Basilisk: Bayesian hierarchical inference of the galaxy–halo connection using satellite kinematics – I. Method and validation

Frank C. van den Bosch 1, * Johannes U. Lange 1 and Andrew R. Zentner 2

1 Department of Astronomy, Yale University, P.O. Box 208101, New Haven, CT 06520-8101, USA
2 Department of Physics and Astronomy & Pittsburgh Particle Physics, Astrophysics, and Cosmology Center (PITT PACC), University of Pittsburgh, Pittsburgh, PA 15260, USA

Accepted 2019 July 19. Received 2019 June 28; in original form 2019 April 24

ABSTRACT
We present a Bayesian hierarchical inference formalism (Basilisk) to constrain the galaxy–halo connection using satellite kinematics. Unlike traditional methods, Basilisk does not resort to stacking the kinematics of satellite galaxies in bins of central luminosity, and does not make use of summary statistics, such as satellite velocity dispersion. Rather, Basilisk leaves the data in its raw form and computes the corresponding likelihood. In addition, Basilisk can be applied to flux-limited, rather than volume-limited samples, greatly enhancing the quantity and dynamic range of the data. And finally, Basilisk is the only available method that simultaneously solves for halo mass and orbital anisotropy of the satellite galaxies, while properly accounting for scatter in the galaxy–halo connection. Basilisk uses the conditional luminosity function to model halo occupation statistics, and assumes that satellite galaxies are a relaxed tracer population of the host halo’s potential with kinematics that obey the spherical Jeans equation. We test and validate Basilisk using mocks of varying complexity, and demonstrate that it yields unbiased constraints on the galaxy–halo connection and at a precision that rivals galaxy–galaxy lensing. In particular, Basilisk accurately recovers the full PDF of the relation between halo mass and central galaxy luminosity, and simultaneously constrains the orbital anisotropy of the satellite galaxies. Basilisk’s inference is not affected by potential velocity bias of the central galaxies, or by slight errors in the inferred, radial profile of satellite galaxies that arise as a consequence of interlopers and sample impurity.

Key words: methods: analytical – methods: statistical – galaxies: haloes – galaxies: kinematics and dynamics – dark matter.

1 INTRODUCTION
Accurately constraining the link between galaxies and dark matter haloes, which goes by the catch-all name ‘halo-occupation modelling’, provides valuable insight regarding the formation and evolution of galaxies in a Λ cold dark matter (ΛCDM) cosmology. It describes the link between what we can see (light) and what governs dynamics (mass), and therefore provides a powerful tool to probe (the evolution of) the matter power spectrum. The main techniques that are being utilized to constrain this galaxy–halo connection are galaxy clustering, gravitational lensing, galaxy group catalogues and, to a lesser extent, satellite kinematics.

Since more massive haloes are more strongly clustered (e.g. Mo & White 1996), the amplitude of galaxy clustering on large, linear scales is often interpreted as indicative of the average mass of the haloes in which the galaxies in question reside. However, this method is severely impeded by the issue of assembly bias; halo bias depends not only on halo mass, but also on numerous other halo properties, such as halo formation time, halo spin, and halo concentration (e.g. Gao et al. 2004; Wechsler et al. 2006; Villarreal et al. 2017; Salcedo et al. 2018). Consequently, and contrary to what is assumed in hundreds of studies, large-scale clustering amplitude cannot be used as an unbiased estimator of halo mass (see Wechsler & Tinker 2018, for a comprehensive review). Rather, one constrains some combination of the various halo properties that correlate with halo bias. Whereas ignoring assembly bias can consequently result in significant, systematic errors (e.g. Zentner, Hearin & van den Bosch 2014), properly accounting for it (e.g. Hearin et al. 2016) not only is extremely challenging, but also requires additional constraints on halo masses in order to break the various degeneracies.

* E-mail: frank.vandenbosch@yale.edu
One of the most powerful alternative methods to constrain the galaxy–halo connection is galaxy–galaxy lensing. As an application of (weak) gravitational lensing, it is one of the most direct probes of halo mass. In practice, though, the signal-to-noise ratio of the tangential shear distortions around individual galaxies is typically far too small for a reliable estimate of the galaxy’s halo mass, and one generally resorts to stacking the data for several thousands of lens-galaxies. Starting with the pioneering work by Blandford, Blandford & Smail (1996), this stacking method has been used in numerous studies, and has resulted in accurate measurements of the average relation between stellar mass (or luminosity) and halo mass (Hoekstra et al. 2001; Sheldon et al. 2004; Mandelbaum et al. 2006; Leauthaud et al. 2012; Velander et al. 2014). Galaxy–galaxy lensing on small, non-linear scales has also been used in combination with galaxy clustering in attempts to simultaneously constrain the galaxy–dark matter connection and cosmological parameters (Cacciato et al. 2009, 2013; More et al. 2015; Leauthaud et al. 2017; Wikking et al. 2019). However, none of these studies have allowed for assembly bias, and their results therefore have to be taken with a grain of salt (but see also Lange et al. 2019c).

One can also constrain the galaxy–halo connection using galaxy group finders, which try to group together those galaxies that reside in a common dark matter host halo. The mass of the host halo is typically estimated from the line-of-sight velocity dispersion of its member galaxies (e.g. Eke et al. 2004; Robotham et al. 2011; Tempel et al. 2014), or from the total luminosity or stellar mass using halo abundance matching (e.g. Yang et al. 2005a, 2007). In addition to providing constraints on the galaxy–dark matter connection (e.g. Berlind et al. 2006; Yang, Mo & van den Bosch 2008, 2009; Nurmi et al. 2013), galaxy group catalogues have proven particularly powerful for studying the impact of environment of galaxy demographics (e.g. Weinmann et al. 2006; van den Bosch et al. 2008b; Wetzel et al. 2013; Hou, Parker & Harris 2014; Wang et al. 2018b; Davies et al. 2019). However, it has also become clear that errors in the group finding algorithm and the halo mass assignment can be appreciable (e.g. Campbell et al. 2015; Calderon & Berlind 2019), and thus that the constraints from group catalogues are best combined with additional, independent constraints such as from clustering and/or lensing (Han et al. 2015; Sinha et al. 2018). Satellite kinematics is yet another method to constrain the galaxy–dark matter connection. It uses the notion that satellite galaxies orbiting within the dark matter haloes of their central galaxies are tracers of the gravitational potential, and can therefore be used to probe the relation between halo mass and central galaxy luminosity (or stellar mass). Since individual galaxies typically only have a few (detectable) satellites (with the exception of massive clusters), this method typically relies on the same stacking approach used in galaxy–galaxy lensing. Although it is the oldest technique used to constrain halo masses, starting with the seminal work of Zwicky (1933), it has been somewhat underutilized since the advent of clustering and galaxy–galaxy lensing. This is somewhat surprising, as the actual measurements (redshifts) are much easier to obtain than in the case of galaxy–galaxy lensing (tangential shear distortions). The main reason is that the kinematics of dark matter subhaloes, which host satellite galaxies, are believed to be inconsistent with a steady-state tracer population in a spherical, equilibrium potential (e.g. Wang et al. 2017, 2018a; Adhikari et al. 2019). Consequently, the general notion is that it must be extremely difficult to extract reliable halo masses. In addition, satellite orbits are likely to be anisotropic (Diemand, Moore & Stadel 2004; Cuesta et al. 2008; Wojtak & Mamon 2013), further complicating the modelling. And finally, most satellite kinematics studies in the past have been extremely conservative in selecting central–satellite pairs, to the extent that the signal-to-noise ratio of the data did not allow for competitive constraints on the galaxy–halo connection (see Lange et al. 2019a, for a detailed, historical overview).

As we demonstrate in this paper, and have demonstrated before (Lange et al. 2019a), these issues are far less severe than has been suggested, and satellite kinematics can be used as a competitive, reliable probe of the galaxy–halo connection. In particular, although individual haloes may not be spherical, and individual satellite galaxies may not obey the spherical Jeans equation, to a good approximation the ensemble of satellite galaxies can be treated as a steady-state tracer population of the ensemble of host haloes. And, as we demonstrate in this paper, modelling such ensembles using the spherical Jeans equation yields unbiased estimates of the galaxy–halo connection, as long as one carefully accounts for scatter (i.e. ‘mass-mixing’, see Section 2.3), sample selection effects (i.e. interlopers, impurity and incompleteness, see Section 3.3), and, to a lesser extent, orbital anisotropy. In addition, van den Bosch et al. (2004) demonstrated that by using iterative, adaptive selection criteria one can boost the number of central–satellite pairs by an order of magnitude, while simultaneously decreasing the fraction of interlopers (galaxies unassociated with the dark matter halo of the central) and increasing the dynamic range of the galaxy–halo connection probed. This sample selection method was subsequently used by More et al. (2009a, 2011) who were able to obtain tight constraints on the galaxy–halo connection.

Although More et al. (2011) found red centrals to reside in more massive haloes than blue centrals of the same stellar mass, a result that has subsequently been confirmed using galaxy–galaxy lensing (Velander et al. 2014; Mandelbaum et al. 2016; Zu & Mandelbaum 2016), their inferred stellar mass-to-halo mass ratios are significantly different (by a factor two to three) than those inferred from clustering and/or galaxy–galaxy lensing (e.g. Dutton et al. 2010; Leauthaud et al. 2012; Mandelbaum et al. 2016; Wechsler & Tinker 2018). In Lange et al. (2019a, b) we improved on the analysis of More et al. (2009a, 2011) by correcting for sample incompleteness due to fibre collisions in the Sloan Digital Sky Survey (York et al. 2000, hereafter SDSS) data, by accounting for covariance in the data, and by using forward modelling to correct the model for small, but significant biases. This alleviates the tension with the lensing results mentioned above, demonstrating that satellite kinematics can yield constraints on the galaxy dark matter connection in good agreement with constraints from galaxy-galaxy lensing and/or clustering.

Here we continue our goal of maturing satellite kinematics into an accurate and precise probe of the galaxy–halo connection. In particular, we develop a Bayesian hierarchical method to analyse satellite kinematics, called Basilisk,\(^1\) which is entirely complementary to the forward-modelling-based method that we recently developed, and applied to SDSS-DR7 data, in Lange et al. (2019a, b). Basilisk has a number of advantages over the standard method for analysing satellite kinematics. First of all, it requires no arbitrary stacking of the data and can be trivially applied to a flux-limited sample, whereas the methodology used by More et al. (2009a, 2011) and Lange et al. (2019a, b) requires volume-limited samples. This drastically increases the quantity and dynamic range of the data. In addition, Basilisk does not make use of any summary statistic (i.e. the satellite velocity dispersion as function of

\(^1\) Bayesian hierarchical inference using satellite kinematics.
central luminosity), but rather leaves the data in its raw form. This has the advantage that all data is used optimally, thereby allowing to simultaneously constrain halo mass and velocity anisotropy. In addition, as a by-product of the method, Basilisk yields estimates for the halo mass of each individual, central galaxy.

In this first paper in a series, we introduce Basilisk and test its performance using mock data. In Section 2 we first discuss the standard method of analysing satellite kinematics, in which we highlight some of its shortcomings. Section 3 presents our new, Bayesian hierarchical framework, and our method for correcting for interlopers and fibre collisions. Section 4 discusses the two main model ingredients; the conditional luminosity function (CLF) that we use to characterize the galaxy–halo connection (Section 4.1), and our model for the phase-space distribution of satellite galaxies within their host haloes (Section 4.2). Section 5 presents our three-tiered validation process, in which we test the performance of Basilisk on a series of mock data sets of increasing complexity and realism. In Section 6 we examine Basilisk’s ability to constrain the anisotropy of satellite galaxies, and we discuss how central velocity bias and errors in the inferred radial profile of satellite galaxies impacts the inference regarding the galaxy–halo connection. Finally, Section 7 summarizes our findings and presents a detailed discussion of pros and cons of Basilisk.

Throughout this work, we assume a $\Lambda$CDM cosmology with $\Omega_m = 0.3071$, $\Omega_b = 0.0483$, $n_s = 0.9611$, $\sigma_8 = 0.8288$, and $h = H_0/100 \text{km s}^{-1} \text{Mpc}^{-1} = 0.6777$, the best-fitting results from the cosmic microwave background analysis of Planck Collaboration XVI (2014).

2 STANDARD APPROACH

Before we outline our new, Bayesian hierarchical approach to satellite kinematics, we first give an overview of what has become the standard method, which basically consists of three steps: (1) selecting a sample of central and satellite galaxies from a galaxy redshift survey, (2) using this data to compute the velocity dispersion of satellite galaxies, with respect to their centrals, as a function of the luminosity or stellar mass of the central, and (3) using these velocity dispersion measurements to constrain the galaxy–dark matter connection. In what follows we describe each of these three steps in detail.

2.1 Selecting centrals and satellites

The standard method to select centrals and satellites, and the one that we will adhere to as well, is to use a cylindrical isolation criterion to identify centrals, and then to assign fainter galaxies within a similar cylindrical volume as corresponding satellites. Due to interlopers and other impurities, discussed below, not every central (satellite) thus selected is indeed a central (satellite). In what follows, we therefore refer to galaxies that are selected as centrals and satellites as primaries and secondaries, respectively.

To be considered a primary, a galaxy must be brighter than any other galaxy in a cylinder defined by radius $R_{\text{ap}}^{\text{pri}}$ and length $\Delta V_{\text{pri},\max}^2$, centred on it. The radius is defined as the physical separation projected on to the sky and the length is measured by the line-of-sight velocity difference (see equation 2 below). We follow Lange et al. (2019a) and apply this criterion in a rank-ordered fashion, starting with the brightest galaxy. Any galaxy located inside the cylinder of a brighter galaxy is removed from the list of potential primaries. All galaxies fainter than the primary and located inside a cylinder defined by $R_{\text{ap}}^{\text{sec}}$ and $\Delta V_{\text{sec},\max}^2$ centred on the primary are identified as secondaries.

The four free parameters that control the selection of primaries and secondaries, $R_{\text{ap}}^{\text{pri}}$, $\Delta V_{\text{pri},\max}^2$, $R_{\text{ap}}^{\text{sec}}$, and $\Delta V_{\text{sec},\max}^2$ determine both the completeness and purity of the sample. Increasing the cylinder used to select primaries, i.e. increasing $R_{\text{ap}}^{\text{pri}}$ and/or $\Delta V_{\text{pri},\max}^2$ boosts the purity among primaries (i.e. it reduces the number of satellites erroneously identified as centrals), but at the cost of a reduced completeness. Similarly, suppressing the number of interlopers, defined as secondaries that are not satellite galaxies within the same halo as the corresponding primary, requires a small secondary-selection cylinder (i.e. small $R_{\text{ap}}^{\text{sec}}$ and/or $\Delta V_{\text{sec},\max}^2$), which also reduces completeness. A reduced completeness not only complicates the modelling, but also results in data of lower signal-to-noise. As first pointed out in van den Bosch et al. (2004), the fact that brighter primaries typically reside in larger haloes, implies that it is advantageous to scale the cylinder sizes with the luminosity of the primary. We do so using the exact implementation of Lange et al. (2019b), who adopt $R_{\text{ap}}^{\text{pri}} = 0.5 \sigma_{200} \text{h}^{-1} \text{Mpc}$, $R_{\text{ap}}^{\text{sec}} = 0.15 \sigma_{200} \text{h}^{-1} \text{Mpc}$, $\Delta V_{\text{pri},\max}^2 = 1000 \sigma_{200} \text{km s}^{-1}$, and $\Delta V_{\text{sec},\max}^2 = 4000 \text{km s}^{-1}$. Here $\sigma_{200}$ is an estimate for the satellite velocity dispersion in units of 200 km s$^{-1}$, which scales with the luminosity of the primary as

$$\log \sigma_{200} = -0.07 + 0.38 \log L_{10} + 0.29 \log^2 L_{10},$$

where $L_{10} = L/(10^{10} \text{h}^{-2} \text{L}_\odot)$. These criteria were optimized for the SDSS, using an iterative approach, as detailed in More et al. (2009a). The values of $R_{\text{ap}}^{\text{pri}}$ and $R_{\text{ap}}^{\text{sec}}$ correspond to roughly 1.25 and 0.375 times the virial radius, respectively, while the value for $\Delta V_{\text{sec},\max}^2$ is large enough to include the vast majority of all satellites, even in massive clusters. Note that we do not scale this parameter with $\sigma_{200}$; although this implies an increasing fraction of interlopers with decreasing central luminosity, these interlopers are easily identified as such. In principle one could follow van den Bosch et al. (2004) and More et al. (2009a) and apply these selection criteria iteratively, each time updating the $\sigma_{200}(L_c)$ relation based on the inference from the satellite kinematics data selected using the previous $\sigma_{200}(L_c)$. However, as detailed in Section 3.1, there is no need for this as Basilisk’s inference is extremely insensitive to moderate changes in equation (1). Furthermore, tests with detailed mock data sets have shown that using equation (1) yields samples that allow for an accurate recovery of the underlying galaxy–halo connection (see Lange et al. 2019a, and Section 5 below). In Section 5.2 we use mock redshift surveys to assess the completeness, the purity, and the interloper contamination of the above selection criteria when applied to a flux-limited SDSS-like survey.

2.2 Characterizing satellite kinematics

Using a sample of primaries and secondaries, the next step is to quantify the kinematics of these secondaries (assumed to be satellite galaxies) within the host haloes of their associated primaries (assumed to be centrals). Since the typical number of satellite galaxies per central is small, except in nearby clusters, this requires stacking whereby one co-adds all central–satellite pairs for centrals in a given range of luminosity,$^2$ [$L_{c,1}$, $L_{c,2}$]. The satellite kinematics are then specified by the line-of-sight velocity distribution

$^2$One may also stack on stellar mass, or any other (combination) of properties of the central galaxies. For brevity we focus on luminosity throughout.
(LOSVD), \(P(\Delta V|L_c)\), where \(L_c\) is a characteristic luminosity for the luminosity bin in question, and
\[
\Delta V = c \frac{(z_s - z_c)}{1 + z_c},
\]
(2)
with \(z_s\) and \(z_c\) the observed redshifts of the satellite and central, and \(c\) the speed of light. The summary statistic that is most often used in the study of satellite kinematics is the satellite velocity dispersion, \(\sigma_{sat}(L_c)\), which characterizes the second moment of this LOSVD. In order to extract \(\sigma_{sat}(L_c)\) from \(P(\Delta V|L_c)\) one has to correct for interlopers, which can be done in a variety of ways, each with its own pros and cons (e.g. Wojtak et al. 2007; Lange et al. 2019a).

There are a number of important shortcomings with this methodology. First, it requires stacking data in some arbitrary luminosity bins in order to measure the corresponding satellite velocity dispersion, \(\sigma_{sat}\), as function of luminosity. Larger bins implies fewer independent measurements of \(\sigma_{sat}\) but at higher signal-to-noise. However, since luminosity correlates with halo mass, it also implies more ‘mass-mixing’, i.e. combining kinematics from satellites orbiting in haloes of different masses. As discussed in Section 2.3, properly addressing the impact of mass-mixing is extremely important, non-trivial, yet often ignored. Secondly, because of interlopers it is virtually impossible to extract an unbiased estimate of the velocity dispersion from the LOSVD (see Becker et al. 2007; Lange et al. 2019a). Thirdly, by relying solely on the velocity dispersion as a summary statistic, one ignores a wealth of additional information encoded in the detailed shape of the LOSVD and in the correlation between \(\Delta V\) and the projected separation, \(R_p\), of individual central–satellite pairs. This additional information allows one to constrain the density profile of the host halo (e.g. Prada et al. 2003), and the orbital anisotropy of satellite galaxies (e.g. Łokas et al. 2006; Wojtak & Mamon 2013). The \(\Delta V - R_p\) correlation also facilitates a more accurate treatment of interlopers (e.g. Wojtak et al. 2007; Lange et al. 2019a).

The method advocated here, and outlined in Section 3, sidesteps all these shortcomings. It utilizes the full \((\Delta V, R_p)\)-data, without any binning and without the use of a summary statistic.

