

# Breaking Beta: A comparison of mass modelling methods for spherical systems

J. I. Read<sup>1\*</sup>, G. A. Mamon<sup>2</sup>, E. Vasiliev<sup>3,4,5</sup>, L. L. Watkins<sup>6,7,8</sup>,  
M. G. Walker<sup>9</sup>, J. Peñarrubia<sup>10</sup>, M. Wilkinson<sup>11</sup>, W. Dehnen<sup>11,12</sup>, P. Das<sup>4,1</sup>

<sup>1</sup>Department of Physics, University of Surrey, Guildford, GU2 7XH, UK

<sup>2</sup>Institut d'Astrophysique de Paris (UMR 7095: CNRS & Sorbonne Université), 75014 Paris, France

<sup>3</sup>Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK

<sup>4</sup>Rudolf Peierls Centre for Theoretical Physics, University of Oxford, OX1 3NP, UK

<sup>5</sup>Lebedev Physical Institute, Leninsky prospekt 53, Moscow, 119991, Russia

<sup>6</sup>AURA for the European Space Agency (ESA), ESA Office, Space Telescope Science Institute, 3700 San Martin Drive, Baltimore MD 21218, USA

<sup>7</sup>European Southern Observatory, Karl-Schwarzschild-Straße 2, D-85748 Garching bei München, Germany

<sup>8</sup>Department of Astrophysics, University of Vienna, Türkenschanzstraße 17, 1180 Vienna, Austria

<sup>9</sup>McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA

<sup>10</sup>Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh, EH9 3HJ, UK

<sup>11</sup>Department of Physics & Astronomy, University of Leicester, Leicester, LE1 7RH, UK

<sup>12</sup>Universitäts-Sternwarte der Ludwig-Maximilians-Universität, Scheinerstrasse 1, München D-81679, Germany

23 November 2020

## ABSTRACT

We apply four different mass modelling methods to a suite of publicly available mock data for spherical stellar systems. We focus on the recovery of the density and velocity anisotropy as a function of radius, using either line-of-sight velocity data only, or adding proper motion data. All methods perform well on isotropic and tangentially anisotropic mock data, recovering the density and velocity anisotropy within their 95% confidence intervals over the radial range  $0.25 < R/R_{1/2} < 4$ , where  $R_{1/2}$  is the half light radius. However, radially-anisotropic mocks are more challenging. For line-of-sight data alone, only methods that use information about the shape of the velocity distribution function are able to break the degeneracy between the density profile and the velocity anisotropy,  $\beta$ , to obtain an unbiased estimate of both. This shape information can be obtained through directly fitting a global phase space distribution function, by using higher order ‘Virial Shape Parameters’, or by assuming a Gaussian velocity distribution function locally, but projecting it self-consistently along the line of sight. Including proper motion data yields further improvements, and in this case, all methods give a good recovery of both the radial density and velocity anisotropy profiles.

**Key words:** cosmology: dark matter; galaxies: dwarf; galaxies: general; galaxies: haloes; galaxies: kinematics and dynamics

## 1 INTRODUCTION

Many stellar systems are spherical or mildly triaxial, from globular clusters (White & Shahl 1987) and tiny gas-poor dwarf spheroidal galaxies (dSphs) (Sanders & Evans 2017) to giant elliptical galaxies (Hubble 1926) and even galaxy clusters (e.g. Limousin et al. 2013). Building mass models of such systems has a wide range of astrophysical applications, from hunting for intermediate-mass and supermassive black holes (e.g. Verolme et al. 2002; Noyola et al. 2008; van der

Marel & Anderson 2010; Vitral & Mamon 2020) to probing the nature of dark matter (e.g. Walker & Peñarrubia 2011; Read et al. 2018, 2019), the stellar mass function (e.g. Cappellari et al. 2012), or the orbital velocity anisotropy (e.g. Mamon et al. 2019). The velocity anisotropy is an interesting quantity to determine as it encodes information about the assembly history of the system (e.g. Eggen et al. 1962). For spherical systems, it is typically described by the ‘anisotropy parameter’:

$$\beta(r) = 1 - \frac{\sigma_{\theta}^2(r)}{\sigma_r^2(r)}, \quad (1)$$

\* E-mail: justin.inglis.read@gmail.com

where  $r$  is the spherically symmetric radial coordinate,  $\sigma_\theta(r)$  is the tangential velocity dispersion,  $\sigma_r(r)$  is the radial velocity dispersion, and  $\beta(r) = 1$  for radial orbits, 0 for isotropic orbits and  $\beta(r) \rightarrow -\infty$  for circular orbits.

