. C., Jr, 1988, A&A, 198, 61

APPENDIX A: SIMULATING PAIRS OF ANGULAR-MOMENTUM-ORIENTATION VECTORS

In this appendix, we explain how to simulate pairs of angular-momentum-orientation vectors which should exhibit a given correlation.

A1 Uncorrelated, orthonormal orientation vectors

As the orientation vectors indicate directions, the samples are drawn from the uniform distributions \( \varphi \in [0, 2\pi) \) and \( \cos \theta \in [-1, 1] \) of the two polar angles \( \varphi \) and \( \theta \). A random orientation vector is then given by

\[
\ell_1 = \begin{pmatrix} \cos \varphi \sin \theta \\ \sin \varphi \sin \theta \\ \cos \theta \end{pmatrix}. \tag{A1}
\]

This vector is normalized, i.e. \( \ell_1 \cdot \ell_1 = 1 \). Sampling a uniform angle \( \phi \in [0, 2\pi) \), a second random orientation vector is

\[
\ell_2 = \sin \phi \begin{pmatrix} -\sin \varphi \\ \cos \varphi \cos \theta \\ \sin \varphi \cos \theta \end{pmatrix} + \cos \phi \begin{pmatrix} \cos \varphi \\ -\sin \phi \cos \theta \\ \sin \phi \sin \theta \end{pmatrix}. \tag{A2}
\]

This vector is again normalized, i.e. \( \ell_2 \cdot \ell_2 = 1 \), and also orthogonal to the first, i.e. \( \ell_1 \cdot \ell_2 = 0 \).
Only marginal alignment of disc galaxies

The autocorrelation is then given by
\[ \langle (L_a \cdot L_a')^2 \rangle = \cos^2 \beta . \] (A10)

A2.2 Computing the other terms

The other three terms in equation (A7) are computed in precisely the same way. We obtain
\[ \langle (L_a \cdot L_b)^2 \rangle = \frac{7}{15} \cos^2 \beta + \frac{4}{15} \sin^2 \beta . \] (A11)

As the correlation estimate is invariant under exchanging the pair, we can directly conclude that
\[ \langle (L_b \cdot L_a)^2 \rangle = \frac{7}{15} \cos^2 \beta + \frac{4}{15} \sin^2 \beta , \] (A12)
as well. The final term is given by
\[
\langle (L_b \cdot L_b')^2 \rangle = \left[ \frac{7}{15} - \frac{8}{5} (e_r \cdot e'_{r})^2 + \frac{32}{15} (e_r \cdot e'_{r})^4 \right] \cos^2 \beta \\
+ \left[ \frac{4}{15} + \frac{4}{5} (e_r \cdot e'_{r})^2 - \frac{16}{15} (e_r \cdot e'_{r})^4 \right] \sin^2 \beta ,
\] (A13)
which depends on the angular separation \( e_r \cdot e'_{r} \) of the galaxy pair that is simulated. This dependence is inherited from flipping the radial component of both angular-momentum-orientation vectors due to an unknown front edge.

A2.3 Mixing angle

Inserting all four terms into equation (A7), we can solve for the mixing angle for a given input correlation. The result is
\[ \cos \beta = \sqrt{\frac{1}{3} + \frac{20 \xi_{\text{input}}}{16 (e_r \cdot e'_{r})^2 - 12 (e_r \cdot e'_{r})^4 + 8}} . \] (A14)
This mixing angle is used in Section 6.3.

This paper has been typeset from a \TeX/LaTeX file prepared by the author.