2.3 Constraining the galaxy–dark matter connection

The final step in using satellite kinematics to constrain the galaxy–halo connection is to translate the data, \(\sigma_{sat}(L_c)\), into corresponding constraints on \(P(M|L_c)\), which characterizes the probability that a central of luminosity \(L_c\) resides in a halo of mass \(M\). Ideally, this is done using forward-modelling based on numerical simulations in which haloes are populated with mock galaxies according to a halo occupation model. This has the advantage that model (mock) and data can be treated in the same way, which makes the analysis less susceptible to biases arising from interlopers, sample incompleteness, and sample impropriety. However, as discussed in Lange et al. (2019a), a full-fledged forward-modelling approach is computationally unfeasible at the present, and all previous studies have therefore relied on simple halo mass estimators based on the virial theorem (e.g. Bahcall & Tremaine 1981; Zaritsky & White 1994; McKay et al. 2002; Brainerd & Specian 2003), or on analytical models that use the Jeans equations to predict the satellite kinematics as a function of halo mass (e.g. van den Bosch et al. 2004; Conroy, Wechsler & Kravtsov 2007; More et al. 2009a, 2011; Wojtak & Mamon 2013). In a recent study, Lange et al. (2019b) combined an analytical model based on the Jeans equations with forward modelling, by using the latter to iteratively calibrate and correct the analytical model for small, systematic biases.

An important issue in trying to infer \(P(M|L_c)\) from satellite kinematics is ‘mass-mixing’. There are good reasons to expect a fair amount of scatter in the galaxy–halo connection, such that central galaxies of a given luminosity occupy haloes of varying masses. Put differently, \(P(M|L_c)\) is not a Dirac delta function. Hence, when stacking the satellite kinematics from a number of centrals of similar luminosity, one is combining the kinematics corresponding to a range in halo masses. The extent of mass-mixing is further compounded by the use of luminosity bins of non-zero width. Surprisingly, mass-mixing has been ignored in many previous studies, including McKay et al. (2002), Brainerd & Specian (2003), Prada et al. (2003), Norberg, Frenk & Cole (2008), and Wojtak & Mamon (2013). As first pointed out in van den Bosch et al. (2004), and further corroborated in More, van den Bosch & Cacciato (2009b), this can result in a very significant, systematic bias in the inferred halo masses. The origin of this bias is easy to understand: typically, more massive haloes contain more satellite galaxies. Hence, when stacking central–satellite pairs residing in different haloes, the more massive ones receive a larger ‘weight’ in that they contribute more data points. This ‘satellite-weighting’ (i.e. giving each satellite equal weight) results in a systematic overestimate of the average halo mass (see also Appendix B). This problem can be avoided, though, by weighting each central–satellite pair by the inverse of the number of satellites around that central. Such this ‘host-weighting’ gives equal weight to each central, that the measured velocity dispersion more fairly represents the average halo mass. In fact, as elucidated in More et al. (2009b), by simultaneously modelling the satellite-weighted and host-weighted velocity dispersions, one can actually constrain the amount of scatter in halo mass at given primary luminosity. This idea has been used by More et al. (2009a), More et al. (2011), and Lange et al. (2019b), all of whom were able to put tight constraints on the scatter in the galaxy–halo connection.

Finally, using dynamics to infer masses is hampered by the well-known mass-anisotropy degeneracy (Binney & Tremaine 2008). With the exception of Wojtak & Mamon (2013) all previous studies of satellite kinematics have simply assumed isotropic orbits for the population of satellite galaxies. Although a clear oversimplification, and one that is likely to be systematically wrong (e.g. Diemand et al. 2004), this does not have an important impact on the standard approach outlined here. The reason is simply that the satellite velocity dispersions are averaged over large parts of the host halo. The mass-anisotropy degeneracy predominantly plagues attempts to infer a mass profile from radially dependent kinematic data. Kinematic tracers that have the same radial distribution but different orbital anisotropies manifest different radial profiles of (projected) velocity dispersion. However, their total velocity dispersion, averaged over the entire system, has little to no dependence on the anisotropy. This is the reason why orbital anisotropy does not enter the virial theorem. As stated above, the cylindrical selection criteria used here to select secondaries only reach out to \(\sim 0.375\) times the virial radius, and the resulting satellite velocity dispersion is therefore not averaged over the entire system. However, as explicitly shown in van den Bosch et al. (2004), even in this case anisotropy only affects the kinematics at the level of a few per cent. Although this is advantageous if one is only interested in constraining halo mass, being able to constrain the orbital anisotropy opens up new avenues to test models for galaxy formation and evolution. The new method outlined below allows one to simultaneously constrain halo mass and orbital anisotropy.
3 METHODOLOGY

The Bayesian hierarchical method for analysing satellite kinematics presented here differs substantially from the ‘standard’ method outlined above. It is developed with the following goals in mind: (i) leave the data in its raw form as much as possible, particularly avoiding the use of summary statistics, binning, and/or stacking; (ii) include as much data as possible, by ensuring that the method can be applied to flux-limited samples, rather than only to volume-limited samples; and (iii) use a sufficiently flexible model that allows for proper treatment of mass-mixing and orbital anisotropy.

After selecting centrals and satellites (or rather, primaries and secondaries) using the same cylindrical isolation criteria as described in Section 2.1, we define a likelihood for the data given the model, and use an affine invariant ensemble sampler, within a Bayesian hierarchical framework, to constrain the posterior of the model parameters. As we detail below, this method achieves all three goals listed above.

3.1 Data format

Running the isolation criteria described in Section 2.1 yields a number of central–satellite pairs characterized by the following set of parameters: \( (L_c, L_s, z_c, \Delta V, R_p) \). Here \( L_c \) and \( L_s \) are the luminosities of the central and satellite, respectively, \( z_c \) is the redshift of the central, \( \Delta V \) is the line-of-sight velocity difference between central and satellite (equation 2), and \( R_p \) is the projected separation, which is related to the angular separation, \( \theta \), according to \( R_p = d_h(z_c) \theta \), with \( d_h \) the angular diameter distance. Finally, associated with each secondary is a weight, \( w \), that accounts for spectroscopic incompleteness in the survey, as described in Section 3.3.

For the purpose of our inference problem, we write our data vector \( D \) as the union of the \( N_c \) data vectors
\[
D_i = \{(\Delta V_{ij}, R_{p,ij})| j = 1, \ldots, N_s,i)\}.
\]

Here \( N_{s,i} \) is the number of secondaries associated with primary \( i \), and we have made it explicit that we only treat \( L_{c,i}, z_{c,i}, \) and \( N_{s,i} \) as \textit{conditionals} for the data \( \{(\Delta V_{ij}, R_{p,ij})| j = 1, \ldots, N_{s,i}\} \). In other words, we consider \( L_{c,i}, z_{c,i}, \) and \( N_{s,i} \) as ‘given’ and shall not use the distributions of these quantities as constraints on our likelihood. Rather, \textit{Basilisk} uses the luminosity function of all galaxies as an additional constraint (see Section 3.4).

The main reason for doing so is to make the method less sensitive to the detailed selection of centrals, which is difficult to model in detail. In particular, this approach makes \textit{Basilisk} insensitive to details regarding the \( \sigma_{200}(L) \) relation (equation 1) used to define the selection cylinders. Finally, we emphasize that we ignore the luminosities of secondaries in our inference, which instead relies entirely on their projected phase-space coordinates \( \Delta V \) and \( R_p \).

3.2 The inference problem

Our goal is to use \( D \) to constrain the galaxy-dark matter connection, which we characterize using the CLF, \( \Phi(L|M) \), and a model for the phase-space distribution of secondaries (satellites plus interlopers).

The CLF is split in a central component, \( \Phi_c(L|M) \), and a satellite component, \( \Phi_s(L|M) \), as detailed in Section 4.1 below. In particular, we seek to constrain the posterior distribution, \( P(\theta|D) \), where \( \theta \) is the vector that describes our model parameters, \( \theta_i \) (\( i = 1, \ldots, N_p \)). From Bayes theorem
\[
P(\theta|D) \propto L_{\text{SK}}(D|\theta) P(\theta),
\]
where \( P(\theta) \) is the prior probability distribution on the model parameters, and \( L_{\text{SK}}(D|\theta) \) is the likelihood of the satellite kinematics data given the model. Throughout we mainly use uniform, non-informative priors for our model parameters. If we make the reasonable assumption that the data for different primaries is independent, we have that
\[
L_{\text{SK}}(D|\theta) = \prod_{i=1}^{N_c} L(D_i|\theta)
\]
\[
= \prod_{i=1}^{N_c} \prod_{j=1}^{N_{s,i}} P(\Delta V_{ij}, R_{p,ij}|L_{c,i}, z_{c,i}, N_{s,i}, \theta).
\]

Computing the probability \( P(\Delta V, R_p|L_c, z_c, N_s) \), that a satellite galaxy in a halo at redshift \( z_c \), with a central galaxy of luminosity, \( L_c \), and with a total of \( N_s \) detected secondaries has projected phase-space parameters \( (\Delta V, R_p) \), requires knowledge of the gravitational potential, \( \Psi(x) \), in which the satellite is moving, as well as some knowledge regarding the phase-space distribution function (DF), \( f(x, v, t) \), of the population of satellites and interlopers. Throughout this work we make the assumption that satellite galaxies are a virialized, steady-state tracer of the gravitational potential well, which implies that \( f(x, v, t) = \hat{f}(x, v) \). In addition, we assume that dark matter haloes are spherical NFW profiles with a concentration–mass relation with zero scatter. This implies that \( \Psi(x) \) is completely specified by a single parameter, which we take to be the halo virial mass, \( M \). Our treatment of the DF is described in Section 4.2.2.

Since we assume that the satellite kinematics are governed solely by host halo mass, the likelihood for data \( D_i \), given model \( \theta \), can be factored as
\[
L(D_i|\theta) = \int dM P(M|L_{c,i}, z_{c,i}, N_{s,i})
\times \prod_{j=1}^{N_{s,i}} P(\Delta V_{ij}, R_{p,ij}|M, L_{c,i}, z_{c,i}),
\]
where, for the sake of brevity, we no longer explicitly write-out that \( P(M|L_{c,i}, z_{c,i}, N_{s,i}) \) and \( P(\Delta V, R_p|M, L_{c,i}, z_{c,i}) \) depend on \( \theta \). This equation describes a marginalization over halo mass, the prior for which is informed by \( L_c, z_c, \) and \( N_s \) according to the model \( \theta \). In addition, by putting the product operator inside the mass integral, as opposed to outside, we have made it explicit that all secondaries are assumed to belong to the same halo. Note also that the likelihood for \( \Delta V \) and \( R_p \) given a halo mass and redshift is conditional on the luminosity of the central, which arises from the fact that the isolation criteria used to select secondaries depend on the luminosity of the primary (see Section 2.1). As is evident from equation (6), and illustrated in Fig. 1, the halo masses for the individual primaries serve as latent variables, accentuating the hierarchical nature of our inference procedure.

---

\footnote{We have explicitly verified this by running \textit{Basilisk} on different (mock) samples, extracted from mock redshift surveys such as the ones described in Section 5 using different values for the coefficients in the \( \sigma_{200}(L) \) relation. The resulting posterior distributions are always in excellent, mutual agreement.}

\footnote{Throughout this paper, we define virial quantities according to the virial overdensities given by the fitting formula of Bryan & Norman (1998).}
and with the spectroscopic weights, $w_i$, which carry two indices.

Using Bayes theorem, we have that

$$P(M|L, z, N_i) = \frac{P(N_i|M, L, z) P(M|L, z)}{\int dM P(N_i|M, L, z) P(M|L, z)},$$

which allows us to write the log-likelihood for the satellite kinematics data as

$$\ln L_{\text{sk}}(D|\theta) = \sum_{i=1}^{N_c} (\ln G_i - \ln F_i),$$

with

$$G_i = \sum_k w_k F_{ik} \exp[Q_{ik}], \quad F_i = \sum_k w_k F_{ik},$$

and

$$F_{ik} = P(N_{i,j}|M_k, L_{c,i}, z_{c,i}) P(M_k, L_{c,i}, z_{c,i}).$$

Here $M_k$ and $w_k$ are the abscissas and weights of the Gaussian quadrature used to evaluate the mass-integral, and

$$Q_{ik} = \sum_{j=1}^{N_s} \sum_{l=1}^{N_s} P(\Delta V_{ij}, R_{p,l}|M_k, L_{c,i}, z_{c,i}).$$

As detailed in Section 3.5 and Appendix A, Gaussian quadrature has the advantage that the integrands are always evaluated at the same $M_k$, which allows for many quantities to be pre-computed, thereby greatly speeding up the Bayesian inference. Throughout we adopt a total of $N_s = 21$ quadrature points, which is sufficient to achieve accurate results.

What remains is to specify the probabilities $P(M, L, z), P(N_i|M, L, z)$, and $P(\Delta V, R_p|M, L, z)$, which we address in the following subsections.

### 3.2.1 The probability $P(M, L, z)$

Within the CLF formalism that we use to model the halo occupation statistics (see Section 4.1), the probability that a halo of mass $M$ at redshift $z$ hosts a central of luminosity $L$ is given by $\Phi_c(L|M, z)$. To take account of the fact that not every central is selected as a primary, we have that

$$P(M, L, z) = P(L|M, z) P(M, z) = C(M, L, z) \Phi_c(L|M, z) n(M, z).$$

Here $n(M, z)$ is the halo mass function at redshift $z$, and $C(M, L, z) = C(M|L, z) C(L, z)$ is a completeness function that expresses the fraction of central galaxies of luminosity $L$ residing in haloes of mass $M$ at redshift $z$ that are selected as primaries by our cylindrical isolation criteria. Since $P(M, L, z)$ appears in both the numerator and the denominator of equation (7), the factor $C(L, z)$ drops out, and the expression for $P(M, L, z)$ only depends on $C(M|L, z)$. In Section 5.2 we use detailed mock data to demonstrate that $C(M|L, z)$ is virtually independent of halo mass. Hence, we may simply model $P(M, L, z)$ as the product of $\Phi_c(L|M)$ and $n(M, z)$, without having to make any corrections for incompleteness (i.e. we can effectively set $C(M|L, z) = 1$). Throughout we compute the halo mass function using the method of Tinker et al. (2008) for our cosmology and halo mass definition, while we assume the CLF to be independent of redshift, at least over the redshift range considered in this study ($0.02 \leq z \leq 0.15$). The functional form that we use to describe $\Phi_c(L|M)$ is presented in Section 4.1.1.
The number of secondaries, \(N_s\), associated with a particular primary consists of both satellites (galaxies that belong to the same dark matter host halo as the primary), and interlopers (those that do not). We now derive the probability \(P(N_s|M, L, z)\), that a primary of luminosity \(L\), residing in a host halo of mass \(M\) at redshift \(z\), has a number of interlopers, \(N_{\text{int}}\), and satellites, \(N_{\text{sat}}\), such that \(N_{\text{int}} + N_{\text{sat}} = N_s\).

If we assume that the number of interlopers and the number of satellite galaxies are independent, then

\[
P(N_{\text{int}} + N_{\text{sat}} = N_s) = \sum_{n=0}^{N_s} P(N_{\text{int}} = n) P(N_{\text{sat}} = N_s - n) .
\]

If we furthermore assume that both interlopers and satellites obey Poisson statistics, we have that

\[
P(N_s|M, L, z) = \sum_{n=0}^{N_s} \frac{e^{-\lambda_{\text{sat}}}}{n!} \lambda_{\text{sat}}^n \left(\frac{\lambda_{\text{int}}}{N_s} \right)^n \left(\frac{N_s}{N_s} - n\right)! .
\]

The function \(\tilde{\phi}(\theta)\) avoids problems with fibre-collisions, as discussed in Section 3.3. Given that \(\Delta V_{\text{max}}\) is much larger than the extent of the halo in redshift space, we have that

\[
\lambda_{\text{int}} = \lambda_{\text{int}}(M, L, z) = f_{\text{ap}}(M, L, z_c) \int_{L_{\text{min}}(z_c)}^{\infty} \Phi_s(L|M) dL .
\]

Here \(\Phi_s(L|M)\) is the satellite component of the CLF (see Section 4.1.2) and \(f_{\text{ap}}\) is the aperture fraction, defined as the probability for true halo members (satellite galaxies) to fall within the secondary selection cylinder specified by \(L_{\text{ap}}\) and \(\Delta V_{\text{max}}\) (see Section 2.1). Given that \(\Delta V_{\text{max}}\) is much larger than the extent of the halo in redshift space, we have that

\[
f_{\text{ap}}(M, L, z_c) = 4\pi \int_{L_{\text{min}}(z_c)}^{\infty} \Phi_s(L|M) dL .
\]

Hence, we obtain the well-known result that the sum of two independent Poisson distributed random variables follows itself a Poisson distribution with a mean that is simply the sum of the means of its components.

We can rewrite this to avoid problems with fibre-collisions, as discussed in Section 3.3. Given that \(\Delta V_{\text{max}}\) is much larger than the extent of the halo in redshift space, we have that

\[
f_{\text{ap}}(M, L, z_c) = 4\pi \int_{L_{\text{min}}(z_c)}^{\infty} \Phi_s(L|M) dL .
\]

Hence, we obtain the well-known result that the sum of two independent Poisson distributed random variables follows itself a Poisson distribution with a mean that is simply the sum of the means of its components.

What remains is to specify \(\lambda_{\text{int}}\) and \(\lambda_{\text{sat}}\). The expectation value for the number of interlopers and satellites, respectively, \(\lambda_{\text{int}} \equiv \lambda_{\text{int}} + \lambda_{\text{sat}}\), and we have used the binomial identity in the last step. Hence, we obtain the well-known result that the sum of two independent Poisson distributed random variables follows itself a Poisson distribution with a mean that is simply the sum of the means of its components.

In order to derive the probability for true halo members (satellite galaxies) to fall within the aperture used to select secondaries around a primary of luminosity \(L_c\), is given by

\[
\lambda_{\text{sat}} = \lambda_{\text{sat}}(M, L_c, z_c) = f_{\text{ap}}(M, L_c, z_c) \int_{L_{\text{min}}(z_c)}^{\infty} \Phi_s(L|M) dL .
\]

Here \(\Phi_s(L|M)\) is the satellite component of the CLF (see Section 4.1.2) and \(f_{\text{ap}}\) is the aperture fraction, defined as the probability for true halo members (satellite galaxies) to fall within the secondary selection cylinder specified by \(L_{\text{ap}}\) and \(\Delta V_{\text{max}}\) (see Section 2.1). Given that \(\Delta V_{\text{max}}\) is much larger than the extent of the halo in redshift space, we have that

\[
f_{\text{ap}}(M, L_c, z_c) = 4\pi \int_{L_{\text{min}}(z_c)}^{\infty} \Phi_s(L|M) dL ,
\]

and

\[
\xi(r, R) = \begin{cases} 1 & \text{if } r \leq R \\ 1 - \sqrt{1 - R^2/r^2} & \text{otherwise.} \end{cases}
\]

Note that this neglects the (small) possibility that some primaries are satellites (i.e. what we refer to as impurity). As we demonstrate in Section 5.2, this impurity is small and does not have a significant impact on any of our results (see also Lange et al. 2019a). The full expression for \(\lambda_{\text{int}}(M, L_c, z_c)\) for our assumed functional forms for the CLF and the radial distribution of satellite galaxies is given in Appendix A.