In all of the above applications, we would like to understand how model biases and systematic errors impact the results. Mock data – a dynamically realistic representation of the real data for which we know the true answer – provides an elegant way to address these questions. Such mocks can be very simple, reproducing all of the assumptions employed in the analysis on real data, or highly sophisticated, simulating the effect of data selection effects and/or departures from key model assumptions. For these latter, developing the mock can be a substantial task in its own right, discouraging broad and deep tests of our methodologies.

The goal of this article is to compare how four different mass-orbit modelling algorithms, applied to mock data for a spherical, single component, dwarf galaxy, recover the radial profiles of total density and velocity anisotropy. We focus on specific physical radii,  $R = [0.25, 0.5, 1, 2, 4] R_{1/2}$ , where  $R_{1/2}$  is the projected half-light radius of the tracer stars. And we consider both line-of-sight (LOS) velocity data for 1,000 and 10,000 tracer stars<sup>1</sup>, and proper motion data for 1,000 stars. The line-of-sight data are similar to the state-of-the-art data that are currently available for nearby dwarf galaxies (e.g. Walker et al. 2009), globular clusters (e.g. Watkins et al. 2015a; Kamann et al. 2018), giant elliptical galaxies (e.g. Napolitano et al. 2014), and clusters of galaxies (with large spectroscopic surveys such as SDSS [York et al. 2000] and GAMA [Robotham et al. 2010]). The proper motion data are similar to what is available for globular clusters (e.g. Bellini et al. 2014; Watkins et al. 2015b; Zocchi et al. 2017) and what will hopefully become available for dwarf galaxies from future planned missions (e.g. Strigari et al. 2007; Theia Collaboration et al. 2017; Massari et al. 2018, 2020).

Our mock data suite is publicly available on the GAIA CHALLENGE wiki site<sup>2</sup> and has already been used to test a number of methods in the literature (Walker & Peñarrubia 2011; Geringer-Sameth et al. 2015; Bonnavard et al. 2015; Richardson & Fairbairn 2014; Zhu et al. 2016; Genina & Fairbairn 2016; Read & Steger 2017; Chiappo et al. 2017; Diakogiannis et al. 2019). These mock data will continue to grow and improve in sophistication over time. Our central philosophy is that *it is much easier to produce a sophisticated mock, than to produce a sophisticated mass modelling tool*. Using such mocks may demonstrate that simple models – e.g. those that assume spherical symmetry – work sufficiently well to test or rule out interesting models, or estimate a given parameter of interest. In this case, increasing the sophistication of the model is not required. Mock data tests can also help us identify problems in the real data – for example, if confidence intervals on the model are much smaller than expected from the mock.

In this paper, we assess the performance of four mass

modelling methods described in §3 through their recovery of the radial density and velocity anisotropy profiles. Three of these methods, GRAVSPHERE, MAMPOSS<sub>T</sub>, and DISCRETEJAM, are based on the Jeans (1922) equation of local dynamical equilibrium, while one, AGAMA, is based on fitting a global distribution function. For stationary spherical systems without streaming motions, the Jeans equation is given by (e.g. Binney & Tremaine 2008):

$$\frac{d(\nu(r)\sigma_r^2(r))}{dr} + 2\frac{\beta(r)}{r}\nu(r)\sigma_r^2(r) = -\nu(r)\frac{GM(r)}{r^2}, \quad (2)$$

where  $\nu(r)$  is the tracer density,  $M(r)$  is the total enclosed mass,  $\beta(r)$  is the anisotropy parameter (1), and  $G$  is Newton’s gravitational constant. Jeans methods have the advantage that they can be efficiently implemented, allowing a wide range of models to be explored (e.g. Read & Steger 2017). Furthermore, since they do not require any assumption about the form of the distribution function, they can be constructed to be formally unbiased (e.g. Read 2014; Read & Steger 2017). However, disadvantages include the fact that the fits can, in principle, require a formally negative (i.e. unphysical) distribution function (e.g. An & Evans 2006). By contrast, methods like AGAMA have the advantage that they provide a full distribution function that is positive definite and that can be easily convolved with errors, binary contamination (e.g. McConnachie & Côté 2010; Spencer et al. 2018) and survey selection functions. However, if the chosen parameterisation of the distribution function does not contain the true solution, such methods run the risk of becoming biased (e.g. Garbari et al. 2011).