For the interlopers, we model the expectation value as the product of an effective bias, \(b_{\text{eff}}\), and the expectation value for the number of galaxies with \(L_{\text{min}}(z_c) < L < L_c\) in a random, cylindrical volume, \(V_{\text{cyl}}(L_c, z_c)\), equal to that used to select the secondaries around the central of luminosity \(L_c\) at redshift \(z_c\):

\[
\lambda_{\text{int}} = \lambda_{\text{int}}(L_c, z_c) = b_{\text{eff}}(L_c, z_c) V_{\text{cyl}}(L_c, z_c) \bar{n}_{\text{gal}}(L_c, z_c) .
\]

Here

\[
\bar{n}_{\text{gal}}(L_c, z_c) = \int_{L_{\text{min}}(z_c)}^{L_c} dL \int_0^{\infty} \Phi(L|M) n(M, z) dM ,
\]

is the average number density of galaxies at redshift \(z_c\) that are fainter than \(L_c\) but brighter than the survey limit \(L_{\text{min}}(z_c)\). Since the cylinder used to select secondaries is specified by an opening angle \(\theta_{\text{ap}} = R_{\text{ap}}(L_c)/d_{\text{ap}}(z_c)\), and accounting for the cut-off radius \(R_{\text{min}} = R_{\text{cut}}(z_c) = \theta_{\text{ap}} d_{\text{ap}}(z_c)\), we have that

\[
V_{\text{cyl}}(L_c, z_c) = \omega_{\text{cyl}} \int_{z_c}^{1} \frac{d^2V}{d\omega dz} dz \approx \omega_{\text{cyl}} \frac{2\Delta V_{\text{max}}}{c} \frac{d^2V}{d\omega dz} (z_c) .
\]

Here \(\omega_{\text{cyl}} = \pm \Delta V_{\text{max}}/c\), the derivative \(d^2V/d\omega dz\) is the comoving volume element at redshift \(z\) corresponding to a solid angle \(d\omega\) and a depth \(dz\), and \(\omega_{\text{cyl}} = 2\pi (\cos \theta_c - \cos \theta_{\text{ap}})\) is the solid angle of the cylinder centred on the primary. Since \(R_{\text{ap}}(L_c) \ll d_{\text{ap}}(z_c)\), we have that both \(\theta_c\) and \(\theta_{\text{ap}}\) are small, which implies that to good approximation

\[
V_{\text{cyl}}(L_c, z_c) = \pi \left[ R_{\text{max}}^2 - R_{\text{min}}^2 \right] \frac{2\Delta V_{\text{max}}}{H(z_c)} (1 + z_c)^2 ,
\]

with \(H(z_c)\) the Hubble parameter. What remains is to model the effective bias, describing how the number density of interlopers around primaries is enhanced or suppressed relative to that in a random volume. We simply model this effective bias as having independent power-law dependences on the luminosity and redshift of the central, i.e.

\[
b_{\text{eff}}(L_c, z_c) = n_0 \left( \frac{L_c}{10^{10.5} L_{\odot}} \right)^{\eta_1} (1 + z_c)^{\eta_2} .
\]

Here \(n_0\), \(\eta_1\), and \(\eta_2\) three free parameters that fully specify our interloper-model, and whose values are to be determined from the data.

### 3.2.3 The probability \(P(\Delta V, R_p|M, L, z)\)

Since interlopers and satellites have distinct phase-space distributions, we write

\[
P(\Delta V, R_p|M, L, z) = f_{\text{int}} P_{\text{int}}(\Delta V, R_p|L, z) + [1 - f_{\text{int}}] P_{\text{sat}}(\Delta V, R_p|L, z)
\]

with

\[
f_{\text{int}} = f_{\text{int}}(M, L, z) = \frac{\lambda_{\text{int}}(L_c, z_c)}{\lambda_{\text{int}}(M, L, z)} .
\]
the interloper fraction. We assume that interlopers have a constant projected number density and a uniform distribution in line-of-sight velocity.\footnote{Although a clear oversimplification (see Section 5.2), this does not significantly impact our inference regarding the galaxy–dark matter connection.} This implies that

\[ P_{\text{int}}(\Delta V, R_p | L, z) = \frac{R_p}{\Delta V_{\text{sec}} \left( R_{\text{max}}^2 - R_{\text{min}}^2 \right)}, \]  

(26)

which is properly normalized, i.e.

\[ \int_{-\Delta V_{\text{sec}}_{\text{min}}}^{+\Delta V_{\text{sec}}_{\text{max}}} d\Delta V \int_{R_{\text{min}}}^{R_{\text{max}}} dR_p \, P_{\text{int}}(\Delta V, R_p | L, z) = 1. \]  

(27)

Finally, the probability \( P_{\text{int}}(\Delta V, R_p | M, L, z) \) is determined by our detailed model for the phase-space distribution of satellite galaxies, which is discussed in detail in Section 4.2.2 below.

### 3.3 Correction for fibre collisions

As demonstrated in Lange et al. (2019a), it is important to include in the analysis of satellite kinematics a correction for fibre-collision induced incompleteness in the spectroscopic data used. In what follows, we use the SDSS Main Galaxy Sample as a characteristic example. In the SDSS, spectroscopic fibres cannot be placed simultaneously on a single plate for objects separated by less than 55 arcsec (Blanton et al. 2003). Although some galaxies are observed with multiple plates, yielding spectroscopic redshifts even for close pairs, roughly 65 per cent of galaxies with a neighbour within 55 arcsec lack redshifts due to this fibre collision effect. We use this fact to mimic fibre collisions in our mock data sets, as discussed in Section 5.2.

In order to correct the data for the presence of fibre collisions, we follow Lange et al. (2019a) and start by assigning each fibre-collided galaxy the redshift of its nearest neighbour (see Blanton et al. 2005; Zehavi et al. 2005). Although we use these during the identification of primaries,\footnote{As shown in Lange et al. (2019a), ignoring fibre-collided galaxies during the selection of primaries results in a much larger sample impurity.} during the subsequent analysis only primary-secondary pairs with spectroscopic redshifts for both are used. In addition, each galaxy is assigned a spectroscopic weight, \( w_{\text{spec}} \), that is computed as follows. For each galaxy we first count the number of galaxies, \( n \), brighter than \( m_r = 17.6 \) within a projected separation less than 55 arcsec. Next, for all galaxies in the survey with \( n \) neighbours, we compute the fraction, \( f_{\text{spec}} \), of those neighbours that have been successfully assigned a redshift. Finally, all galaxies with \( n \) neighbours are then assigned a spectroscopic weight equal to \( w_{\text{spec}} = 1/f_{\text{spec}} \).

In Lange et al. (2019a) we used these weights to compute fibre-collision-corrected satellite velocity dispersions, \( \sigma_{\text{sat}}(L_c) \), and projected surface densities, \( \Sigma_{\text{sat}}(R_p) \). This works extremely well, except on scales below the fibre-collision scale of \( \theta_c = 55 \) arcsec. Therefore, Lange et al. (2019a) decided to exclude all secondaries with a projected separation from their primary less than \( 60 \, h^{-1} \) kpc, which is roughly the fibre collision scale at the maximum redshift of their volume-limited sample. Using the Tier-2 mocks described below (Section 5.2), we have tested a number of different fibre-collision-correction schemes for \( B_{\text{sec}} \leq 1.1 \, \epsilon_{\text{sec}} \). We find that the following scheme works extremely well; rather than up-weighting the number of secondaries in the data, we down-weight the expectation value, \( \lambda_{\text{sat}} = \lambda_{\text{tot}} - \lambda_{\text{int}} \), for the number of secondaries in the model. In particular, we multiply \( \lambda_{\text{sat}} \) (see Section 3.2.2) with the correction factor

\[ f_{\text{corr},i} = \frac{N_{i,j}}{\sum_{j=1}^{N_{i}} w_{ij}}, \]  

(28)

where \( w_{ij} \) is the spectroscopic weight, \( w_{\text{spec}} \), for secondary \( j \) associated with primary \( i \). Since \( w_{ij} \geq 1 \) we have that \( f_{\text{corr}} < 1 \), thereby correcting the expected number of secondaries for the fact that some are lost as a consequence of fibre collisions. In addition, since correction for fibre collisions is extremely difficult on scales below the fibre-collision scale, we remove all secondaries with \( R_p < R_{\text{sat}}(\theta_c) \equiv d_{\theta}(\theta_c) \theta_c \), with \( \theta_c \) the redshift of primary \( i \). Tests with mock data show that this typically removes of order 5 (11) per cent of the secondaries when satellite galaxies are assumed to have the phase-space distribution of subhaloes (dark matter particles). Since most primaries only have a single secondary, this cut in \( R_p \) also reduces the number of primaries, by roughly the same percentage. Tests with detailed mock data sets indicates that this cut does not significantly affect the constraining power regarding the galaxy–halo connection (see Sections 5.2 and 5.3).

### 3.4 Additional observational constraints

#### 3.4.1 Primaries without secondaries

The data vector \( \mathbf{D} \) described thus far only contains primaries with at least one secondary. However, running the selection criteria over a spectroscopic redshift survey also yields a complementary data vector, \( \mathbf{D}_0 = (\{L_{c,i}, z_{c,i}\}) | i = 1, 2, ..., N_0 \) listing all \( N_0 \) primaries with zero secondaries. This additional data vector provides additional constraints on the galaxy–halo connection, in particular regarding the satellite component of the CLF, and we therefore include it in our analysis. Since \( N_0 \) is typically much larger than the number of primaries with at least one primary, \( N_s \), we bin this data using a \( 5 \times 5 \) uniformly-spaced grid in \( \log L_c, z_c \) covering the range \([9.5, 11.0] \) in \( \log L_c \) and \([0.02, 0.15] \) in \( z_c \). For each bin we compute the probability \( P_0 \equiv N_0/(N_0 + N_s) \) that a primary in that bin has zero secondaries. Since the error distribution of \( P_0 \) can be very non-Gaussian, the actual constraint that we use in our modelling is \( f_0 \equiv \log (N_0/N_s) = \log [P_0/(1 - P_0)] \). Assuming that both \( N_0 \) and \( N_s \) follow Poisson statistics, we compute the corresponding errors on \( f_0 \) as \( \sigma_{f_0} = \sqrt{1/N_0 + 1/N_s}/\ln 10 \).

We define the log likelihood corresponding to this data as

\[ \ln L_0 \equiv -\frac{1}{2} \sum_{i,j} \left( \frac{f_{ij} - f_{0}(L_{ij}, z_{ij})}{\sigma_{f_0}(L_{ij}, z_{ij})} \right)^2. \]  

(29)

Here \( \log L_c \) and \( z_c \) are the centres of the bins used to compute \( f_0 \), and \( f_0 \) is the corresponding model prediction. The latter is computed using \( f_0 = \log [P_0/(1 - P_0)] \) with

\[ P_{ij} = \int dM \, P(M | \{M_i, z_i \}, P(M | L_{ij}, z_{ij}) \) 

\[ = \frac{\int dM \, P(L_{ij} | M, z_{ij}) n(M, z_{ij}) e^{-\lambda_{M,z}}}{\int dM \, P(L_{ij} | M, z_{ij}) n(M, z_{ij})}, \]  

(30)

the probability that a primary with luminosity \( L_i \) at redshift \( z_i \) has zero secondaries. Here we have once again used the fact that the completeness, \( C(M | L, z) \), does not depend significantly on halo mass (see Section 3.2), and we assume that both satellites and interlopers follow Poisson statistics, such that \( P(N_s = 0 | M, L, z) = e^{-\lambda_{M,z}} \) with \( \lambda_{M,z} \) the sum of the expectation values for the number of satellites (equation 15) and the number of interlopers (equation 19).
As always, we evaluate the mass integrals in equation (30) using Gaussian quadrature, as detailed in Appendix A.

3.4.2 Galaxy number densities

Since our main goal is to constrain the galaxy–halo connection, it is also advantageous to include constraints from the overall number density of galaxies. In particular, the luminosity function provides important constraints on the CLF (e.g. van den Bosch, Yang & Mo 2003; Yang, Mo & van den Bosch 2003; Cooray & Milosavljević 2005; Cooray 2006), which greatly helps to tighten the posterior in our inference problem.

We follow Lange et al. (2019a,b) and use the number density of galaxies in 10, 0.15 dex bins in luminosity, ranging from $10^{9.5}$ to $10^{11} h^{-2} \text{M}_\odot$. For our model, these number densities are computed according to

$$n_{\text{gal}}(L_1, L_2) = \int_{L_1}^{L_2} dL \int_0^\infty \Phi(L|M) n(M, \bar{z}_{\text{surv}}) \, dM,$$

where $\bar{z}_{\text{surv}}$ is a characteristic redshift for the survey in question. For the mock data samples discussed in Section 5, which cover the redshift range [0.02, 0.15], we set $\bar{z}_{\text{surv}}$ equal to the redshift of the simulation output used to construct the mock. When analysing SDSS data (van den Bosch et al., in preparation), we adopt $\bar{z}_{\text{surv}} = 0.1$. We have verified that our results do not depend significantly on this choice; using $\bar{z}_{\text{surv}} = 0.05$ instead of 0.0 or 0.1 yields results that are virtually indistinguishable.

We include the data on $n_{\text{gal}}(L_1, L_2)$ in our inference problem by defining the corresponding log-likelihood

$$\ln L_{\text{KL}}(n_{\text{obs}}|\theta) = -\frac{1}{2} \left[ n(\theta) - n_{\text{obs}} \right] \Psi \left[ n(\theta) - n_{\text{obs}} \right].$$

Here $n_{\text{obs}}$ is the data vector for $n_{\text{gal}}$ for the ten bins in luminosity, $n(\theta)$ is the corresponding model prediction given by equation (31), and $\Psi$ is the precision matrix, which is the inverse of the covariance matrix. The latter is computed using 1000 SDSS-like mocks and the unbiased estimator as described in Lange et al. (2019a).

3.5 Numerical implementation

Probing the posterior $P(\theta|D)$ over our 17-dimensional parameter space requires millions of likelihood evaluations, each of which involves many numerical integrations (see Appendix A). In order to make this problem feasible, we follow Lange et al. (2019a,b) and perform the Bayesian inference under the assumption of a fixed normalized, radial number density distribution of satellite galaxies, $\bar{n}_{\text{sat}}(r|M, z)$, to be defined in Section 4.2.2 below. This has the advantage that $f_{\text{gal}}(M, L, z)$ and $P_{\text{gal}}(\Delta V, R_0|L, z)$ are all independent of the model, $\theta$, while $P_{\text{sat}}(\Delta V, R_0|L, z)$ only depends on a single anisotropy parameter (see Section 4.2). Combined with the fact that we perform the mass integration using a Gaussian quadrature with fixed abscissas, $M_i$, this implies that we only need to compute (and store) these quantities once for each primary and/or secondary. And the same applies for the halo mass function, $n(M, z)$, which appears in equations (9) and (10). The probabilities $P_{\text{sat}}(\Delta V, R_0|L, z)$ are computed using linear-interpolation over a grid of values that are pre-computed for different anisotropy parameters, as detailed in Section 4.2.3. As a consequence, a single evaluation of the full likelihood

$$L_{\text{tot}}(D + D_0 + n_{\text{obs}}|\theta) = L_{\text{SK}}(D|\theta) + L_0(D_0|\theta) + L_{\text{KL}}(n_{\text{obs}}|\theta),$$

for a mock data set with 5000 satellite galaxies, takes only of order 10 milliseconds using a single, run-of-the-mill CPU. This is sufficiently fast, that it easily allows one to run many different Monte Carlo Markov Chains for different assumptions regarding $\bar{n}_{\text{sat}}(r|M, z)$, or to find the best-fitting radial profile, marginalized over all other model parameters, using a straightforward $\chi^2$-minimization algorithm.

The method that we use to construct Monte Carlo Markov Chains is the affine invariant ensemble sampler proposed by Goodman & Weare (2010). This is the same method that is used by the popular PYTHON code EMCEE developed by Foreman-Mackey et al. (2013) and we refer the interested reader to these two papers for details. Throughout we use 1000 walkers and the proposal density advocated by Goodman & Weare (2010). This results in typical acceptance fractions between 0.3 and 0.4. We start the walkers in a small region of parameter space centred on the best-fitting model obtained during the ‘burn-in’ stage. Throughout we adopt a Metropolis–Hastings burn-in of 10 000 steps in which we use independent Gaussian proposal distributions for each model parameter. The best-fitting model at the end of the burn-in period is always close to the best-fitting model subsequently obtained from the entire MCMC. Most of our MCMC chains contain 5 million elements (post burn-in), corresponding to 5000 steps for each of the 1000 walkers, and are well converged.

4 MODEL INGREDIENTS

This section describes the model ingredients to be used in combination with the method outlined in the previous section. These include a model for the galaxy–halo connection, and a model for the phase-space distributions of central and satellite galaxies as a function of halo mass.

4.1 Galaxy–halo connection

We model the galaxy occupation using the CLF (van den Bosch et al. 2003; Yang et al. 2003) approach. The CLF, $\Phi(L|M)dL$, specifies the average number of galaxies with luminosities in the range $L \pm dL/2$ residing in a dark matter halo of virial mass $M$. As already eluded to in Section 3.2, we assume that galaxies can be separated into centrals and satellites, each with their own CLF,

$$\Phi(L|M) = \Phi_c(L|M) + \Phi_s(L|M).$$

Here, as always, subscripts ‘c’ and ‘s’ refer to central and satellite, respectively. These two populations are described in more detail below.

4.1.1 Central galaxies

The CLF of centrals is parametrized using a lognormal distribution,

$$\Phi_c(L|M)dL = \frac{\log e}{\sqrt{2\pi}\sigma_c} \exp \left[-\frac{(\log L - \log \bar{L}_c)^2}{2\sigma_c^2}\right] dL.$$

The mass dependence of the median luminosity, $\bar{L}_c$, is parametrized by a broken power law:

$$\bar{L}_c(M) = L_0 \left(\frac{M}{M_1}\right)^{\gamma_1} \left(1 + \frac{M}{M_1}\right)^{-\gamma_2},$$

which is characterized by three free parameters; a normalization, $L_0$, a characteristic halo mass, $M_1$, and two power-law slopes, $\gamma_1$ and $\gamma_2$. 

MN148 488, 4984–5013 (2019)
Motivated by the fact that several hydrodynamical simulations suggest that the scatter, $\sigma_c$, increases with decreasing halo mass (e.g. Sawala et al. 2017; Pillepich et al. 2018), we allow for a mass-dependent scatter using

$$\sigma_c(M) = \begin{cases} \sigma_{12} & \text{if } \log M \leq 12 \\ \sigma_{12} + \frac{\log M - 12}{2} (\sigma_{14} - \sigma_{12}) & \text{if } 12 < \log M < 14 \\ \sigma_{14} & \text{if } \log M \geq 14 \end{cases} \quad \text{(37)}$$

Hence, the scatter is characterized by two free parameters, $\sigma_{12}$ and $\sigma_{14}$, that indicate the lognormal scatter in haloes of mass $M = 10^{12}$ and $10^{14} h^{-1} M_{\odot}$, respectively.

### 4.1.2 Satellite galaxies

We model the satellite CLF as a modified Schechter function:

$$\Phi_s(L|M) = \frac{\phi_s^*}{L^{*}_s} \left( \frac{L}{L^{*}_s} \right)^{\alpha_s} \exp \left[ - \left( \frac{L}{L^{*}_s} \right)^2 \right]. \quad \text{(38)}$$

Thus, the luminosity function of satellites, for a given halo mass, follows a power law with slope $\alpha_s$ with an exponential cut-off above a critical luminosity, $L^{*}_s(M)$, which is related to the characteristic luminosity of central galaxies in haloes of the same mass according to

$$L^{*}_s(M) = 0.562 L_{c}(M). \quad \text{(39)}$$

As shown in Yang et al. (2009), this relation provides a good description of the luminosities of centrals and satellites as inferred from the SDSS galaxy group catalogue of Yang et al. (2007). Motivated by the results of Yang et al. (2008), who found evidence for a steeper slope (more negative value of $\alpha_s$) in more massive groups, we allow for a mass-dependent power-law slope using

$$\alpha_s(M) = \begin{cases} \alpha_{12} & \text{if } \log M \leq 12 \\ \alpha_{12} + \frac{\log M - 12}{2} (\alpha_{14} - \alpha_{12}) & \text{if } 12 < \log M < 14 \\ \alpha_{14} & \text{if } \log M \geq 14 \end{cases} \quad \text{(40)}$$

Hence, similar to the scatter, the logarithmic slope is characterized by two free parameters, $\alpha_{12}$ and $\alpha_{14}$, that indicate the slope in haloes of mass $M = 10^{12}$ and $10^{14} h^{-1} M_{\odot}$, respectively. Finally, the normalization $\phi_s^*(M)$ is parametrized by

$$\log \left[ \phi_s^*(M) \right] = b_0 + b_1 \log M_{12} + b_2 (\log M_{12})^2, \quad \text{(41)}$$

where $M_{12} = M/(10^{12} h^{-1} M_{\odot})$.