We focus here on a suite of spherical, idealised, mocks with minimal measurement uncertainties as a starting point. As we shall see, already this presents a challenge for some methods, primarily due to the mass-anisotropy degeneracy (e.g. Binney & Mamon 1982; Read & Steger 2017). More sophisticated mocks that break spherical symmetry and include the impact of unbound tidally stripped stars, lower numbers of tracer stars, and multiple independent tracer populations are also available through the GAIA CHALLENGE wiki site. These have already been presented in the literature for the GRAVSPHERE method (Read & Steger 2017; Read et al. 2018; Gregory et al. 2019), with the result that – at least for 1,000 tracers with small velocity uncertainties – these additional complications do not dominate the error budget. Other groups have reached similar conclusions with their own mock data tests, including on cosmologically realistic mocks (e.g. Sanchis et al. 2004; Wojtak et al. 2009; Mamon et al. 2013; Kowalczyk et al. 2013; Hayashi et al. 2020; Genina et al. 2020). We will consider such more sophisticated tests on the methods presented here in future work.

This paper is organised as follows. In §2, we describe the mock data suite and how it was set up. In §3, we describe the different mass modelling techniques we apply to these mock data. In §4, we apply several modelling methods to the spherical mocks to examine the accuracy of their inferred density  $\rho(r)$  and velocity anisotropy  $\beta(r)$  profiles. Finally, in §5, we discuss our results and present our conclusions.

<sup>1</sup> We define ‘tracer’ stars as massless particles orbiting in the combined gravitational potential of all stars, gas, dark matter (DM) etc. in a stellar system.

<sup>2</sup> <http://astrowiki.ph.surrey.ac.uk/dokuwiki/> (under the ‘Spherical & Triaxial’ suite tab).

Label	$[r_*, \alpha_*, \beta_*, \gamma_*]$	$[(\rho_0, r_D, \alpha_D, \beta_D, \gamma_D)]$	$\beta_0, \beta_\infty, n, r_a$	$R_{1/2}$
PlumCuspIso	[0.25, 2, 5, 0.1]	[6.4, 1, 1, 3, 1]	[0, 0, 1, 0]	0.25
PlumCoreIso	[1, 2, 5, 0.1]	[40, 1, 1, 3, 0]	[0, 0, 1, 0]	1.0
NonPlumCuspIso	[0.25, 2, 5, 1]	[6.4, 1, 1, 3, 1]	[0, 0, 1, 0]	0.20
NonPlumCoreIso	[1, 2, 5, 0.1]	[40, 1, 1, 3, 0]	[0, 0, 1, 0]	0.79
PlumCuspOM	[0.1, 2, 5, 0.1]	[6.4, 1, 1, 3, 1]	[0, 1, 2, 0.10]	0.1
PlumCoreOM	[0.25, 2, 5, 0.1]	[40, 1, 1, 3, 0]	[0, 1, 2, 0.25]	0.25
NonPlumCuspOM	[0.1, 2, 5, 1]	[6.4, 1, 1, 3, 1]	[0, 1, 2, 0.10]	0.079
NonPlumCoreOM	[0.25, 2, 5, 1]	[40, 1, 1, 3, 0]	[0, 1, 2, 0.25]	0.20
PlumCuspTan	[0.5, 0.5, 5, 0.1]	[2.39, 2, 1, 4, 1]	[-0.5, -0.5, 2, 0]	0.86
PlumCoreTan	[1.75, 0.5, 5, 0.1]	[3.0, 4.0, 1, 4, 0]	[-0.5, -0.5, 2, 0]	3.0
NonPlumCuspTan	[0.5, 0.5, 5, 1.0]	[2.39, 2, 1, 4, 1]	[-0.5, -0.5, 2, 0]	0.41
NonPlumCoreTan	[1.75, 0.5, 5, 1.0]	[3.0, 4.0, 1, 4, 0]	[-0.5, -0.5, 2, 0]	1.43

**Table 1.** Parameters of the spherical mock data that we use in this paper. The columns give, from left to right: the mock data label; the tracer density parameters (see equation 3); the DM halo density parameters (see equation 3); the velocity anisotropy profile; and the projected half stellar mass radius. The length units are given in kpc, while  $\rho_0$  has units of  $10^7 M_\odot \text{kpc}^{-3}$ . The mock labelling convention is: <light profile><Cusp/core><Anisotropy>, where: the <light profile> can be Plummer-like (Plum;  $\gamma_* = 0.1$ ) or cusped (NonPlum;  $\gamma_* = 1$ ); the DM halo can be cusped (Cusp;  $\gamma_D = 1$ ) or cored (Core;  $\gamma_D = 0$ ); and the <Anisotropy> can be isotropic (Iso;  $\beta = 0$ ), of generalized Osipkov-Merritt form (OM; Osipkov 1979; Merritt 1985, see equation 4) or tangentially anisotropic (Tan;  $\beta = -0.5$ ).