Note that our particular characterization of the CLF is very similar to that adopted in a number of previous studies (Cacciato et al. 2009, 2013; More et al. 2009a; van den Bosch et al. 2013; Lange et al. 2019a, b).

### 4.2 Phase-space distributions

The CLF described above specifies the abundance of central and satellite galaxies as function of luminosity and halo mass. We now describe our model for the positions and velocities of these galaxies with respect to their host halo.

---

8 We have tested that treating the ratio $L^*_s(M)/L_c(M)$ as a free parameter in Basilisk does not significantly impact any of our results.

4.2.2 Satellite galaxies

In the case of satellite galaxies, the phase-space model determines the probability $P_{sat}(\Delta V, R_p|M, z, L_s)$, which characterizes the satellites’ projected phase-space distribution and plays the key role in our likelihood evaluation (see Section 3.2.3). In the following, we assume that satellites have a spherically symmetric radial profile $n_{sat}(r|M, z)$. It is known that satellite populations of individual haloes can have varying degrees of asphericity (e.g. Zentner et al. 2005b; Azzaro et al. 2007; Wang et al. 2008). However, since we combine the data from a large number of individual haloes with random orientations, this assumption of spherical symmetry will not affect our inferences substantially. We assume that the radial profile as a function of the radial distance $r$ from the halo centre is given by a generalized Navarro–Frenk–White (gNFW) profile,

$$n_{sat}(r|M, z) \propto \left( \frac{r}{R} \right)^{-\gamma} \left( 1 + \frac{r}{R_{\lambda}} \right)^{-\gamma - 3}. \quad \text{(42)}$$

Here $R$ and $\gamma$ are free parameters and $R_{\lambda}$ is the scale radius of the dark matter halo, which is related to the halo virial radius via the concentration parameter $c_{con} = R_{\lambda}/R$. This gNFW profile has sufficient flexibility to adequately describe a wide range of radial profiles, from satellites being unbiased tracers of their dark matter halo ($\gamma = 1$, to cored profiles that resemble the radial profile of surviving subhaloes in numerical simulations ($\gamma = 0$, $R \sim 2$). This also brackets the range of observational constraints on the radial distribution of satellite galaxies in groups and clusters (e.g. Carlberg et al. 1997; van der Marel et al. 2000; Lin, Mohr & Stanford 2004; Yang et al. 2008b; Chen 2008; More et al. 2009a; Watson et al. 2010, 2012; Guo et al. 2012; Cacciato et al. 2013; Lange et al. 2019b).

We also assume that the host haloes of satellite galaxies are spherical NFW haloes that are completely specified by their mass, i.e. we adopt the concentration–mass relation of Macciò, Dutton & van den Bosch (2008) without scatter. In the most general case, under the assumption of spherical symmetry, one then has that

$$P_{sat}(\Delta V, R_p|M, L, z) = \frac{2\pi R_p}{\lambda_{sat}} \int \int dz \int d\Omega f(E, J^2). \quad \text{(43)}$$

Here $z$ is the coordinate along the line of sight, not to be confused with the redshift $z$, $\lambda_{sat} = \lambda_{sat}(M, L, z)$ is the expectation value for the number of satellite galaxies (equation 15), and $f(E, J^2)$ is the DF, which for a spherically symmetric system is a function of energy, $E$, and angular momentum, $J = |J|$. Typically one of three assumptions is made: (i) the DF is isotropic, such that $f(E, J^2) = f(E)$, (ii) the DF depends on energy and angular momentum only through the quantity $Q = E + J^2/(2c^2)$, where $r_s$ is a free parameter known as the ‘anisotropy radius’, such that $f(E, J^2) = f(Q)$, or (iii) the DF is separable, such that $f(E, J^2) = g(E)h(J^2)$. Models that make assumption (ii) are known as Osipkov–Merritt
models (Osipkov 1979; Merritt 1985) and have an anisotropy profile that increases from isotropic in the centre ($r < r_a$) to radially anisotropic at larger radii. Models that make assumption (iii) can have constant anisotropy or a radially varying anisotropy (e.g. Louis 1993; Cuddeford & Louis 1995; Wojtak et al. 2009). In each case, the computation of the DF for a given $n_{sat}(r|M, z)$ and halo potential, $\Psi(r|M, z)$, involves at least a 1D integration. Together with equation (43), which involves a 3D integration, this makes the computation of $P(\Delta V, R_p|M, L, z)$ prohibitively expensive. In addition, we lack a good prior on the functional form of $\beta (E, P)$, further disincentivizing the use of equation (43).

We therefore opt for an alternative, approximate method to compute $P(\Delta V, R_p|M, L, z)$. Rather than using DFs, we write

$$P_{\text{sat}}(\Delta V, R_p|M, L, z) = P(R_p|M, L, z) P(\Delta V|R_p, M, z),$$

(44)

and make the simplified assumption that $P(\Delta V|A|M, z)$ is a Gaussian with a projected velocity dispersion, $\sigma_{\text{los}} = \sigma_{\text{los}}(R_p|M, L, z)$:

$$P(\Delta V|R_p, M, z) = \frac{1}{\sqrt{2\pi} \sigma_{\text{los}}(R_p|M, L, z)} \exp \left[-\frac{(\Delta V)^2}{2\sigma_{\text{los}}^2(R_p|M, L, z)}\right],$$

(45)

The division by the error function is required by the normalization condition, and the fact that our selection criterion only includes satellites for which $|\Delta V| < \Delta V_{\text{max}}$. Although there is no a priori reason for the assumption of Gaussianity, the LOSVDs of (non-rotating) dynamical systems often are very close to Gaussian. Indeed, as we demonstrate in this paper, this assumption is adequate for the purpose of constraining the galaxy–halo connection and allows for an extremely efficient computation of $P(\Delta V, R_p|M, L, z)$.

The probability $P(R_p|M, L, z)$ in equation (44) derives from the (normalized) radial number density distribution of satellite galaxies, $\bar{n}_{\text{sat}}(r|M, z)$, according to

$$P(R_p|M, L, z) = \frac{2 \pi R_p \Sigma(R_p|M, z)}{f_{\text{ap}}(M, L, z)},$$

(46)

where

$$\Sigma(R_p|M, z) = \int_{R_p}^{\infty} \bar{n}_{\text{sat}}(r|M, z) \frac{r \, dr}{\sqrt{r^2 - R_p^2}}$$

(47)

is the projected, normalized number density distribution of satellite galaxies. The division by the aperture fraction is required by the normalization condition. The projected velocity dispersion is related to the intrinsic, radial velocity dispersion, $\sigma^2_{r}(r|M, z)$, according to the following Abel integral

$$\sigma^2_{\text{los}}(R_p|M, z) = \frac{2}{\Sigma(R_p)} \int_{R_p}^{\infty} \left[1 - \beta(r) \frac{R_p^2}{r^2}\right] \frac{\bar{n}_{\text{sat}}(r|M, z) \sigma^2_{r}(r|M, z)}{\sqrt{r^2 - R_p^2}} \frac{r \, dr}{\sqrt{r^2 - R_p^2}},$$

(48)

where

$$\beta(r) = 1 - \frac{\sigma^2_{\text{los}}(r)}{2\sigma^2_{r}(r)},$$

(49)

is the local anisotropy parameter, relating the tangential and radial velocity dispersions.

The radial velocity dispersion follows from the Jeans equation. If we assume a constant orbital anisotropy, such that $\beta(r) = \beta$, then this Jeans equation reduces to

$$\sigma^2_{r}(r|M, z) = \frac{G}{r^{2\beta}} \frac{1}{\bar{n}_{\text{sat}}(r|M, z)} \times \int_{r}^{\infty} \frac{\bar{n}_{\text{sat}}(r'|M, z)}{M(r') \, dr'},$$

(50)

where $M(r)$ is the halo mass enclosed by radius $r$. In addition to these constant anisotropy (hereafter CA) models, we will also consider Osipkov–Merritt (hereafter OM) models, for which the anisotropy parameter scales with radius as

$$\beta(r) = \frac{r^2}{r^2 + z^2}.$$  

(51)

Hence, the orbits are close to isotropic ($\beta \approx 0$) at small radii ($r \ll r_a$), and become more and more radially anisotropic at larger radii. This is very reminiscent of the orbital anisotropy of dark matter particles (e.g. Ascasibar & Gottlöber 2008; Wojtak et al. 2008; Wojtak, Gottlöber & Klypin 2013) and subhaloes (e.g. Diemand et al. 2004) in numerical simulations, and is therefore a realistic model to describe the kinematics of satellite galaxies (but see Cuesta et al. 2008; Sawala et al. 2017, and Appendix C).

For an OM-model, the Jeans equation for the radial velocity dispersion becomes

$$\sigma^2_{r}(r|M, z) = \frac{G}{r^{2\beta}} \frac{1}{\bar{n}_{\text{sat}}(r|M, z)} \times \int_{r}^{\infty} \frac{\bar{n}_{\text{sat}}(r'|M, z) M(r') \, dr'}{r'^2 + z^2}.$$  

(52)

(Merritt 1985). Note that an OM-model with $r_a \rightarrow \infty$ is equivalent to a CA-model with $\beta = 0$ (both are isotropic throughout). Detailed expressions for $\sigma^2_{\text{los}}(R_p|M, z)$ for a tracer population with a gNFW profile orbiting within an NFW host halo are given in Appendix A, for both the CA and OM-model.

Finally, in order to account for non-zero redshift errors in the data, the line-of-sight velocity dispersion used in equation (45) is the quadratic sum of $\sigma_{\text{los}}$ given by equation (48) plus two times the typical redshift error, $\sigma_{\text{err}}$.

### 4.2.3 Summary

Altogether, our model has a total of 15 free parameters: 11 to describe how galaxies populate dark matter haloes, of which 6 describe the CLF of centrals ($M_1, L_0, \gamma_1, \gamma_2$, $\sigma_1$, $\sigma_2$) and 5 quantify the satellite component ($\alpha_1$, $\alpha_1^*$, $b_0$, $b_1$, $b_2$), 3 parameters to specify the effective bias that characterizes the number density of interlopers ($\eta_0$, $\eta_1$, $\eta_2$), and 1 parameter to characterize the orbital anisotropy (either $\beta$ or $r_a$).

As already mentioned in Section 3.5, the two additional parameters, $R$ and $\gamma$, that describe the radial profile of satellite galaxies, are not treated as free parameters in the MCMC analysis. Rather, we separately constrain the posteriors for different sets of $(R, \gamma)$, as this vastly increases the speed. In the case of the anisotropy parameter, we pre-compute $P_{\text{sat}}(\Delta V, R_p|M, L, z)$ over a range of anisotropy parameters, and then use linear-interpolation to compute these quantities for given $\beta$ or $r_a$. In the case of the CA-model, we pre-compute the matrix $M_{ijkl} \equiv P_{\text{sat}}(\Delta V_{ij}, R_{p,k})|M_{i}, L_{e,j}, z_{c,i}, \beta_i)$.
for 10 values of $\beta$ that uniformly sample the parameter
\begin{equation}
B \equiv -\log[1 - \beta],
\end{equation}
over the interval $[-1.0, 1.0]$, which corresponds to $\beta$ covering the range $[-9.0, 0.9]$. In the case of the $\Omega M$-model, we pre-compute $\mathcal{M}_{ij}(\rho) \equiv \mathcal{P}_{\text{sat}}(\Delta V_{ij}, R_{ij}|M_1, L_{\text{c},1}, \gamma_{ij}, \eta_{ij})$ using 10 values of $\rho$, that uniformly sample $\log[r_i/r_0]$ over the range $[-1.0, 1.5]$. The Table 1 lists all our model parameters used to quantify the galaxy–halo connection, the number density of interlopers, and the phase-space distribution of satellite galaxies. It also lists their prior ranges used when fitting data and their default values used to create the mock catalogues described in Section 5. These default values are similar to the constraints inferred by More et al. (2011) and Cacciato et al. (2009) analysing satellite kinematics, galaxy–galaxy lensing, and galaxy clustering in the SDSS, and therefore give a realistic description of the galaxy–halo connection at low redshift. Note that we adopt non-informative, uniform priors for all parameters except $\gamma_1$, for which instead we use a Gaussian prior with a mean of 3.5 and a standard deviation of 0.2. This is motivated by the fact that the slope of the $L_*(M_\delta)$ relation is poorly constrained at the low mass end, which in turn owes to the fact that we have very few satellites for primaries with $L_\delta \lesssim 10^{10.5}h^{-2}L_\odot$ (see Fig. 2). The value of 3.5 is consistent with constraints from a variety of independent studies (e.g. Yang et al. 2009; Cacciato et al. 2013; Lange et al. 2019b), all of which find best-fitting values in the range $3.2 \lesssim \gamma_1 \lesssim 3.7$. We have verified that this prior has no impact on any of our results, other than restricting the posterior constraint on $\gamma_1$.

5 VALIDATION

We now proceed with a three-tiered validation process of our method. In Tier-1 we use highly idealized mocks in which we draw dark matter haloes from an analytical halo mass function, and in which we assume perfect identification of centrals and satellites and ignore survey incompleteness effects (i.e. fibre collisions). In addition, we use the same model to assign satellites and interlopers their phase-space coordinates as used in our analysis. In Tier-2 mocks we add complexity by constructing the mocks from dark-matter-only $N$-body simulations, yielding more realistic interlopers, and by including spectroscopic incompleteness due to fibre collisions and other redshift failures. When populating the dark matter haloes with mock galaxies, though, we still adopt the same analytical model for their phase-space distributions as in our model. Finally, in the third and final tier, we construct mocks by assigning satellite galaxies the locations and velocities of dark matter subhaloes in the $N$-body simulations. All mocks are constructed to resemble the SDSS Main Galaxy Sample. In particular, we adopt an apparent magnitude limit of $m_\text{lim} = 17.6$ in the $r$ band and we assume that galaxy redshifts have a velocity error of $\sigma_\text{err} = 15$ km s$^{-1}$ (Guo et al. 2015b). For the Tier-2 and Tier-3 mocks we also mimic the SDSS footprint on the sky and model the impact of incompleteness due to fibre collisions. We now describe each tier in detail, and highlight some important aspects of our modelling approach that provide valuable insight.

5.1 Tier-1: idealized mocks

In the first step of our validation process we consider highly idealized mock data sets that are constructed as follows:

(i) Draw a redshift, $z_c$, in the range $[0.02, 0.15]$, sampled according to the corresponding comoving volume, i.e. $P(z) \propto \text{d}V/\text{d}z$, and compute the corresponding luminosity limit, $L_{\text{lim}}(z_c)$, defined as the minimum luminosity for a galaxy at that redshift to have an apparent magnitude brighter than the survey limit $m_{\text{lim}}$.

(ii) At this redshift, draw a halo mass from the halo mass function, $n(M, z_c)$ covering the range $\log [M/\text{h}^{-1}M_\odot] \in [10, 15]$. These mass limits are purely numerical convenience; the upper limit is large enough that the abundance of more massive haloes is sufficiently small, while the lower limit is low enough that the probability that its central is brighter than the apparent magnitude limit of the survey is negligible.

(iii) Draw a luminosity for the central galaxy from $\Phi_c(L_c|M)$. If $L_c < L_{\text{lim}}(z)$ or $L_c < 10^{10.5}h^{-2}L_\odot$, discard the halo and galaxy and go back to (i). The reason for discarding primaries with $L_c < 10^{10.5}h^{-2}L_\odot$ is that the number is small and virtually all their secondaries are interlopers. Hence, they add little in terms of constraining power for the model.

(iv) Compute the aperture radius, $R_{\text{ap}}(L_c)$, and the expectation value for the number of interlopers in the secondary-selection-cylinder, $\lambda_{\text{int}} \equiv \langle N_{\text{int}}(L_c, z) \rangle$. The latter is computed using equation (19) with the effective bias set to unity, i.e. $(\theta_0, \eta_1, \eta_2) = (1, 0, 0)$. Draw the actual number of interlopers from a Poisson distribution with a mean equal to $\lambda_{\text{int}}$, and for each of these interlopers draw a $R_p$ and $\Delta V$ with respect to the primary from the phase-space probability distribution given by equation (26).

(v) Use equation (15) with $f_{\text{ap}} = 1$ to compute the expectation value, $\lambda_{\text{sat}}$, for the number of satellites in the halo in question, and draw the actual number of satellites from a Poisson distribution with a mean equal to $\lambda_{\text{sat}}$. For each satellite draw a position within the halo, assuming a spherically symmetric distribution characterized by $n_{\text{sat}}(r|L_c, z)$. Next, compute the local velocity dispersion, $\sigma^2(r|L_c, z)$, using the Jeans equation for an isotropic DF (equation 50 with $\beta = 0$), and draw the component of the velocity vector along the line of sight, $\Delta V$, from a Gaussian with a velocity dispersion equal to $\sigma^2(r|L_c, z)$. Compute the corresponding $R_p$, draw a luminosity for the satellite, $L_s$, from $\Phi_s(L_s|L_c)$, and correct $\Delta V$ for redshift errors by adding a random velocity drawn from a Gaussian with a dispersion of $\sqrt{2}\sigma_{\text{err}}$. If $L_s < L_c$ and $R_p < R_{\text{ap}}(L_c)$ and $|\Delta V| < \Delta V_{\text{lim}}$, add this satellite to the list of secondaries for the primary in question.

(vi) Repeat this procedure until the total number of secondaries exceeds the target number. Note that we keep track of primaries that end up with zero secondaries (zero interlopers and zero satellites), which we use to compute $f_d(L_c, z)$ for our mock data set.

We use this method to construct a mock that has 5000 secondaries around 2379 primaries covering the redshift range $[0.02, 0.15]$. The mock also contains 15 361 primaries with zero secondaries. Satellites are assumed to be unbiased, isotropic tracers of the mass distribution of their host haloes (i.e. $\beta = 0$), and the halo occupation statistics are given by a CLF with the fiducial parameters listed in Table 1. Hence, this mock data set is generated using exactly the same model as used to compute the likelihood, and the results of the likelihood analysis discussed below therefore merely serves as a sanity check of Basilisk’s inference procedure.

Fig. 2 shows some properties of this mock data set. The upper-left panel plots the luminosity as a function of redshift. The two vertical lines mark the redshift limits, while the red, solid curve corresponds to an apparent $r$-band magnitude of 17.6. Black and cyan dots correspond to primaries and secondaries, respectively. The red dotted–dashed lines demarcate the volume-limited survey that was used in the satellite kinematics analyses of More et al. (2009a), More et al. (2011), and Lange et al. (2019b). The method developed here can
Table 1. Model parameters that quantify the galaxy–halo connection, the interloper fraction, and the phase-space distribution of satellite galaxies within their host haloes. Column (2) gives a description, while column (3) lists the equation in which the parameter is introduced. Column (4) indicates the prior used in our Bayesian analysis, where $U[a, b]$ indicates a uniform prior over the range $[a, b]$ and $G[a, b]$ indicates a Gaussian prior with a mean $a$ and dispersion $b$. Finally, column (5) indicates the default value used when constructing our mock data sets. Note that for $\eta_0$, $\eta_1$, and $\eta_2$, the parameters characterizing the effective bias parameter of the interlopers, these fiducial values only apply to the Tier-1 mock.

| Parameter | Description | Equation | Prior | Default |
|-----------|-------------|----------|-------|---------|
| $\log M_1/(h^{-1}M_\odot)$ | Characteristic mass of mass–luminosity relation for centrals | (36) | $U[8.0, 15.0]$ | 11.20 |
| $\log L_0/(h^{-2}L_\odot)$ | Normalization of mass–luminosity relation for centrals | (36) | $U[8.0, 12.0]$ | 9.95 |
| $\gamma_1$ | Low-mass slope of mass–luminosity relation for centrals | (36) | $G[3.5, 0.2]$ | 3.5 |
| $\gamma_2$ | High-mass slope of mass–luminosity relation for centrals | (36) | $U[0.0, 5.0]$ | 0.25 |
| $\sigma_{12}$ | Logarithmic scatter in luminosity at a halo of mass $10^{12} h^{-1} M_\odot$ | (37) | $U[0.001, 1.0]$ | 0.15 |
| $\sigma_{14}$ | Logarithmic scatter in luminosity at a halo of mass $10^{14} h^{-1} M_\odot$ | (37) | $U[0.001, 1.0]$ | 0.15 |
| $\alpha_{12}$ | The logarithmic slope of the satellite CLF at a halo of mass $10^{12} h^{-1} M_\odot$ | (40) | $U[-2.0, 2.0]$ | $-1.2$ |
| $\alpha_{14}$ | The logarithmic slope of the satellite CLF at a halo of mass $10^{14} h^{-1} M_\odot$ | (40) | $U[-2.0, 2.0]$ | $-1.2$ |
| $b_0$ | Determines the normalization of the satellite CLF | (41) | $U[-3.0, 3.0]$ | 1.5 |
| $b_1$ | Determines the normalization of the satellite CLF | (41) | $U[-3.0, 3.0]$ | 0.2 |
| $b_2$ | Determines the normalization of the satellite CLF | (41) | $U[-3.0, 3.0]$ | 0.2 |
| $\eta_0$ | Normalization of effective bias of interlopers | (23) | $U[0.0, 100.0]$ | 0.0 |
| $\eta_1$ | Power-law dependence of effective bias of interlopers on luminosity of primary | (23) | $U[-2.0, 2.0]$ | 0.0 |
| $\eta_2$ | Power-law dependence of effective bias of interlopers on redshift of primary | (23) | $U[-2.0, 2.0]$ | 0.0 |
| $R$ | Ratio of scale radius of satellite distribution wrt that of dark matter | (42) | $U[0.1, 1.0]$ | 1.0 |
| $\gamma$ | Central slope of radial profile of satellite distribution | (42) | $U[0.0, 1.5]$ | 1.0 |
| $\beta$ | Anisotropy parameters ($\Lambda$CDM models) | (49) | $U[-9.0, 0.9]$ | 0.0 |
| $\log [r_a/r_s]$ | Anisotropy radius ($\Omega_M$ models) | (51) | $U[-1.0, 1.5]$ | $-$ |

Figure 2. Tier-1 mock data. The upper left-hand panel plots luminosity as a function of redshift, with black and cyan dots indicating primaries and secondaries, respectively. The solid red curve indicates the apparent magnitude limit of the (mock) survey, while the red dot–dashed lines mark the volume-limited subsample used in previous SDSS-based analyses of satellite kinematics (in particular More et al. 2009a, 2011; Lange et al. 2019b). The latter is shown to underscore that Basilisk bases its analysis on much more data. The upper right-hand panel plots the luminosity of the primary as a function of its halo mass, with the red solid line marking the expectation value $\langle L_c \mid M \rangle$ corresponding to the CLF used to construct the mock. Note the large amount of scatter in $M$ at given $L_c$, highlighting the importance of properly taking ‘mass-mixing’ into account. Finally, the lower panels plot $\Delta V$ as a function of the primary luminosity (left) and the projected separation between primary and secondary (right). Note the obvious increase in the variance of $\Delta V$ with increasing $L_c$, reflecting that more luminous primaries reside in more massive haloes, the contribution of individual clusters and large groups at the luminous end, and the uniform ‘background’ due to interlopers.
be applied to a flux-limited sample, thereby greatly increasing the amount of data that can be used. Note that, when evaluating $\ln L_0$ using equation (29), we only sum over those $\log L_c$ and $z_c$ bins that lie entirely above the flux limit of the survey, i.e. for which $\log L_i - \Delta \log L/2 > \log L_{\text{min}}(z_i + \Delta z/2)$ with $\Delta \log L = 0.15$ and $\Delta z = 0.026$ the bin widths. This is the case for 16 out of the total of 25 bins. The upper-right panel of Fig. 2 plots the luminosity of primaries in the mock as a function of their halo mass, with the red, solid line indicating the expectation value, $\log \langle L_c | M \rangle$, computed from the central CLF, $\Phi_c(L|M)$, used to construct the mock. Although the mock only assumes a scatter in $\log L_c$ at fixed halo mass of 0.15 dex (see Table 1), it is clear that at fixed $L_c$, the primaries cover a huge range in halo mass (see More et al. 2009b, for a detailed discussion). Properly accounting for this ‘mass-mixing’ is one of the main challenges in satellite kinematic. Finally, the lower two panels of Fig. 2 plot $\Delta V$ of the primary–secondary pairs in the mock as functions of $\log L_c$ (lower-left) and $R_c$ (lower-right). In addition to an obvious increase in the dispersion of $\Delta V$ with increasing $L_c$, which is the signal of interest, a roughly uniform contribution from interlopers is apparent.

We analyse this mock data using the method outlined in Section 3. The first step is to determine the best-fitting radial profile of the satellite galaxies, $n_{\text{sat}}(r|M)$, properly marginalized over all other model parameters. Using a $10 \times 9$ grid in ($\gamma$, $\mathcal{R}$)-space, we use the downhill simplex method (Nelder & Mead 1965) to find the best-fitting CA-model for each ($\gamma$, $\mathcal{R}$)-model (results for the OM-model are very similar). Since this Tier-1 mock has no fibre collisions, or other form of incompleteness, we set the weights of all secondaries, $w_{ij}$, to unity, and the cut-off radius, $R_c$, to zero. The left-hand panel of Fig. 3 shows the 68, 95, and 99 percent confidence levels thus obtained. Confidence levels are computed assuming that $\chi^2_{\text{tot}}$ obeys a Chi-square distribution with two degrees of freedom, where $\chi^2_{\text{tot}} = -2 \ln L_{\text{tot}}$. The contours trace out a narrow region centred on the input model ($\gamma = \mathcal{R} = 1$), indicated by a solid, black dot. Hence, Basilisk yields an unbiased, and well-constrained estimate of the radial profile of the satellites, at least for this highly idealized mock.

The next step is to quantify the full posterior distribution of our model parameters using the best-fitting radial profile (i.e. $\mathcal{R} = \gamma = 1.0$). We do so constructing an MCMC of 5 million elements (5000 steps for 1000 walkers). The resulting medians and 95 percent confidence intervals for all parameters are listed in Table 2, while Fig. 4 shows a number of posterior predictions. In particular, the solid dots in panels (a)–(e) indicate the true values, while the shaded regions always mark the 95 percent confidence intervals inferred from our MCMC. Panel (a) shows the constraints on the luminosity–halo mass relation for central galaxies. For completeness, the grey dots show the actual mock data. Clearly, the model is successful in recovering the expectation value for the central luminosity given the mass of its halo. Panel (b) shows the cumulative distributions $P(<M(L_c))$ for three different central luminosities, as indicated. Again, the posterior predictions are in excellent agreement with the true distributions, indicating that Basilisk recovers not only the average relation between light and mass, but the full distribution. This is also apparent from Table 2, which shows that both $\sigma_{15}$ and $\sigma_{14}$ are tightly constrained, and in excellent agreement with the true values. An important reason for this success is the fact that we include the number of secondaries, $N_c$, as a constraint. As we demonstrate in Appendix B, ignoring this observable, as has been done in several previous studies, results in a strongly biased inference on the galaxy–halo connection.

Panels (c) and (d) of Fig. 4 plot the additional data used to constrain the model: the former plots $f_0 = \log [N_p/N_c]$ (see Section 3.4.1) and the latter plots the luminosity function of all galaxies (centrals plus satellites). Note that $f_0$ is plotted as a function of the luminosity of the primary and for three different redshift bins, as indicated. The same holds for the interloper fractions plotted in panel (e). The posterior predictions for $f_0(L_c, z_c)$, $\Phi(L)$, and $f_{\text{int}}(L_c, z_c)$ are all in excellent agreement with their true values. In particular, as is evident from Table 2, the posterior constraints on the parameters $\eta_0$, $\eta_1$, and $\eta_2$ that model the interlopers are in excellent agreement with their input values, further elucidating the success.

Figure 3. Constraints on the two gNFW parameters, $\mathcal{R}$ and $\gamma$, that characterize the radial distribution of satellite galaxies. Different panels show results for different tier mocks, as indicated, and contours, from dark to light, correspond to the 68, 95, and 99 per cent confidence levels obtained from $\Delta \chi^2_{\text{tot}}$ as described in the text. Thick, solid black dots indicate the true input values, while filled pentagons indicate the best-fitting model. In the case of the Tier-1 mock, the latter coincides with the former, and is therefore not visible.
Table 2. Confidence intervals for posteriors of model parameters inferred from Basilisk for the three tier mocks described in the text. Column (1) lists the model parameter, column (2) the input value used to construct the mock, and the remaining columns list the 2.5, 50, and 97.5 percentiles of the corresponding posterior distributions. Results for the Tier-1 mock are in columns (3)–(5), for the Tier-2 mock in columns (6)–(8), and for the Tier-3 mock in columns (9)–(11). Note that these percentiles all correspond to the MCMCs obtained using the CA-model. Results for the OM-model are extremely similar, and therefore not shown. However, the last row lists the 2.5, 50, and 97.5 percentiles for log \(r_a/r_s\) obtained from a separate MCMC that uses the OM-model. Note that the input values for the effective bias parameters of the interlopers, \(\eta_0\), \(\eta_1\), and \(\eta_2\), are only valid for the Tier-1 mock, while the input value for \(\beta\) only applies to the Tier-1 and Tier-2 mocks.

| Parameter Input | Tier-1 | Tier-2 | Tier-3 |
|-----------------|--------|--------|--------|
| \(\log M_1/(h^{-1} M_\odot)\) | 11.20 | 11.15 | 11.27 | 11.40 | 11.13 | 11.24 | 11.35 | 11.07 | 11.18 | 11.30 |
| \(\log L_0/(h^{-2} L_\odot)\) | 9.95 | 9.91 | 9.99 | 10.07 | 9.87 | 9.94 | 10.02 | 9.89 | 9.96 | 10.03 |
| \(\gamma_1\) | 3.50 | 3.02 | 3.45 | 3.87 | 3.08 | 3.49 | 3.91 | 3.04 | 3.45 | 3.87 |
| \(\gamma_2\) | 0.25 | 0.20 | 0.23 | 0.26 | 0.23 | 0.25 | 0.28 | 0.23 | 0.25 | 0.27 |
| \(\sigma_{12}\) | 0.15 | 0.13 | 0.15 | 0.18 | 0.14 | 0.17 | 0.19 | 0.12 | 0.15 | 0.17 |
| \(\sigma_{14}\) | 0.15 | 0.14 | 0.17 | 0.19 | 0.13 | 0.14 | 0.16 | 0.13 | 0.15 | 0.17 |
| \(\alpha_{12}\) | -1.20 | -1.61 | -1.28 | -0.94 | -1.21 | -0.74 | -0.32 | -1.01 | -0.69 | -0.37 |
| \(\alpha_{14}\) | -1.20 | -1.37 | -1.16 | -0.90 | -1.37 | -1.16 | -0.84 | -1.35 | -1.11 | -0.82 |
| \(b_0\) | -1.20 | -1.47 | -1.28 | -1.12 | -1.26 | -1.06 | -0.90 | -1.06 | -0.93 | -0.82 |
| \(b_1\) | 1.50 | 1.41 | 1.62 | 1.86 | 1.39 | 1.60 | 1.85 | 1.24 | 1.41 | 1.59 |
| \(b_2\) | -0.20 | -0.30 | -0.23 | -0.17 | -0.33 | -0.27 | -0.20 | -0.26 | -0.20 | -0.14 |
| \(\eta_0\) | 1.00 | 0.81 | 0.96 | 1.18 | 0.49 | 0.58 | 0.71 | 0.40 | 0.48 | 0.58 |
| \(\eta_1\) | 0.00 | -0.24 | -0.13 | 0.01 | -0.20 | -0.06 | 0.07 | -0.30 | -0.16 | 0.02 |
| \(\eta_2\) | 0.00 | -0.98 | 0.96 | 1.93 | -1.73 | 1.23 | 1.95 | -0.52 | 1.33 | 1.96 |
| \(\beta\) | 0.00 | -0.13 | 0.05 | 0.22 | -0.13 | 0.18 | 0.44 | 0.07 | 0.27 | 0.46 |
| \(\log [r_a/r_s]\) | - | 0.64 | 0.92 | 1.40 | 0.89 | 1.25 | 1.48 | 1.12 | 1.38 | 1.49 |

Figure 4. Results for the analysis of the Tier-1 Mock data. In panels (a)–(e) solid dots always reflect the true input values of the mock, while shaded regions mark the 95 per cent confidence interval inferred from the MCMC. Panel (a) plots the luminosity of central galaxies, \(L_c\), as a function of halo mass, \(M\). Grey dots indicate the actual primaries from the mock, and are shown for comparison and to highlight the amount of scatter in this relation. Panel (b) plots the cumulative distributions for halo mass, \(M\), for three different central luminosities, as indicated. Panel (c) plots \(\log [f_0]\), which is the statistic we use to describe the fraction of primaries with zero secondaries (see Section 3.4.1), as function of the luminosity of the primary, for three different redshift bins, the median values of which are indicated. The luminosity function of all galaxies (centrals plus satellites) is shown in panel (d), while panel (e) plots the interloper fractions as function of luminosity and redshift. Finally, panel (f) plots for each primary the predicted halo mass, \(M_{\text{pred}}\) (equation 54) versus the true halo mass, \(M_{\text{true}}\). Errorbars reflect the 95 per cent confidence intervals on \(M_{\text{pred}}\) as inferred from the MCMC, while the solid, red line indicates the running average. See text for a detailed discussion.
of Basilisk. Finally, panel (f) plots the predicted halo mass, \( M_{\text{pred}} \), versus the true halo mass, \( M_{\text{true}} \), for each primary. The former is computed using

\[
M_{\text{pred}} = \int_0^\infty P(M|L_c, z_c, N_i) M \, dM,
\]

with \( P(M|L_c, z_c, N_i) \) given by equation (7), and serves as a latent variable in our hierarchical Bayesian framework. Errorbars reflect the 95 per cent confidence intervals on \( M_{\text{pred}} \) as inferred from the MCMC, while the solid, red line indicates the running average. Although there is a small systematic bias, in that \( M_{\text{pred}} \) is too high (low) when \( M_{\text{true}} \) is small (large), the bias is small compared to the primary-to-primary variance; averaged over all 2373 primaries, we obtain \( \langle \log (M_{\text{pred}}/M_{\text{true}}) \rangle = 0.12 \), while the halo-to-halo scatter is 0.31. This demonstrates that our inferences regarding the probability distribution \( P(M|L_c, z_c, N_i) \), which we marginalize over in our evaluation of the likelihood for the satellite kinematics data (cf. equation 6), is not significantly biased.

5.2 Tier-2: simulation-based mocks

The idealized Tier-1 mock discussed above is based on a number of unrealistic oversimplifications. First of all, it is assumed that all primaries are centrals, and that each central is selected as a primary (i.e. it is effectively assumed that purity = completeness = 100 per cent). In reality, though, the selection criteria are imperfect and some selected primaries will be satellites, giving rise to impurities. In addition, not all centrals will pass our selection criteria giving rise to sample incompleteness. Realistic redshift surveys also suffer from fibre collisions, which results in an additional, spectroscopic incompleteness that is correlated with local (projected) density. Another shortfall of the Tier-1 mocks is that they assume that interlopers are distributed randomly and uniformly in space, which ignores clustering and redshift space distortions. Finally, when constructing the Tier-1 mocks, we assume the same zero-scatter concentration–mass relation for dark matter haloes as used in the modelling; in reality, haloes have a fair amount of (roughly lognormal) scatter in concentrations (e.g. Bullock et al. 2001; Macciò et al. 2007).

To allow for all these complications, we construct our Tier-2 mocks using a high-resolution N-body simulation from which we construct a mock redshift survey that is similar to the SDSS DR7. Our Tier-2 mocks are based on the SMDPL simulation (Klypin et al. 2016), which uses 38403 particles to trace structure formation in a cubic volume of \( 400 \, h^{-1} \text{Mpc}^3 \), adopting cosmological parameters that are compatible with the CMB constraints from Planck Collaboration XVI (2014). We use halotools (Hearin et al. 2017) to populate dark matter haloes at \( z = 0.0 \), identified with ROCKSTAR, according to our fiducial CLF model (see Table 1).

We populate each host halo with \( M_{\text{vir}} \geq 3 \times 10^{10} h^{-1} \text{M}_\odot \) with mock galaxies using the same method as outlined above for the Tier-1 mocks; i.e. we draw centrals and satellites from the CLF, all above a luminosity limit of \( 10^{8.5} L_\odot \), and assign them phase-space coordinates within the host halo. In particular, each central galaxy is given the position and velocity of its halo core, defined as the region that encloses the innermost 10 per cent of the halo virial mass. These positions and velocities are calculated by ROCKSTAR as detailed in Behroozi et al. (2013). For the satellites, we draw positions from a spherical distribution with radial profile \( n_{\text{sat}}(r) \), and one-dimensional velocities from a Gaussian distribution with dispersion \( \sigma(r|M) \), given by equation (50) with \( \beta = 0 \) (i.e. we assume isotropy). Both the positions and the velocities are with respect to the core of the host halo, and we use the measured concentration of each halo to determine individual halo profiles and satellite kinematics.

In the next step, we simulate the SDSS observations, following the procedure outlined in Lange et al. (2019a). First, we place a virtual observer with a random position and orientation into the simulation volume. We use this virtual observer to convert the \((x, y, z)\) coordinates of each galaxy into sky coordinates plus a cosmological redshift. If necessary, the simulation box is repeated periodically until the entire cosmological volume out to \( z = 0.17 \) is filled. Next, we only keep galaxies with \( m_r \leq 17.6 \) that lie within the SDSS DR7 survey mask. Redshift-space distortions are simulated by adding \((1 + z)v_{\text{los}}/c\) to each galaxy with cosmological redshift \( z \) and line-of-sight peculiar velocity \( v_{\text{los}} \). A random redshift error from a Gaussian with scatter \( \sigma_{\text{err}} = 15 \text{ km s}^{-1} \) is added in order to simulate spectroscopic redshift errors in the SDSS (Guo et al. 2015b). Finally, we simulate the effect of spectroscopic incompleteness. As discussed in Section 3.3, the SDSS suffers from fibre collisions whereby galaxies with a neighbour within 55 arcsec have a 65 per cent chance of not having a spectroscopic redshift. We first construct a decollided set of target galaxies (Blanton et al. 2003), defined as galaxies without neighbouring targets within 55 arcsec. We randomly assign 65 per cent of all galaxies that are not part of this decollided set a redshift, with the remaining 35 per cent making up our ‘fibre-collided’ set (galaxies that lack a redshift due to fibre collisions). Finally, we randomly remove an additional 1 per cent of all redshift to simulate other redshift failures. As demonstrated in Lange et al. (2019a) this approach captures all the salient features of spectroscopic incompleteness in the SDSS DR7. Once the mock is completed, we select primaries and secondaries as described in Section 2.1, and assign spectroscopic weights to all secondaries using the method described in Section 3.3.