## 2 MOCK DATA

In this section, we describe our mock data; the full ‘default spherical’ suite is summarised in Table 1 and is publicly available at the GAIA CHALLENGE wiki site<sup>2</sup>. All of the mock data have a single population of massless tracer stars orbiting within a host DM halo. They are available either without errors or with normally distributed errors with a standard deviation of  $2 \text{ km s}^{-1}$  on the line-of-sight velocities for each star (larger errors can easily be added to explore their effect, but in this paper we will focus on the case where the measurement uncertainties do not dominate over the sampling error). We also include proper motion data, assuming similarly small measurement errors.

As discussed in §1, the Gaia Challenge suite also includes a host of more realistic triaxial, tidally stripped and split-population mocks. We will consider more realistic mocks with individual star-by-star velocity errors, binary star contamination, contamination from Milky Way stars, unbound tidally stripped stars, and a realistic survey selection function, in future work. We note, however, that – at least for mock galaxies similar to the Milky Way ‘classical’ dwarfs – previous works have found these additional effects to be small (e.g. Walker & Peñarrubia 2011; Read et al. 2018; Genina et al. 2020).

We label the mocks using the following naming convention:

< light profile >< DM cusp/core >< Anisotropy >

where: the <light profile> can be Plummer-like (Plummer 1911; Plum) or cusped (NonPlum); the DM halo can be cusped (Cusp) or cored (Core); and the <Anisotropy> can be isotropic (Iso), of generalized Osipkov-Merritt form (OM; see next section) or tangentially anisotropic (Tan).

The isotropic and radially anisotropic spherical mocks are a subset of the mocks presented in Walker & Peñarrubia (2011) (see §2.1); the tangentially anisotropic mocks were generated using the method outlined in Dehnen (2009) (see §2.2). In all cases, we assume a double-power-law profile (Zhao 1996) for both the tracer distribution and the host

DM halo:

$$\rho_X(r) = \rho_0 \left( \frac{r}{r_X} \right)^{-\gamma_X} \left[ 1 + \left( \frac{r}{r_X} \right)^{\alpha_X} \right]^{(\gamma_X - \beta_X)/\alpha_X} \quad (3)$$

where  $\rho_0$  is a normalisation parameter;  $r_X$  sets the scale length;  $\gamma_X$  is the inner asymptotic logarithmic slope;  $\beta_X$  is the outer asymptotic logarithmic slope; and  $\alpha_X$  controls the sharpness of the transition at  $r_X$ . For the tracers, we use the notation:  $\nu = \nu_*(r; \nu_0, r_*, \alpha_*, \beta_*, \gamma_*)$ ; for the DM we write similarly:  $\rho_D = \rho_D(r; \rho_0, r_D, \alpha_D, \beta_D, \gamma_D)$ .

### 2.1 Radially anisotropic mocks

The velocity anisotropy coefficient (1) is assumed to depend on radius as follows:

$$\beta(r) = \frac{\beta_\infty r^n + \beta_0 r_a^n}{r^n + r_a^n}, \quad (4)$$

where  $\beta_0$  is the asymptotic anisotropy at small radii,  $\beta_\infty$  is the asymptotic anisotropy at large radii,  $r_a$  is the transition radius, and  $n$  controls the sharpness of the transition. A familiar special case is the Osipkov–Merritt anisotropy profile (Osipkov 1979; Merritt 1985), used already above for the mock data, in which  $\beta_0 = 0$ ,  $\beta_\infty = 1$ , and  $n = 2$ .

For the radially-anisotropic mocks, the tracers are set up in equilibrium inside their host DM halo assuming an OM distribution function, as in Walker & Peñarrubia (2011).

### 2.2 Tangentially anisotropic mocks

The tangentially-anisotropic mocks are set up using the ‘made to measure’ code from Dehnen (2009) that is an evolution of the method described in Syer & Tremaine (1996). In such methods, an  $N$ -body system is evolved for some time under its own self-gravity, adjusting the masses of each particle so as to move towards a target phase space distribution. The tangentially anisotropic mocks also assume the form given in equation 3 for the tracer and DM density profiles.

Tangential anisotropy is not expected theoretically, unless significant angular momentum is imparted to the system through mergers, or the system is orbiting within a strong

tidal field<sup>3</sup> (e.g. [Read et al. 2006b](#)). Given the lack of any strong theoretical motivation for a particular form of tangential anisotropy, we assume a constant  $\beta = -0.5$  at all radii to explore whether tangential mocks can or cannot be successfully recovered.