We analyse the Tier-2 mocks in exactly the same fashion as the Tier-1 mocks described above. Note, though, that the spectroscopic weights, \( w_{\text{g}} \), used for the fibre collision correction are no longer unity, and that we use a non-zero cut-off radius, \( R_c(z_c) \), equal to 55 arcsec. Before showing the results from such an analysis, we first discuss some statistics of the mocks. A full Tier-2 mock, which mimicks the SDSS-DR7, contains \( \sim 37 \, 000 \) secondaries around \( \sim 19 \, 000 \) primaries (with at least one secondary). From such a mock, we construct eight subsamples of roughly equal size, whereby each primary plus its corresponding secondaries are randomly assigned to one of the subsamples. Since the CPU-cost for the MCMC analysis is proportional to the number of secondaries, analysing smaller subsamples is faster, and therefore ideal for testing. Furthermore, as is evident from Figs 4 and 7, samples with of order 5000 secondaries already yield extremely tight constraints. This implies that we can easily subdivide the actual SDSS-DR7 data in of order 10 subsamples, and still achieve exquisite constraints on the galaxy–dark matter connection for each. These subsamples can be random, in which case we can test for consistency among them, or based on secondary properties of the primaries (i.e. split by colour, size, bulge-to-disc ratio, etc.). The latter will allow for a much richer characterization of the galaxy–halo connection, and give valuable insight regarding galaxy assembly bias (Zentner et al. 2014; Hearin et al. 2016; Zentner et al. 2019). In what follows, we focus on one of the subsamples of our Tier-2 mock, consisting of 4567 secondaries around 2373 primaries. In addition, the subsample contains 23,234 primaries with zero secondaries. Note that this is almost twice as many as for the Tier-1 mock with 5000 secondaries. As we will see below, this is mainly because the (more realistic) number density of interlopers in the Tier-2 mock is roughly half of that in the Tier-1 mock.
The left-hand panel of Fig. 5 plots the completeness \(C(M|L, z)\) in our Tier-2 mock, defined as the fraction of centrals of luminosity \(L\) residing in haloes of mass \(M\) at redshift \(z\) that are selected as primaries by our cylindrical isolation criterion. The completeness is plotted as function of halo mass for seven different luminosity bins (different colours, as indicated) and two different redshift bins (dashed and solid lines correspond to \(z = [0.02, 0.09]\) and \(z = [0.09, 0.15]\), respectively). In each case we plot \(C(M|L, z)\) over the 5–95 percentile range of the corresponding halo mass. Note that the completeness of centrals is fairly high, increasing from \(\sim 65\) per cent at \(L_c = 10^{8.5} h^{-2} L_\odot\) to \(\sim 90\) per cent at \(L_c = 10^{11} h^{-2} L_\odot\). Most importantly, though, the completeness at given \(L\) and \(z\) has virtually no halo mass dependence. As discussed in Section 3.2.1 this implies that we can simply ignore \(C(M|L, z)\) altogether, or, equivalently, assume that it is equal to unity throughout.

The typical impurity in our Tier-2 mocks, defined as the fraction of primaries that are not centrals, plotted as function of the luminosity of the primary. Results are shown for our fiducial Tier-2 and Tier-3 mocks (solid blue and red curves, respectively), and for corresponding mocks in which we remove satellites that are brighter than their centrals, i.e. in which, by construction, each central is its brightest halo galaxy (dashed lines). Errorbars are computed assuming Poisson statistics. Note that the impurity fraction for the Tier-1 mock is zero by construction.

Finally, it is illustrative to compare to properties of interlopers in our Tier-1 and Tier-2 mocks. In the former, these were 'put in by hand' by drawing line-of-sight velocities and projected radii assuming that interlopers have a uniform phase-space distribution. In the latter, these arise from true projection effects in redshift space. The middle row of panels in Fig. 6 shows that the interlopers in the Tier-2 mocks have velocity distributions that differ notably from a purely uniform distribution. In particular, their \(\Delta V\)-distributions reveal a pronounced peak near \(\Delta V = 0\). This aspect of the phase-space distribution of interlopers has been pointed out in several previous studies (e.g. van den Bosch et al. 2004; Wojtak et al. 2007; Mamon, Biviano & Murante 2010). These have shown that these ‘peaks’ are mainly due to satellite galaxies between one and two halo virial radii that are \textit{bound} to the halo of the primary. Hence, their kinematics reflect the underlying gravitational potential of the host halo, and failing to identify them as ‘interlopers’ is not expected to cause a significant bias in the inferred halo masses. This was also pointed out in van den Bosch et al. (2004) and will be confirmed below. The right-hand panels of Fig. 6 shows that the interloper fraction as function of central luminosity in the Tier-2 mock is very similar to that in the Tier-1 mock.

The middle panel of Fig. 3 plots the constraints on \(\mathcal{R}\) and \(\gamma\), marginalized over all other model parameters and obtained using the same method as for the Tier-1 mock. As is evident, the parameters are now less well constrained, and reveal a slight, systematic bias. In fact, the best-fitting parameters are \(\gamma = 1.0\) and \(\mathcal{R} = 1.34\). Note, though, that the true input-model, which has \(\mathcal{R} = \gamma = 1.0\) (indicated by the solid, black dot), falls well within the 95 per cent confidence interval. The main reason for the reduced sensitivity is the fact that, as part of the mitigation of fibre-collisions, we exclude all secondaries separated from their primaries by less than 55 arcsec. Especially for fainter primaries at higher redshifts, this corresponds to an appreciable fraction of the halo’s virial radius, and it should not come as a surprise that this significantly diminishes the data’s potential to yield precise constraints on the phase-space distribution of satellite galaxies. However, this should not necessarily result in a

Figure 5. Left-hand panel: The completeness, \(C(M|L, z)\), in our Tier-2 mock for different bins of central luminosity (different colours, as indicated), and different redshifts bins; \(z = [0.02, 0.09]\) (dashed lines) and \(z = [0.09, 0.15]\) (solid lines). \(C(M|L, z)\) is defined as the fraction of centrals of luminosity \(L\) at redshift \(z\) in haloes of mass \(M\) that are selected as primaries by our cylindrical isolation criterion (cf. equation 12). Note that this completeness has only a very weak dependence on halo mass, which implies that it can be ignored in the modelling (see Section 3.2.1). Right-hand panel: Impurity fraction, \(f_{\text{imp}}\), defined as the fraction of primaries that are not centrals, plotted as function of the luminosity of the primary. Results are shown for our fiducial Tier-2 and Tier-3 mocks (solid red and blue curves, respectively), and for corresponding mocks in which we remove satellites that are brighter than their centrals, i.e. in which, by construction, each central is its brightest halo galaxy (dashed lines). Errorbars are computed assuming Poisson statistics. Note that the impurity fraction for the Tier-1 mock is zero by construction.
Figure 6. Interlopers in the Tier-1 (upper panels), Tier-2 (middle row of panels), and Tier-3 (lower panels) mocks. Panels on the left show the velocity distributions of secondaries around primaries with \( \log L_c \) in the range indicated at the top of each columns. The contribution due to interlopers is marked in red. Right-hand panels plot the interloper fraction as function of the luminosity of the primary.

Figure 7. Same as Fig. 4, but for the Tier-2 Mock. Note that we have excluded secondaries with \( R_p < R_{\text{cut}}(z_c) \) from the analysis in order to avoid problems due to fibre-collisions; see text for details.

significant, systematic bias, which instead arises from the combined effect of interlopers and impurity. Our model does not account for the excess of interlopers at small \( |\Delta V| \). Since these follow a radial profile that is more extended than that of the true satellites, this causes a systematic bias in \( R \). And the same applies for impurities: the radial distribution of secondaries around primaries that in reality are satellites is less centrally concentrated than that around true centrals. Fortunately, as we demonstrate below and in more detail in Section 6.2, the resulting, systematic overestimate of \( R \) does not have a significant impact on the inference regarding the galaxy–dark matter connection, which is our prime objective.

Fig. 7 shows the posterior results from the MCMC analysis of our fiducial Tier-2 subsample (see Table 2 for the medians and 95 percent confidence intervals for all model parameters). Results for other subsamples are extremely similar. Here we have adopted the CA-model with the best-fitting \( n_{\text{sat}}(r|M) \) (i.e. \( \gamma = 1.0 \) and \( R = 1.33 \)). As for the Tier-1 mock, we infer the posterior constraints from an MCMC consisting of 5 million elements. A comparison with Fig. 4 shows that the posterior constraints are
Figure 8. Marginalized posteriors obtained by Basilisk for the Tier-2 mock (assuming the best-fitting radial profile for the satellites, with $\gamma = 1$ and $R \approx 1.34$). Results are shown for the 11 parameters that characterize the CLF, and for the anisotropy parameter, $\beta$. To avoid having panels that are too small, we do not show the nuisance parameters $\eta_0, \eta_1,$ and $\eta_2$ that quantify the number density of interlopers. The diagonal shows marginalized 1D posteriors and off diagonal panels the 2D posteriors. In the latter case, contours demarcate the 68 per cent, 95 per cent, and 99 per cent containment of the posterior, while the red lines plus dot indicate the true input values used to create the mock data set. All parameters adopted a uniform prior, using the ranges indicated in Table 1, except for $\gamma_1$, which specifies the slope of the $L_c$–$M$ relation at the low-mass end, and which is poorly constrained. The Gaussian prior that was adopted instead is indicated with a thick, solid curve in the panel showing the 1D posterior for $\gamma_1$.

remarkably similar, despite the many additional complications that come with a Tier-2 mock and the fact that the value of $R$ adopted differs from that used to create the mock. In particular, the method yields constraints on $P(L_c|M)$ that are in good agreement with the input, both in terms of the median (panel a), and in terms of the full distributions (panel b). As for the Tier-1 mock, the posterior predictions for $f_0$ (panel c) and the luminosity function (panel d) are in excellent agreement with the data. However, the predictions for the interloper fractions (panel e) are slightly too low. This is due to the fact that the model assumes a uniform velocity distribution for interlopers, which does not account for the central ‘peak’ near $\Delta V = 0$ (Fig. 6). Most importantly, though, other than a systematic underestimate of the number density of interlopers, which is merely a nuisance parameter, our oversimplified treatment of interlopers does not result in a systematic error in the inferred galaxy–halo connection. Finally, as is evident from panel (f), the predicted halo masses, which act as latent variables in our hierarchical Bayesian framework, are again in good agreement, in a statistical sense, with the true masses with $\langle \log (M_{\text{pred}}/M_{\text{true}}) \rangle = 0.15$ and a halo-to-halo scatter of 0.35. Note that there are a small number of clear outliers, for which $M_{\text{pred}} \ll M_{\text{true}}$. These systems, which are absent in the Tier-1 mock, correspond to impurities in which the primary
is a satellite in the outskirts of a massive halo. At \(\sim 5\) per cent the impurity fraction is sufficiently low that this does not notably impact our overall inference.

To get some insight as to how the various parameters are correlated, Fig. 8 plots one- and two-dimensional marginalized posterior distributions, but only for the 11 parameters that characterize the CLF and the anisotropy parameter \(\beta\). In order to avoid having the panels be too small, we do not show the results for the three (nuisance) parameters that characterize the effective bias of the interlopers (\(\eta_0, \eta_1, \eta_2\)). The yellow histograms along the diagonal show the marginalized 1D posteriors, while the off-diagonal panels show marginalized 2D distributions, with the contours demarcating the 68 per cent, 95 per cent, and 99 per cent containment of the posterior. Red lines indicate the true input values used to create the mock data set. In general, the posteriors are in better agreement with the input values for the parameters that characterize the central part of the CLF (\(M_1, L_0, \gamma_1, \gamma_2, \sigma_{12}, \sigma_{14}\)), than for those that characterize the satellite component (\(\alpha_{12}, \alpha_{14}, b_0, b_1, b_2\)). This is true for all subsamples we examined, and also holds for the Tier-3 mock discussed below: whereas \(\text{Basilisk}\) always yields constraints on \(\Phi_1(L|M)\) in excellent agreement with the input, often it will yield constraints on one or more parameters characterizing \(\Phi_1(L|M)\) that are inconsistent with the input values. This owes to the impurities and the oversimplified treatment of interlopers (i.e., uniform velocity distribution) that is also responsible for the small but systematic error in \(n_{\text{gal}}(r|M)\). We suspect that the inference regarding the parameters of the satellite component of the CLF may significantly improve if the observed luminosities of the secondaries are included as data in our likelihood evaluation. Since the primary goal of analysing satellite kinematics is to constrain the halo masses of central galaxies, we leave such an extension for future work.

Note how all 1D posterior distributions depicted in Fig. 8 closely resemble Gaussians, and how most parameters are only weakly covariant with one another. Some notable exceptions are the sets \(\{\log M_1, \log L_0, \gamma_1, \gamma_2\}\) and \(\{b_0, b_1, b_2\}\), which reveal non-negligible covariance among each other (see also Cacciato et al. 2013). These mainly owe to the limiting dynamic range in halo mass covered by the (mock) data, especially the lack of data sampling the low-mass end (\(M \lesssim 10^{13} \, h^{-1} M_\odot\)). Mainly for this reason we decided to adopt a restricting, Gaussian prior for \(\gamma_1\), which is poorly constrained by the data. Indeed, as is evident from Fig. 8, the posterior distribution of \(\gamma_1\) is in perfect agreement with its prior, indicated by the solid, black curve.

To summarize, we conclude that neither impurities, nor the complicated phase-space distribution of interlopers, nor fibre-collisions, nor realistic scatter in the concentration–mass relation (which is ignored in our modelling) prohibit \(\text{Basilisk}\) from inferring accurate and precise constraints regarding the galaxy–halo connection of central galaxies. Although the aforementioned complications can cause a slight systematic bias in the inferred radial profile of satellite galaxies, and in the parameter \(\alpha_{12}\), these do not significantly impact the inference regarding \(\Phi_1(L|M)\). As a cautionary note, though, we find that not including a treatment for fibre-collisions, i.e. not applying the correction factor, \(f_{\text{corr}}\), given by equation (28), and not removing secondaries with \(R_p < R_s(z)\), typically results in significant biases that are very similar to the biases one incurs when not accounting for the number of secondaries (see Appendix B).

5.3 Tier-3: using subhaloes to model the phase-space distribution of satellites

At the third and final stage of our validation, we move to mock data sets in which we no longer make assumptions regarding the phase-space distributions of satellite galaxies. We construct Tier-3 mocks exactly as in the case of Tier-2 mocks, but rather than assuming a radial profile, \(n_{\text{sat}}(r|M, z)\), or assuming that satellite galaxies obey the spherical Jeans equations, we now assume that satellite galaxies reside in subhaloes, and their phase-space distribution therefore follows that of resolved subhaloes in the SMDPL \(N\)-body simulation.

After drawing a satellite number \(N_{\text{sat}}\) for each halo, we assign those \(N_{\text{sat}}\) satellites the phase-space coordinates of the \(N_{\text{sat}}\) subhaloes with the highest \(M_{\text{peak}}\). It is possible that \(N_{\text{sat}}\) exceeds the number of resolved subhaloes in a specific halo. In that case, we randomly take phase-space positions of subhaloes hosted by other haloes of a similar mass. We then proceed to generate a mock SDSS-like catalogue and analyse it in the same way as the Tier-2 mock described in the previous section. In particular, we select a random subsample containing 4521 secondaries around 2431 primaries, while an additional 23 360 primaries have zero secondaries. We have verified that the results for other random subsamples of similar size are again extremely similar.

As is evident from Figs 5 and 6 the impurity and interloper contamination in the Tier-3 mock are very similar to those in the Tier-2 mock. The overall impurity is 5.2 per cent, while the interloper fraction decreases from \(\sim 80\) per cent for primaries with \(L_\text{c} = 10^{10.5} h^{-2} L_\odot\) to \(\sim 10\) per cent for the brightest primaries.

The right-hand panel of Fig. 3 shows that the best-fitting radial profile for the satellites is very different from that of the Tier-1 and Tier-2 mocks. The data clearly prefers a large, constant density core with \(\gamma = 0\) and \(\mathcal{R} = 2.37\) (close to the limits on our adopted prior ranges). For comparison, the true, radial profile of satellite galaxies in the Tier-3 mock is best-fitting with a gNFW profile with \(\gamma = 0\) and \(\mathcal{R} = 2.57\) (indicated with a solid, black dot). This is not only in good agreement with the best-fitting value inferred from \(\text{Basilisk}\), but also with the well-known fact that the radial distribution of dark matter subhaloes in dark-matter-only simulations, such as the SMDPL simulation used here, is strongly antibiased with respect to the dark matter particles (e.g. Springel et al. 2008).\(^{12}\)

We analyse the Tier-3 mock, again using a MCMC with 5 million elements, keeping \(\gamma\) and \(\mathcal{R}\) fixed at these best-fitting parameters. The results are shown in Fig. 9, while Table 2 lists the medians and 95 per cent confidence intervals for all model parameters. The posterior constraints for the galaxy–halo connection of central galaxies, as characterized by \(\Phi_1(L|M)\), are once again in excellent agreement with the input model. In particular, the posteriors accurately reflect the full PDF of the relation between central luminosity and halo mass. Similar to what was found for the Tier-2 mock, the inferred fraction of interlopers is somewhat too low, while the predicted halo masses are in good agreement with the true masses, with \((\log (M_{\text{pred}}/M_{\text{true}})) = 0.10 \pm 0.37\). Results for other subsamples are very similar. Hence, we conclude that \(\text{Basilisk}\) is extremely reliable and robust in constraining the galaxy–halo connection of centrals, despite the oversimplified assumptions that satellite galaxies reside in spherical NFW haloes, obey the Jeans equation, and have Gaussian LOSVDs.

\(^{11}\)Neither of these parameters reveals significant covariance with any of the parameters shown in Fig. 8.

\(^{12}\)This is, at least partially, an artefact due to artificial numerical disruption (van den Bosch & Oguya 2018; van den Bosch et al. 2018).
6 CONSTRAINING THE PHASE-SPACE DISTRIBUTION OF SATELLITE GALAXIES

In this section we discuss the accuracy and precision with which Basilisk can constrain the orbital anisotropy of satellite galaxies, and we investigate how central velocity bias and errors in the radial profile of satellite galaxies impact Basilisk’s inference regarding the galaxy–halo connection.

6.1 The orbital anisotropy of satellite galaxies

Fig. 10 shows the posterior distributions inferred from the mocks from all three of the tiers. Left- and right-hand panels show the results obtained from MCMCs in which we use the constant anisotropy (CA) model and the OM model, respectively. The upper panels correspond to Tier-1 and reveal posteriors in good agreement with the isotropy assumed, i.e. $\beta = 0$, which corresponds to a large $r_a/r_s$ in the case of an OM-model. As indicated in Table 2, Basilisk yields $\beta = 0.05^{+0.17}_{-0.14}$ and $\log[r_a/r_s] = 0.92^{+0.48}_{-0.28}$. The uncertainties indicate the 95 per cent confidence intervals centred on the medians. Note that the posterior distribution for $\log[r_a/r_s]$ is restricted by our assumed prior, which is uniform over the range $[-1.0, 1.5]$. The middle row of panels show the results for the Tier-2 mock, for two values of $R_\text{tot}$; dark-blue histograms correspond to the best-fitting value of $R_\text{tot}$, which corresponds to a $\Delta \chi^2 = \chi^2_{\text{tot,OM}} - \chi^2_{\text{tot,CA}} = 8.6$. As we demonstrate in Appendix C, the orbital anisotropy of dark matter subhaloes in the SMDPL simulation has a strong mass dependence (see also Cuesta et al. 2008). Whereas the average anisotropy is well described by $\beta \sim 0.2$ in excellent agreement with the constraints shown in the lower-left panel of Fig. 10, the anisotropy parameter has a strong mass and radius dependence. In massive hosts, with $M_{\text{vir}} \gtrsim 10^{13} h^{-1} M_\odot$, the anisotropy parameter $\beta$ increases with increasing radius, in good agreement with the findings by Diemand et al. (2004). However, in less massive haloes the anisotropy parameter is found to decrease with increasing radius (see also Sawala et al. 2017, who studied the orbital anisotropy of Milky Way sized host haloes). Since a negative radial gradient in $\beta$ is inconsistent with an Osipkov–Merrit model, for which $d\beta/dr > 0$, this explains why the best-fitting OM-model for the Tier-3 mock is significantly worse than the CA-model.