### 3 METHODS

In this section, we briefly describe each of the four mass-orbit modelling methods that will be applied to the mock data.

#### 3.1 GRAVSPHERE

GRAVSPHERE<sup>4</sup> is described and tested in detail in [Read & Steger \(2017\)](#), [Read et al. \(2018\)](#) and [Genina et al. \(2020\)](#). It fits the line-of-sight velocity variance, which can be written by combining the Jeans equation (2) with an equation projecting the velocities along the line of sight, yielding ([Binney & Mamon 1982](#)):

$$\frac{\Sigma(R) \sigma_{\text{LOS}}^2(R)}{2} = \int_R^\infty \left(1 - \beta \frac{R^2}{r^2}\right) \nu \sigma_r^2 \frac{r \, dr}{\sqrt{r^2 - R^2}}, \quad (5)$$

where  $\Sigma(R)$  denotes the tracer surface mass density at projected radius  $R$ .

The radial velocity variance is given by ([van der Marel 1994](#), in this form by [Mamon & Lokas 2005](#)):

$$\sigma_r^2(r) = \frac{1}{\nu(r)g(r)} \int_r^\infty \frac{GM(\tilde{r})\nu(\tilde{r})}{\tilde{r}^2} g(\tilde{r}) d\tilde{r} \quad (6)$$

where

$$g(r) = \exp\left(2 \int_0^r \frac{\beta(\tilde{r})}{\tilde{r}} d\tilde{r}\right) \quad (7)$$

and  $M(r)$  is the cumulative mass of the stellar system (due to all stars, gas, DM etc.).

GRAVSPHERE uses a free-form, i.e. non-parametric, model for  $M(r)$  that comprises a contribution from all visible matter and a contribution from ‘dark matter’ that is described by a sequence of power laws defined on a set of radial bins. By default, these bins are defined at  $R = [0.25, 0.5, 1, 2, 4] R_{1/2}$  where  $R_{1/2}$  is the projected half light radius of the tracer stars. (This can be changed if the data quality warrant it, but in this paper we will use only this default choice.) The tracer light profile is also non-parametric, using a series sum of Plummer spheres, as in [Rojas-Niño et al. \(2016\)](#). This has the advantage that the mapping between the spherically-averaged tracer density,  $\nu(r)$  and the projected light profile,  $\Sigma_*$  is analytic ([Read & Steger 2017](#)). Finally, the velocity anisotropy,  $\beta(r)$ , is assumed to have a form (4) that ensures broad generality while making the function  $g(r)$  (equation 7) analytic. For assigning priors on

the parameters in equation 4, we transform  $\beta(r)$  to a symmetrised velocity anisotropy, as in [Read et al. \(2006a\)](#) and [Read & Steger \(2017\)](#):

$$\tilde{\beta} = \frac{\sigma_r^2 - \sigma_t^2}{\sigma_r^2 + \sigma_t^2} = \frac{\beta}{2 - \beta} \quad (8)$$

This is bounded on  $-1 < \tilde{\beta} < 1$  ( $\tilde{\beta} = -1$  corresponds to full tangential anisotropy;  $\tilde{\beta} = 1$  to full radial anisotropy; and  $\tilde{\beta} = 0$  to isotropy), unlike  $\beta$  that tends to  $-\infty$  for full tangential anisotropy.

GRAVSPHERE can also use split population or proper motion data, where available. By default, it also fits the two higher order ‘Virial Shape Parameters’ (VSPs; [Merrifield & Kent 1990](#); [Richardson & Fairbairn 2014](#)):

$$v_{s1} = \frac{2}{5} \int_0^\infty GM(5 - 2\beta) \nu \sigma_r^2 r \, dr \quad (9)$$

$$= \int_0^\infty \Sigma \langle v_{\text{LOS}}^4 \rangle R \, dR \quad (10)$$

and:

$$v_{s2} = \frac{4}{35} \int_0^\infty GM(7 - 6\beta) \nu \sigma_r^2 r^3 \, dr \quad (11)$$

$$= \int_0^\infty \Sigma \langle v_{\text{LOS}}^4 \rangle R^3 \, dR. \quad (12)$$

The key advantage of these VSPs is that they involve fourth-order moments of the line-of-sight velocities ( $\langle v_{\text{LOS}}^4 \rangle$ ), but depend only on  $\beta(r)$  and not on its fourth-order counterparts ([Merrifield & Kent 1990](#); [Richardson & Fairbairn 2014](#); [Read & Steger 2017](#)). Thus,  $v_{s1}$  and  $v_{s2}$  allow us to obtain additional constraints on  $\beta(r)$  via the line-of-sight velocities, alleviating the mass-anisotropy degeneracy between the spherically averaged density profile  $\rho(r)$  and the velocity anisotropy  $\beta(r)$ .