6.2 The radial profiles of satellite galaxies

As shown in Section 5.2, the combination of fibre collision incompleteness, sample impurities and interlopers can cause a systematic error in the inferred radial profile of the satellite galaxies, as characterized by $R$ and $\gamma$. However, in the case of the Tier-2 mock, we also found that the MCMC that adopts these biased, best-fitting
Bayesian inference with satellite kinematics

The best-fitting model (lower-right panel) depends very strongly on \((R, \gamma)\), which is why we were able to obtain tight, albeit slightly biased, constraints on these parameters (cf. Fig. 3).

It may seem surprising that errors in \(R\) and/or \(\gamma\) have so little impact; after all, the radial profile of the satellite galaxies enters the Jeans equation that is used to compute the radial velocity dispersion, \(\sigma_r(\vec{r}|M, z)\) (equations 50 and 52). The main reason is that the line-of-sight velocity dispersion, which is used to compute the probability \(P(\Delta V)|\vec{r}, M, z\), is less sensitive to changes in \(n_{\text{rad}}(\vec{r}|M, z)\) than the radial velocity dispersion. To first order, these changes are degenerate with (relatively modest) changes in orbital anisotropy and interloper contribution. Interestingly, though, there seems to be virtually no degeneracy with the parameters that characterize \(\Phi_c(L|M)\). Such degeneracy is inhibited by the additional constraints that we use in our inference (i.e. the galaxy luminosity function and the fraction of primaries with zero secondaries). Hence, we conclude that although fibre collisions, impurity, and an oversimplified treatment of interlopers can cause a slight, systematic error in \(\text{Basilisk}'s\) inference of the radial distribution of satellite galaxies, this has no discernible impact on the inference regarding the halo occupation statistics of central galaxies. And although it may impact the constraints on the velocity anisotropy, unless the systematic errors on \(\gamma\) and/or \(R\) are very large, which is unlikely to be the case, the magnitude of this effect is small compared to the typical statistical uncertainty.

6.3 Central velocity bias

Throughout we have assumed that central galaxies reside at rest at the centre of the host halo. However, as pointed out in a number of studies (van den Bosch et al. 2005b; Reid et al. 2014; Guo et al. 2015a,b, 2016; Ye et al. 2017), this is not necessarily the case. Several studies have investigated this so-called central velocity bias in hydrodynamical simulations of galaxy formation (e.g. Berlind et al. 2003; Yoshikawa, Jing & Börner 2003). More recently, Ye et al. (2017) used hydrodynamical simulations from the Illustris-1 suite (Vogelsberger et al. 2014) to examine the motion of central galaxies with respect to the bulk velocity of their host haloes. They defined the central velocity bias

\[
\alpha_c \equiv \frac{\sqrt{|v_c - v_h|^2}}{\sigma_{\text{halo}}^2},
\]

where \(v_c\) is the velocity of the central galaxy, \(v_h\) is the centre-of-mass velocity of the host halo, \(\sigma_{\text{halo}}\) is the 3D velocity dispersion of the dark matter particles of the host halo, and the angle brackets indicate an average over an ensemble of central galaxies. If central galaxies are at rest with respect to their host haloes, then \(\alpha_c = 0\) (‘no bias’). Ye et al. (2017) find that centrals in the Illustris-1 simulation have an average central velocity bias \(\alpha_c \sim 0.2\). In fact, \(\alpha_c\) depends strongly on halo age and the stellar mass-to-host halo mass ratio, \(M_{\text{stellar}}/M_{\text{halo}}\), with younger haloes and smaller \(M_{\text{stellar}}/M_{\text{halo}}\) corresponding to larger velocity bias, reaching values for \(\alpha_c\) as large as 0.6.

If we assume that the average velocity (but not necessarily the rms velocity) of subhaloes coincides with that of the bulk of the host halo, then a non-zero central velocity bias will boost the root-mean-square velocity difference \(\sqrt{\Delta V^2}\) by roughly a factor \(\sqrt{1.0 + \alpha_c^2}\), which can systematically bias the inference of \(\text{Basilisk}\). For \(\alpha_c = 0.2\) the boost is only \(\sim 1.02\), but since it is systematic and since, to good approximation, \(M \propto (\Delta V)^{3/2}\), the impact might be non-negligible. In order to test this, we proceed as follows. For each central galaxy, \(i\), in a mock sample we

![Figure 10. Posterior distributions (normalized) for the anisotropy parameters inferred from the three tier mocks (different rows) for both the CA- and the OT-model (right). The Tier-1 and Tier-2 mocks were constructed with isotropic orbital distributions, corresponding to \(\beta = 0\) (indicated by the vertical, dashed line in the left-hand panels). For the Tier-2 mock, we show the posterior distributions for both our fiducial analysis, which adopts the best-fitting value for \(R\) (light-blue colour), and for an analysis that adopts the true input value of \(R = 1.0\) (dark-blue colour). The grey, vertical bar in the lower-left panel indicates the typical range of anisotropy parameters for subhaloes in the SMDPL simulation (see Appendix C). In the Tier-3 mock, satellite galaxies are placed on subhaloes, and the grey bar therefore indicates the true, underlying anisotropy of the satellite galaxies in this mock.](https://academic.oup.com/mnras/article/488/4/4984/5539546)
Figure 11. The impact of changes in $R$ and $\gamma$, characterizing $n_{\text{sat}}(r|M)$, on the inference of Basilisk. Each panel plots the best-fitting values of a different CLF parameter as a function of $R$, with different colours corresponding to different values of $\gamma$ (as indicated in the upper-left panel). The shaded regions indicate the 95 per cent confidence intervals obtained in the case of our fiducial MCMC that adopts the best-fitting parameters for $R$ and $\gamma$ (filled pentagon in the middle panel of Fig. 3), while the horizontal, dotted lines mark the input values used to construct the mock data. The lower-right panel plots $\Delta \chi^2 \equiv \chi^2_{\text{tot}}(\alpha_c) - \chi^2_{\text{tot}}(\alpha_c = 0)$, with $\chi^2_{\text{tot}}$ the $\chi^2$-value of the best-fitting model. For $\alpha_c \lesssim 0.4$, central velocity bias has a negligible impact on the inference, with best-fitting parameters that are almost indistinguishable from the case with no velocity bias. Even for $\alpha_c = 0.6$, corresponding to the largest values found in the Illustris-1 simulation suite, the impact of the central velocity bias is extremely modest, in that the differences in the best-fitting parameters remain small compared to the 95 per cent confidence intervals. The main impact of a non-zero velocity bias is a decrease in the goodness of fit, which is evident from the rapid increase of $\Delta \chi^2$ with increasing $\alpha_c$. Together with the fact that $\alpha_c$ does not display a significant degeneracy with any of the other model parameters, this therefore suggest that it should even be possible to include $\alpha_c$ as a free parameter in our model, without significantly impacting Basilisk’s ability to constrain the galaxy–halo connection.
7 DISCUSSION AND CONCLUSIONS

As part of our ongoing efforts to mature satellite kinematics into an accurate, competitive probe of the galaxy–halo connection, complementary to galaxy clustering and galaxy–galaxy lensing, this paper presents a new Bayesian hierarchical method for analysing the kinematics of satellite galaxies. Basilisk uses the spherically symmetric Jeans equations to model the kinematics of large ensembles of satellite galaxies associated with central galaxies that span a wide range in halo mass and luminosity (or stellar mass). The halo masses of the individual centrals act as latent variables in a hierarchical Bayesian framework that uses data from galaxy redshift surveys to constrain the detailed galaxy–halo connection, characterized by the CLF.

Unlike traditional methods for analysing satellite kinematics, Basilisk does not resort to stacking the kinematics of satellite galaxies in bins of central luminosity, and does not make use of any summary statistic, such as satellite velocity dispersion. Rather, Basilisk leaves the data in its raw form, which has the advantage that all data is used optimally. In addition, whereas traditional methods typically require volume-limited samples, Basilisk can be applied to flux limited samples, thereby greatly enhancing the quantity and dynamic range of the data. And finally, Basilisk is the only available method that simultaneously solves for halo mass and orbital anisotropy of the satellite galaxies, while properly accounting for ‘mass-mixing’.

Starting from a sample of primary and secondary galaxies selected from a galaxy redshift survey, representing centrals and satellites, respectively, Basilisk uses the projected phase-space coordinates of the secondaries with respect to their primaries to constrain the galaxy–halo connection under the following assumptions:

(i) Dark matter haloes are characterized by spherical NFW density profiles that follow the concentration–mass relation of Macciò et al. (2008) with zero scatter.

(ii) Satellite galaxies are a virialized, steady-state tracer population of their underlying host halo potential.

(iii) The LOSVD, $P(\Delta V|R_p, M, z)$, is a Gaussian with zero mean and a dispersion, $\sigma_{\text{los}}$, that follows from the spherical Jeans equations.

(iv) Interlopers have a uniform distribution in redshift space.

Each of these assumptions is expected to be violated at some level. Dark matter haloes are triaxial, rather than spherical (e.g. Jing & Suto 2002; Allgood et al. 2006), and their concentration–mass relation has an appreciable amount of (lognormal) scatter (e.g. Bullock et al. 2001; Macciò et al. 2007). In addition, due to the continued disruption and merging of existing satellites, and the accretion of new ones, the system of satellite galaxies is not expected to be perfectly virialized or to be in a steady-state, and their kinematics therefore do not necessarily obey the Jeans equations (e.g. Wang et al. 2017, 2018a; Ye et al. 2017; Adhikari et al. 2019). Furthermore, there is no reason why the local LOSVD, $P(\Delta V|R_p, M, z)$, be perfectly Gaussian. Finally, several studies have shown that interlopers do not have a uniform distribution in redshift space (e.g. van den Bosch et al. 2004; Wojtak et al. 2007). Clearly, then, there are numerous reasons why one might expect Basilisk to fail, and its performance therefore needs to be tested and validated in detail.
We have done so using a three-tiered validation process, in which we test the performance of Basilisk on a series of mock data sets of increasing complexity and realism. The Tier-1 mocks are highly idealized, abiding by all four assumptions listed above, and are mainly used to test the main engine of Basilisk and to gauge the constraining power given a particular amount of data. For the Tier-2 mocks we place mock galaxies inside dark matter haloes in a large, cosmological N-body simulation. The satellite galaxies are given phase-space coordinates within their host haloes that, by construction, still obey assumptions (ii) and (iii), but the masses, concentrations, positions, and velocities of the dark matter haloes are taken directly from the simulation. A mock redshift survey is constructed by placing a virtual observer at a random location within the simulation volume, after which primaries and secondaries are selected using the cylindrical isolation criteria described in Section 2.1. Consequently, the Tier-2 mocks have realistic distributions of interlopers, and suffer from impurities and incompleteness effects (in particular fibre-collisions) in the same way as data extracted from a realistic galaxy redshift survey. Finally, the Tier-3 mocks are similar to the Tier-2 mocks, except that now the phase-space distributions of the satellite galaxies are assumed to be identical to those of the subhaloes (i.e. satellite galaxies are placed on subhaloes within the N-body simulation). Hence, in the Tier-3 mocks the satellite galaxies only obey assumptions (ii) and (III) in as far as subhaloes do.

As shown in Section 5, Basilisk is able to yield precise and accurate constraints on the galaxy–halo connection in the case of all three tier mocks. The combined effect of impurities and fibre-collisions causes a small, systematic bias in the inferred radial profile of satellite galaxies, but we have demonstrated that this does not significantly impact the inference regarding the galaxy–halo connection. As demonstrated in Appendix B, it is crucial, though, to include the number of secondaries per primary as a constraint on the data; failing to do so means that the model is unable to constrain the scatter in the galaxy–halo connection, and results in posteriors that are systematically biased. In line with our previous studies (Lange et al. 2019a,b), we therefore conclude that the kinematics of satellite kinematics are a powerful probe of the galaxy–halo connection, complementary to and competitive with galaxy clustering and galaxy–galaxy lensing.

In fact, satellite kinematics has several advantages with respect to these alternative methods. Unlike galaxy–galaxy lensing, which requires tangential shear measurements that rely on accurate photometry, accurate characterization of the point-spread function (‘seeing’), and a non-trivial method to extract reliable shape measurements, satellite kinematics can be measured from any redshift survey without the need for any additional data. And unlike galaxy clustering, which only probes halo mass indirectly through the mass dependence of the halo bias, satellite kinematics, similar to galaxy–galaxy lensing, directly probes the gravitational potential of the dark matter haloes, thus giving a more direct handle on the galaxy–halo connection.

Although the discussion presented here has focused exclusively on constraining the CLF, Basilisk is easily modified so that it can constrain the conditional stellar mass function. This is most easily done by first defining a redshift-dependent, stellar-mass complete sample as in van den Bosch et al. (2008b), and subsequently replacing the minimum luminosity, \( L_{\text{min}}(z) \), used in the computation of the expectation value for the number of satellites (equation 15) with the corresponding minimum stellar mass, \( M_{\text{star,min}}(z) \). In addition, one has to account for the fact that stellar masses are not directly measurable, but instead are inferred from (multiwavelength or spectroscopic) data in a model-dependent fashion. This implies that the data is effectively ‘convolved’ with an unknown probability function, \( P(M_{\text{obs}} | M_{\text{true}}) \), relating the true and ‘observed’ stellar masses, \( M_{\text{true}} \) and \( M_{\text{obs}} \), respectively. The Bayesian hierarchical framework that underlies Basilisk is ideally suited to account for such a complication (see Sonnenfeld & Leauthaud 2018).

We end this paper by discussing potential future advances and applications. There are several opportunities for further development of the Basilisk framework. First of all, Basilisk in its current form assumes that the halo occupation statistics depend only on halo mass. If, instead, the occupation statistics also depend on other halo properties that impact satellite kinematics, such as halo concentration, the inference of Basilisk may be significantly impacted. A particular concern is that haloes of a given mass that are more concentrated have fewer subhaloes on average (e.g. van den Bosch, Tormen & Giocoli 2005a; Zentner et al. 2005a; Giocoli et al. 2010; Jiang & van den Bosch 2017; Fielder et al. 2019).13 This implies a correlation between the number of satellite galaxies and host halo concentration, at fixed host halo mass. Since the kinematics of satellites depend on the concentration of the host halo, this correlation, which is currently not accounted for, could potentially impact Basilisk’s inference. We intend to examine this issue in the near-future, and upgrade Basilisk accordingly, if needed. Another possible extension of Basilisk is to include satellite luminosities as constraints on the model. In principle this should tighten the constraints on the satellite component of the CLF, and may well be important for improving the constraints on \( \alpha_{12} \) and \( \alpha_{14} \). On the other hand, this also causes complications, as it probably requires a careful treatment of luminosity-segregation, i.e. the fact that satellites of different luminosities have different radial profiles (e.g. Rood & Turnrose 1968; Biviano et al. 2002; van den Bosch et al. 2008a; Balogh et al. 2014). Since this introduces additional degrees of freedom, i.e. \( n_{\text{sat}}(r, M) \rightarrow n_{\text{sat}}(r, M, L_s) \), it remains to be seen to what extent including the satellite luminosities actually aids in constraining the galaxy–halo connection. And finally, based on the strong mass and radius dependence of the orbital anisotropy of dark matter subhaloes (cf. Fig. C1), it may be worthwhile to consider more sophisticated anisotropy models for the satellite galaxies, going beyond the constant anisotropy (CA) and Osipkov–Merritt (OM) models considered here.

As for applications, we will use Basilisk to analyse existing and forthcoming galaxy redshift surveys, including the SDSS Main Galaxy Sample (York et al. 2000), the Baryon Oscillation Spectroscopic Survey (BOSS, Dawson et al. 2013), the Galaxy and Mass Assembly Survey (GAMA, Driver et al. 2011), and the Dark Energy Spectroscopic Instrument (DESI). In particular, since Basilisk can yield accurate constraints for relatively small subsamples, compared to the full extent of these surveys, such an analysis will yield a much richer, multidimensional characterization of the galaxy–halo connection. We will also compare the inferred relation between galaxies and their dark matter haloes to constraints inferred from a combined analysis of galaxy clustering and galaxy–galaxy lensing. The latter consistently reveals tension with the cosmological parameters inferred from the Planck cosmic microwave background data (see e.g. Cacciato et al. 2013; Mandelbaum et al. 2013; Leauthaud et al. 2017; Abbott et al. 2018), and it will

13These studies reveal a strong anticorrelation between halo formation redshift and subhalo occupation; the dependence on halo concentration follows from its strong correlation with formation time (Wechsler et al. 2002).
be interesting to see whether such tension persists when satellite kinematic data are included in the analysis. Another interesting opportunity is to compare the halo masses inferred for samples of primaries from satellite kinematics, using Bayesian approach, and galaxy–galaxy lensing. This can test the law of gravity on the scale of galaxy–group-sized haloes by constraining the gravitational slip (see e.g. Daniel et al. 2008; Pizzuti et al. 2019). In short, we envision a bright future for the hitherto underutilized method of satellite kinematics as a probe of the galaxy–halo connection.

ACKNOWLEDGEMENTS

We are grateful to the referee for constructive comments that helped to improve the presentation. FvdB and JUL are supported by the US National Science Foundation through grant AST 1516962. ARZ is funded by the Pittsburgh Particle physics Astrophysics and Cosmology Center (Pitt PACC) at the University of Pittsburgh and by the NSF through grant NSF AST 1517563. This research was supported in part by the National Science Foundation under Grant No. NSF PHY-1125915 and NSF PHY-1748958. FvdB received additional support from the Klaus Tschira foundation, and from the National Aeronautics and Space Administration through Grant No. 17-ATP17-0028 issued as part of the Astrophysics Theory Program.