The errors on  $\sigma_{\text{LOS}}$ ,  $\langle v_{\text{LOS}}^4 \rangle$ ,  $v_{s1}$  and  $v_{s2}$  are estimated as in [Read et al. \(2018\)](#), using a Monte-Carlo method that incorporates the individual measurement errors on each star.

Finally, if proper motion data are available, GRAVSPHERE can fit also for the radial and tangential plane-of-sky velocity dispersions,  $\sigma_{\text{POSr}}$  and  $\sigma_{\text{POSt}}$ , satisfying (e.g. [Strigari et al. 2007](#); [van der Marel & Anderson 2010](#); [Read & Steger 2017](#)):

$$\frac{\Sigma(R) \sigma_{\text{POSr}}^2(R)}{2} = \int_R^\infty \left(1 - \beta + \beta \frac{R^2}{r^2}\right) \frac{\nu(r) \sigma_r^2 r \, dr}{\sqrt{r^2 - R^2}}, \quad (13)$$

$$\frac{\Sigma(R) \sigma_{\text{POSt}}^2(R)}{2} = \int_R^\infty (1 - \beta) \frac{\nu(r) \sigma_r^2 r \, dr}{\sqrt{r^2 - R^2}}. \quad (14)$$

This provides an alternative route to breaking the mass-anisotropy degeneracy since equations (5), (13) and (14) each have different dependencies on  $\beta$  (see [Read & Steger 2017](#) for further details).

GRAVSPHERE fits the above model to the surface density profile of tracer stars,  $\Sigma_*(R)$ , their line-of-sight projected velocity dispersion profile  $\sigma_{\text{LOS}}(R)$  and their VSPs using the EMCEE affine invariant Markov Chain Monte Carlo (MCMC) sampler from [Foreman-Mackey et al. \(2013\)](#). We assume uncorrelated Gaussian errors such that the Likelihood function is given by  $\mathcal{L} = \exp(-\chi^2/2)$ . We use as default 1,000 walkers, each generating 5,000 models and we throw out the first half of these as a conservative ‘burn in’

<sup>3</sup> Note that in the case of strong tides, such tangential anisotropy would indicate a departure from the steady-state pseudo-equilibrium assumed by all of the methods explored here.

<sup>4</sup> A public version of the code, PYGRAVSPHERE, can be downloaded from <https://github.com/AnnaGenina/pyGravSphere>.

criteria. Please see [Read & Steger \(2017\)](#) for further details of our likelihood function, methodology and priors.

### 3.2 DISCRETEJAM

The DISCRETEJAM method uses an alternative approach to solve the Jeans equations, which were laid out in [subsection 3.1](#). There are three key features of this method: (1) both the tracer number density profile  $\nu(r)$  and the underlying mass density profile  $\rho(r)$  are parametrised in the form of Multi-Gaussian Expansions (MGEs, [Emsellem et al. 1994](#)); (2) we fit to the discrete position measurements for each star, not to a binned tracer-density profile; and (3) we fit to the discrete velocity measurements for each star, not to a binned velocity-dispersion profile.

We model the dwarf galaxies using the Jeans Anisotropic MGE (JAM) models ([Cappellari 2008](#); [Watkins et al. 2013](#); [Cappellari 2015](#)). For the spherical mocks that are the focus of this paper, we use the spherical JAM models, which assume spherical symmetry and allow for anisotropy (defined in spherical coordinates) but do not include rotation.

An MGE is characterised by a set of Gaussian components, each with a width and a weight. If the tracer distribution is made up of  $N_t$  components where each component  $j$  has a weight  $\nu_j$  and a width  $s_j$ , then the tracer density profile of the MGE is:

$$\nu(r) = \sum_{j=1}^{N_t} \nu_j \exp\left(-\frac{r^2}{2s_j}\right). \quad (15)$$

Similarly, the mass density  $\rho(r)$  can be parametrised by a set of  $N_m$  Gaussians with each component  $k$  having weight  $\rho_k$  and width  $s_k$ . The density profiles can be fit analytically (where we fit MGEs to assumed functional forms and fit for the parameters of the functional forms) or non-parametrically (where we fit the MGE properties directly). Here we choose an analytic approach, but see [Hénault-Brunet et al. \(2019\)](#) for a non-parametric methodology.

We assume that both the tracer density and DM density are described by [Equation 3](#). The models are insensitive to the normalisation of the tracer density  $\nu_0$ , so this leaves us with four free parameters relating to the tracer distribution ( $r_*$ ,  $\alpha_*$ ,  $\beta_*$ ,  $\gamma_*$ ), and five relating to the DM distribution ( $\rho_0$ ,  $r_D$ ,  $\alpha_D$ ,  $\beta_D$ ,  $\gamma_D$ ). We assume that the anisotropy profile is of the generalized OM form ([Equation 4](#)) with fixed  $n = 2$ . As in the GRAVSPHERE method, we symmetrise  $\beta_0$  and  $\beta_\infty$  using [Equation 8](#) to avoid the high anti-symmetry and infinite lower bound for  $\beta$ . This leaves three free components for the anisotropy: ( $r_a$ ,  $\tilde{\beta}_0$ ,  $\tilde{\beta}_\infty$ ).

We sample scale density  $\rho_0$  and scale radii  $r_*$ ,  $r_D$  and  $r_a$  in log-space as negative values would be unphysical, these parameters have the potential to explore many orders of magnitude in size when model fitting, and a dimensionless parameter space is generally more efficient to search. The other parameters are already dimensionless by definition.

We use an unbounded flat prior on  $\log \rho_0$ . We restrict the range of the tracer scale radii to  $-2 \leq \log_{10}(r_*/\text{kpc}) \leq 1$  and the DM scale radii to  $0.25 \leq r_D/r_* \leq 50$ , but otherwise assume flat priors within these ranges. For transition parameters  $\alpha_*$  and  $\alpha_D$ , we use a flat prior in  $0.5 \leq (\alpha_*, \alpha_D) \leq 2.5$ , this eliminates models that change too sharply or too slowly.

For the outer density slopes, we use flat priors in  $2 \leq \beta_* \leq 7$  and  $2 \leq \beta_D \leq 5$ , where the upper limits restrict the steepness based on observed values. Models with  $(\beta_*, \beta_D) \leq 3$  are unphysical as they are not finite, however, some mocks have outer slopes of 3, meaning that the correct answer is at the edge of the parameter space. To allow the fits to fully explore the parameter space, we use lower limits of 2 for the outer slopes.

For the inner density slopes, we use flat priors in  $-0.5 \leq \gamma_* \leq \beta_*$  and  $-0.5 \leq \gamma_D \leq \beta_D$ , where the upper limits ensure that the inner slope is shallower than the outer slope. Formally, a lower limit of 0 on the inner slopes we would be appropriate here, otherwise we have a negative density component in the MGE fit. However, as above, some mocks have inner slopes of 0, thus at the edge of the parameter space. So to allow the models to fully explore the region of parameter space around the true value, we use lower limits of  $-0.5$  on the inner slopes.

We assume a flat prior on the anisotropy scale radii in  $0.5 \leq r_a/r_* \leq 2$ ; both the isotropic and tangential mocks have flat anisotropy profiles in which case the notion of an anisotropy transition radius becomes somewhat meaningless, with this choice of prior we give the models some flexibility, but avoid fitting a potentially undefined parameter. In theory,  $-1 \leq \tilde{\beta} \leq 1$ . In practice, we set a lower bound of  $\beta = -40$ , which corresponds to  $\tilde{\beta} \sim -0.95$  as extremely tangential models with  $\tilde{\beta} \sim -1$  are computationally challenging (and physically unlikely). We use a flat prior on  $\tilde{\beta}$  within these limits.

For a given set of density parameters, we calculate  $\nu(r)/\nu_0$  or  $\rho(r)$  using [Equation 3](#). To this we fit an MGE using MGEFIT ([Cappellari 2002](#)), which we then deproject to obtain the projected surface mass and tracer density profiles as needed for the JAM models.<sup>5</sup> To each component  $j$  of the tracer MGE we assign an anisotropy  $\beta_j$  by calculating the value of the generalized OM profile at the radius of each Gaussian component  $s_j$ .

Our observables are the spatial distribution of the tracer particles and the velocity measurements of the tracer particles. We fit to both discretely, that is we do not bin the tracers but consider each star individually. We follow the approach laid out in [Watkins et al. \(2013\)](#), modified as described below. This method is flexible and can be used to fit only LOS velocity information, only proper motions, or full 3D velocity information. The original method was designed to fit a binned surface brightness profile, whereas here we fit the tracer density discretely as well.