REFERENCES

Abbott T. M. C. et al., 2018, Phys. Rev. D, 98, 043526
Adhikari S., Dalal N., More S., Wenzel A., 2019, ApJ, 878, 9
Allgood B., Flores R. A., Primack J. R., Kravtsov A. V., Wechsler R. H., Faltenbacher A., Bullock J. S., 2006, MNRAS, 367, 1781
Ascasibar Y., Gottlober S., 2009, MNRAS, 386, 2022
Azaro M., Patiri S. G., Prada F., Zentner A. R., 2007, MNRAS, 376, L43
Bahcall J. N., Primack J. R., 2002, ApJ, 577, 244
Balogh M. L. et al., 2014, MNRAS, 443, 2679
Becker M. R. et al., 2007, ApJ, 669, 905
Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, ApJ, 762, 109
Berlind A. A. et al., 2003, ApJ, 593, 1
Berlind A. A. et al., 2006, ApJS, 167, 1
Binney J., Tremaine S., 2008, Galactic Dynamics, 2nd edn. Princeton Univ. Press, Princeton
Biviano A., Katgert P., Thomas T., Adami C., 2002, A&A, 387, 8
Blanton M. R., Lin H., Lupton R. H., Maley F. M., Young N., Zehavi I., Loveday J., 2003, AJ, 125, 2276
Blanton M. R. et al., 2005, AJ, 129, 2562
Brainerd T. G., Blandford R. D., Smail I., 1996, ApJ, 466, 623
Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Primack J. R., Dekel A., 2001, MNRAS, 321, 559
Cacciato M., van den Bosch F. C., More S., Mo H. J., Yang X., 2009, MNRAS, 394, 929
Cacciato M., van den Bosch F. C., More S., Mo H., Yang X., 2013, MNRAS, 430, 767
Calderon V. F., Berlind A. A., 2019, preprint (arXiv:1902.08680)
Campbell D., van den Bosch F. C., Hearin A., Padmanabhan N., Berlind A. A., Primack J. R., Dekel A., 2001, MNRAS, 321, 559
Leauthaud A., et al., 2012, ApJ, 744, 159
Leauthaud A. et al., 2017, MNRAS, 467, 3024
Lin Y.-T., Mohr J. J., Stanford S. A., 2004, ApJ, 610, 745
Lokas E. L., Mamon G. A., 2001, MNRAS, 321, 155
Lokas E. L., Wojtak R., Gottlober S., Mamon G. A., Prada F., 2006, MNRAS, 367, 1463
Louis P. D., 1993, MNRAS, 263, 283
Macciò A. V., Dutton A. A., van den Bosch F. C., Moore B., Potter D., Stadel J., 2007, MNRAS, 378, 55
Macciò A. V., Dutton A. A., van den Bosch F. C., 2008, MNRAS, 391, 1940
Mamon G. A., Biviano A., Murante G., 2010, A&A, 520, A30
Mandelbaum R., Seljak U., Kauffmann G., Hirata C. M., Brinkmann J., 2006, MNRAS, 368, 715
Mandelbaum R., Slosar A., Baldauf T., Seljak U., Hirata C. M., Nakajima R., Reyes R., Smith R. E., 2013, MNRAS, 432, 1544
Mandelbaum R., Wang W., Zuo Y., White S., Henriques B., More S., 2016, MNRAS, 457, 3200
McKay T. A. et al., 2002, ApJ, 571, L85
Merritt D., 1985, AJ, 90, 1027
Mo H. J., White S. D. M., 1996, MNRAS, 282, 347
More S., van den Bosch F. C., Cacciato M., Mo H. J., Yang X., Li R., 2009a, MNRAS, 392, 801
More S., van den Bosch F. C., Cacciato M., 2009b, MNRAS, 392, 917
More S., van den Bosch F. C., Cacciato M., Skibba R., Mo H. J., Yang X., 2011, MNRAS, 410, 210
More S., Miyatake H., Mandelbaum R., Takada M., Spergel D. N., Brownstein J. R., Schneider D. P., 2015, ApJ, 806, 2
Nelder J. A., Mead R., 1965, Computer J., 7, 308
Norberg P., Frenk C. S., Cole S., 2008, MNRAS, 383, 646
Nurmi P. et al., 2013, MNRAS, 436, 380
Osipkov M. P., 1979, Pisma v Astronomicheskii Zhurnal, 5, 77
Pillepich A. et al., 2018, MNRAS, 475, 648
APPENDIX A: COMPUTING LIKELIHOODS

In the main text we have given the general expressions that are relevant for computing the likelihood $L_{\text{SK}}(D|\theta)$, where $\theta$ is the model vector, and $D$ is the data vector given by

$$D = \sum_{i=1}^{N_c} \left( [\Delta V_{ij}, R_{\theta,ij}] | j = 1, ..., N_z, i = 1, ..., N_s, L_{c,i}, z_{c,i}, N_{s,i} \right). \quad (A1)$$

Here we give the corresponding expressions for the specific model adopted throughout. As shown in Section 3 the log-likelihood can be written in compact form as $\ln L_{\text{SK}}(D|\theta) = \sum_{i=1}^{N_c} \left( \ln G_i - \ln F_i \right)$ (equation 8). Here $G_i$ and $F_i$ are integrals over halo mass, which we compute by integrating $\ln M$ from $M = 10^{10}$ to $10^{15} h^{-1} M_{\odot}$ using Gaussian quadrature. This allows us to write

$$F_i = \sum_k w_k \exp(\tilde{F}_{ik}), \quad G_i = \sum_k w_k \exp(\tilde{F}_{ik} + Q_{ik}), \quad (A2)$$

with $\tilde{F}_{ik} = \ln F_{ik}$ (cf. equations 9–10). Using the expressions for $P(L_{c,i} | N_{s,i}, z_{c,i})$ and $P(N_{s,i} | M_i, L_{c,i}, z_{c,i})$, we have that

$$\tilde{F}_{ik} = \ln M_k + \ln \left[ n(M_k, z_i) \right] - \ln \Gamma(N_{s,i} + 1) - \ln \sigma_c(M_k) - \left( \frac{\ln L_{c,i} - \ln L_c(M_k)}{\sqrt{2} \sigma_c(M_k)} \right)^2 + N_{s,i} \ln \lambda_{ik} - \lambda_{ik}. \quad (A3)$$

Wang E. et al., 2018b, ApJ, 860, 102
Watson D. F., Berlind A. A., McBride C. K., Masjedi M., 2010, ApJ, 709, 115
Watson D. F., Berlind A. A., McBride C. K., Hogg D. W., Jiang T., 2012, ApJ, 749, 83
Wechsler R. H., Tinker J. L., 2018, ARA&A, 56, 435
Wechsler R. H., Bullock J. S., Primack J. R., Kravtsov A. V., Dekel A., 2002, ApJ, 568, 52
Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, ApJ, 652, 71
Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006, MNRAS, 366, 2
Wetzel A. R., Tinker J. L., Conroy C., van den Bosch F. C., 2013, MNRAS, 432, 336
Wibking B. D. et al., 2019, MNRAS, 484, 989
Wojtak R., Mamon G. A., 2013, MNRAS, 428, 2407
Wojtak R., Lokas E. L., Mamon G. A., Gottl¨ober S., Prada F., Moles M., 2007, A&A, 466, 437
Wojtak R., Lokas E. L., Mamon G. A., Gottl¨ober S., Klypin A., Hoffman Y., 2008, MNRAS, 388, 815
Wojtak R., Lokas E. L., Mamon G. A., 2009, MNRAS, 399, 812
Wojtak R., Gottl¨ober S., Klypin A., 2013, MNRAS, 434, 1576
Yang X., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057
Yang X., Mo H. J., van den Bosch F. C., Jing Y. P., 2005a, MNRAS, 356, 1293
Yang X., Mo H. J., van den Bosch F. C., Weinmann S. M., Li C., Jing Y. P., 2005b, MNRAS, 362, 711
Yang X., Mo H. J., van den Bosch F. C., Pasquali A., Li C., Barden M., 2007, ApJ, 671, 153
Yang X., Mo H. J., van den Bosch F. C., 2008, ApJ, 676, 248
Yang X., Mo H. J., van den Bosch F. C., 2009, ApJ, 695, 900
Ye J.-N., Guo H., Zheng Z., Zehavi I., 2017, ApJ, 841, 45
York D. G. et al., 2000, AJ, 120, 1579
Yoshikawa K., Jing Y. P., B¨orner G., 2003, ApJ, 590, 654
Zaritsky D., White S. D. M., 1994, ApJ, 435, 599
Zehavi I. et al., 2005, ApJ, 630, 1
Zentner A. R., Berlind A. A., Bullock J. S., Kravtsov A. V., Wechsler R. H., 2005a, ApJ, 624, 505
Zentner A. R., Kravtsov A. V., Gnedin O. Y., Klypin A. A., 2005b, ApJ, 629, 219
Zentner A. R., Hearin A. P., van den Bosch F. C., 2014, MNRAS, 443, 3044
Zentner A. R., Hearin A. P., van den Bosch F. C., Lange J. U., Villarreal A., 2019, MNRAS, 485, 1196
Zu Y., Mandelbaum R., 2016, MNRAS, 457, 4360
Zwicky F., 1933, Helv. Phys. Acta, 6, 110
Here $\lambda_N$ is the expectation value for the total number of secondaries (satellites plus interlopers) around a primary of luminosity $L_{c_1}$ residing in a halo of mass $M_1$ at redshift $z_{c_1}$, and is given by

$$\lambda_N = \lambda_{\text{sat}}(M_1, L_{c_1}, z_{c_1}) = \lambda_{\text{int}}(M_1, L_{c_1}, z_{c_1}) + \lambda_{\text{res}}(L_{c_1}, z_{c_1}), \quad \text{(A4)}$$

with $\lambda_{\text{sat}}(M, L, z)$ and $\lambda_{\text{int}}(L, z)$ given by equations (15) and (19), respectively. For our particular model for the CLF we have that

$$\int_{L_{\text{min}}(c)}^{\infty} \Phi(L|M) \, dL = \frac{\phi_c^2(M)}{2} \Gamma \left( \alpha_{c_1} + 1, \frac{1}{2} \left[ \frac{L_{\text{min}}(c)}{L_c^*(M)} \right]^2 \right), \quad \text{(A5)}$$

with $\Gamma(a, x)$ the incomplete Gamma function, while the generalized NFW profile implies an aperture fraction (equation 16)

$$f_{ap}(M, L, z_c) = \frac{1}{\mu_p(c/R)} \int_{0}^{c/R} \frac{x^2-y}{(1+y)^{3-y}} \left( \xi[x R_{p}, R_{\text{max}}] - \xi[x R_{p}, R_{\text{min}}] \right) \, dx. \quad \text{(A6)}$$

Here $R_{\text{max}} = R_{\text{ap}}^\text{max}(L_c)$ and $R_{\text{min}} = R_{\text{ap}}(z_c)$, while

$$\mu_p(x) = \int_{0}^{\infty} \frac{y^{1-y} \, dy}{(1+y)^{3-y}}$$

and $\xi(r, R)$ is given by equation (18).

The expression for $Q_{\text{ap}}$ is given by

$$Q_{ik} = \sum_{j=1}^{N_{s_1}} \ln \left[ f_{\text{sat}}(\Delta V_{ij}, R_{ij}|L_{c_1}) + (1 - f_{\text{sat}}) P_{\text{int}}(\Delta V_{ij}, R_{ij}|M_1, L_{c_1}, z_{c_1}) \right]. \quad \text{(A8)}$$

Here $f_{\text{sat}} = f_{\text{sat}}(M_1, L_{c_1}, z_{c_1})$ is the interloper fraction (equation 25), $P_{\text{sat}}(\Delta V, R_{ij}|L, z)$ is given by equation (26), and $P_{\text{int}}(\Delta V, R_{ij}|M, L, z)$ is written as the product of two probabilities, $P(R_{ij}|M, L, z)$ (equation 46), and $P(\Delta V|R_{ij}, M, z)$ (equation 45). These depend on the normalized, projected number density, $\Sigma(R_{ij}|M, z)$ (equation 47), which for the gNFW profile (equation 42) is given by

$$\Sigma(R_{ij}|M, z) = \frac{Q(\gamma, R)}{2\pi R^2 \mu_p(c/R)}$$

where

$$Q(\gamma, R) \equiv \int_{r_p/R_{c_1}}^{c/R_{c_1}} \frac{y^{1-y} \, dy}{(1+y)^{3-y} \sqrt{y^2 - (r_p/R_{c_1})^2}}$$

and on the projected line-of-sight velocity dispersion, $\sigma_{\text{los}}^2(R_{ij}|M, z)$. In the case of a constant anisotropy (CA-model), the latter is given by equation (50), which for a gNFW tracer population in an NFW halo becomes

$$\sigma_{\text{los}}^2(R_{ij}|M, z) = V_{\text{vir}}^2 \frac{c}{\mu_p(c)} \frac{R_{\text{cut}}^{2\beta-2}}{Q(\gamma, R)} \int_{r_p/R_{c_1}}^{c} \frac{1 - \beta s_p^2}{s_p^2} \frac{s^{2-2\beta}}{\sqrt{s^2 - s_p^2}} \frac{s^{2-\gamma}}{(1+s_p/R_{c_1})^2} \mu_1(s, R) \, ds. \quad \text{(A11)}$$

Here $s_p = r_p/R_{c_1}$ is the projected radius in units of the scale radius of the dark matter host halo, $V_{\text{vir}}^2 = GM/r_{\text{vir}}$, and both $c_{\text{vir}}$ and $r_{\text{vir}}$ are functions of $M$ and $z$. For the former we adopt the fitting function of Macciò et al. (2008), ignoring scatter, and the latter follows directly from our definition of halo mass. In the case of an $\Omega M$ model, the expression for the line-of-sight velocity dispersion becomes

$$\sigma_{\text{los}}^2(R_{ij}|M, z) = V_{\text{vir}}^2 \frac{c}{\mu_p(c)} \frac{1}{Q(\gamma, R)} \int_{r_p/R_{c_1}}^{c} \frac{s^{2} + s_p^2 - s_p^2}{(s^2 + s_p^2)^2} \frac{s}{\sqrt{s^2 - s_p^2}} \frac{s^{2-\gamma}}{(1+s_p/R_{c_1})^2} \mu_1(s, R) \, ds. \quad \text{(A12)}$$

Here $s_p = r_p/R_{c_1}$ is the anisotropy radius in units of the scale radius of the dark matter host halo.

Along similar lines, when computing the log-likelihood $\ln L_0$ (equation 29), we compute the probabilities $P_{\text{ij}}$ that a primary with luminosity $L_d$ at redshift $z_d$ has zero secondaries (equation 30) using

$$P_{\text{ij}} = \frac{\sum_{k} w_k \exp(\tilde{E}_{ijk} - \lambda_{ijk})}{\sum_{k} w_k \exp(\tilde{E}_{ijk})}. \quad \text{(A13)}$$

Here

$$\tilde{E}_{ijk} = \ln M_k + \ln \left[ n(M_k, z_j) \right] - \ln \sqrt{\sigma_0(M_k)} \left( \frac{\log L_i - \log \bar{L}_i(M_k)}{2 \sigma_0(M_i)} \right)^2 \quad \text{(A14)}$$

and $\lambda_{ijk}$ is the expectation value for the total number of secondaries (satellites plus interlopers) around a primary of luminosity $L_{c_1}$ residing in a halo of mass $M_1$ at redshift $z_{c_1}$.

By keeping the Gaussian quadrature points $(M_i, w_i)$ fixed, we can pre-compute the first two and three terms of $\tilde{E}_{ijk}$ and $\tilde{F}_{ijk}$, respectively, as they are independent of the model, $\Theta$. Furthermore, if we keep the satellite distribution, $n_{\text{sat}}(r|M)$, and anisotropy, $\beta(r|M)$, fixed, $P(\Delta V|\theta)$,
Here we study the orbital anisotropy of dark matter subhaloes in the SMDPL simulation (Klypin et al. 2016). We select all subhaloes with peak halo masses \( M_{\text{peak}} > 3 \times 10^{10} h^{-1} M_{\odot} \) and calculate their coordinates \( r \) and velocities \( v \) with respect to the halo core identified by ROCKSTAR. We then determine the distance to the core \( r = |r| \), as well as the radial velocity \( v_r = v \cdot r / r \) and tangential velocity \( v_t = (v^2 - v_r^2)^{1/2} \). We use these to compute the local anisotropy parameter \( \beta \) (see equation 49) as a function of radius. The results are shown in Fig. C1, with different curves corresponding to different bins in host halo mass, as indicated. We confirm previous findings on the mass

APPENDIX B: THE IMPORTANCE OF ACCOUNTING FOR THE NUMBER OF SECONDARIES

As described in Section 3, the data vector used in our analysis is \( D = \sum_{i=1}^{N_c} D_i \), with

\[
D_i = \left( |\Delta v_j|, R_{ij} | j = 1, \ldots, N_{c,i} | \right) \left( M_{c,i}, z_{c,i} \right).
\]

Here \( N_{c,i} \) is the number of secondaries associated with primary \( i \), which in the model is used to inform the halo mass in the conditional probability \( P(M|L_{c,i}, z_{c,i}, N_{c,i}) \) given by equation (7). Using Bayes theorem, this implies that the likelihood for each primary is multiplied with the probability \( P(N_{c,i}|M, z_{c,i}, L_{c,i}) \), described in detail in Section 3.2.2.

In principle, we could also opt to not include \( N_i \) as part of the data vector. This implies that equation (7) simplifies to

\[
P(M|L_{c}, z_{c}) = \frac{P(L_{c}|M, z_{c}) P(M, z_{c})}{\int dM P(L_{c}|M, z_{c}) P(M, z_{c})},
\]

which removes \( P(N_{c,i}|M, z_{c,i}, L_{c,i}) \) from the likelihood. Effectively this means that the number of secondaries per primary is no longer used as a constraint in the inference. In addition, ignoring \( N_i \) also implies that we have no constraints on \( f_{\beta} \) (i.e. the fraction of primaries with zero secondaries). Although this speeds up the likelihood evaluation by roughly a factor of three, this is not a viable option, as it results in a large, systematic bias in the inferred galaxy–halo connection. This is demonstrated in Fig. B1, where we show the results of an analysis of our Tier-1 mock (cf. Fig. 4), in which we have ignored \( N_{c,i} \) as observational constraints. Clearly, the predicted masses at given central luminosity are now systematically and significantly too high, while the predicted luminosities of centrals at given halo mass are too low. The systematic bias arises from the non-negligible scatter in the galaxy–halo connection, and from the fact that more massive haloes contribute, on average, more satellites. As described in Section 2.3, unless one accounts for this, the massive haloes receive more weight in the analysis, causing an overestimate in the predicted halo masses. In the standard analysis of satellite kinematics, described in Section 2, this problem can be circumvented using host-weighting (ideally in combination with satellite weighting), while in \textbf{Basilisk} it is avoided by using the number of secondaries as constraints in the hierarchical Bayesian inference.

APPENDIX C: THE ORBITAL ANISOTROPY OF DARK MATTER SUBHALOES IN THE SMDPL SIMULATION

Here we study the orbital anisotropy of dark matter subhaloes in the SMDPL simulation (Klypin et al. 2016). We select all subhaloes with peak halo masses \( M_{\text{peak}} > 3 \times 10^{10} h^{-1} M_{\odot} \) and calculate their coordinates \( r \) and velocities \( v \) with respect to the halo core identified by ROCKSTAR. We then determine the distance to the core \( r = |r| \), as well as the radial velocity \( v_r = v \cdot r / r \) and tangential velocity \( v_t = (v^2 - v_r^2)^{1/2} \). We use these to compute the local anisotropy parameter \( \beta \) (see equation 49) as a function of radius. The results are shown in Fig. C1, with different curves corresponding to different bins in host halo mass, as indicated. We confirm previous findings on the mass
Figure C1. The orbital anisotropy $\beta$ as a function of the 3D radial distance $r$ for resolved subhaloes with $M_{\text{peak}} > 3 \times 10^{10} h^{-1} M_\odot$ in the SMDPL simulation. Different lines correspond to different host halo masses, as indicated by the colour bar. All radii are scaled by the virial radius of the host halo. Overall, the orbits of subhaloes are radially anisotropic, but with a radial dependence that depends strongly on host halo mass. The grey-shaded region, which is reproduced as a vertical band in Fig. 10, corresponds to $\beta = [0.08, 0.38]$ and roughly indicates the range of anisotropy parameters found in the majority of host haloes in the mass range $12 \lesssim \log[M_{\text{vir}}/(h^{-1} M_\odot)] \lesssim 14.5$. Dependence of the subhalo anisotropy parameter (Cuesta et al. 2008). For massive haloes with $M_{\text{vir}} \gtrsim 10^{13} h^{-1} M_\odot$, the velocity anisotropy increases with increasing radius (Diemand et al. 2004), while the trend reverses for less massive haloes (Sawala et al. 2017). Interestingly, the anisotropy at $r \sim 0.4 r_{\text{vir}}$ is roughly constant for all halo masses studied here. We checked that assuming different lower mass limits for $M_{\text{peak}}$ does not qualitatively affect these findings. Overall, if we were to assume a halo mass and radius-independent anisotropy parameter, as done in our modelling, $\beta \sim 0.2$ would best describe the anisotropy parameter of subhaloes and therefore also the anisotropy of satellites of the Tier-3 mocks. This is in excellent agreement with the constraints on $\beta$ as inferred from the Tier-3 mock data (cf. lower-left panel of Fig. 10). Note that the results of Fig. C1 are consistent with an Osipkov–Merritt model, which transits from isotropic at small radii to radially anisotropic at $r > r_a$, only for the most massive haloes ($M_{\text{vir}} \gtrsim 10^{14} h^{-1} M_\odot$). For less massive host haloes, the radial trend of $\beta$ is not adequately described by an $\Omega M$-model. This explains why, in the case of the Tier-3 mock, the $\Omega A$-model provides a significantly better fit to the data than the $\Omega M$-model (see discussion in 6.1).