Consider a tracer particle  $i$  at a projected distance  $R_i$  from the centre of the dwarf with velocity  $v_{X,i} \pm \delta_{X,i}$ , where X can be the Right Ascension PM ( $\alpha$ ), the Declination PM ( $\delta$ ) or the LOS ( $z$ ), and the uncertainties  $\delta_{X,i}$  are assumed to be Gaussian. To fit the spatial distribution of the tracers, we use the projected tracer density MGE to calculate the normalised probability  $P_i = P(R_i)$  of observing a star at projected radius  $R_i$ . Then the likelihood of all  $N$  tracer

<sup>5</sup> A python package MGETOOLS for manipulating MGEs is available at <http://www.github.com/lauralwatkins/mgetools>.

particles is:

$$\mathcal{L}_\star = \prod_{i=1}^N P_i. \quad (16)$$

To fit the velocity measurements, we assume that the LOS velocity distribution at a given radius is Gaussian with mean velocity 0 (due to the assumptions of spherical symmetry and no rotation) and velocity dispersion  $\sigma_{X,i}$ , which we calculate from the JAM model. Then the likelihood of all  $N$  measurements is:

$$\mathcal{L}_{\text{kin},X} = \prod_{i=1}^N \frac{1}{\sqrt{2\pi(\delta_{X,i}^2 + \sigma_{X,i}^2)}} \exp\left(\frac{-v_{X,i}^2}{2(\delta_{X,i}^2 + \sigma_{X,i}^2)}\right), \quad (17)$$

which accounts for the fact that the observed dispersion is a convolution of the true dispersion  $\sigma_{X,i}$  and the measurement uncertainties  $\delta_{X,i}$ . When using only LOS velocity information, the total likelihood  $\mathcal{L}$  for the spatial and kinematic distributions combined is then:

$$\mathcal{L} = \mathcal{L}_\star \mathcal{L}_{\text{kin},z}. \quad (18)$$

When using PM and LOS kinematic information, the total likelihood  $\mathcal{L}$  becomes:

$$\mathcal{L} = \mathcal{L}_\star \mathcal{L}_{\text{kin},\alpha} \mathcal{L}_{\text{kin},\delta} \mathcal{L}_{\text{kin},z}, \quad (19)$$

assuming that the measurements and uncertainties are not correlated.

The posterior probability of the model given the data is then obtained by multiplying together the likelihood and parameter priors. By maximising the posterior, we can locate the family of models that best fit the data. To efficiently explore our parameter space and locate the high-posterior region, we use EMCEE (Foreman-Mackey et al. 2013) with 100 walkers and for 20,000 steps.

### 3.3 MAMPOSSt

MAMPOSSt (Mamon, Biviano, & Boué 2013) is similar to DISCRETEJAM, as it fits the full distribution of the individual tracers in projected phase space. But while DISCRETEJAM assumes that the line-of-sight velocities are Gaussian-distributed, MAMPOSSt avoids this assumption, since anisotropy affects the shape of the line-of-sight velocity distribution (Merritt 1987), and assumes instead that the local three-dimensional velocity distribution is Gaussian-distributed. More precisely, MAMPOSSt assumes parametric forms for the mass, tracer density and anisotropy profiles. In situations where only line-of-sight velocities are available, the density of tracers in projected phase space (projected radius  $R$  and the line-of-sight velocity  $v_{\text{LOS}}$ ) is the integral of the local velocity distributions along the line of sight:

$$g(R, v_{\text{LOS}}) = 2 \int_R^\infty h(v_{\text{LOS}}|R, r) \nu(r) \frac{r \, dr}{\sqrt{r^2 - R^2}}, \quad (20)$$

MAMPOSSt assumes that the *local* velocity distribution function  $h$  is a Gaussian of zero mean:

$$h(v_{\text{LOS}}|R, r) = \mathcal{G}(v_{\text{LOS}}, 0, \sigma_{\text{LOS}}^2), \quad (21)$$

where  $\sigma_{\text{LOS}}^2(R, r) = [1 - \beta(r)R^2/r^2] \sigma_r^2$  as in the integrand of equation (5). The likelihood is given by:

$$-\ln \mathcal{L} = - \sum_i \ln q(R_i, v_i | \boldsymbol{\theta}), \quad (22)$$

where  $i$  are the indices of the individual tracers,  $v_i$  is the LOS velocity for tracer particle  $i$ ,  $\boldsymbol{\theta}$  is the vector of parameters, while  $q$  is the probability density:

$$q(R, v) = \frac{2\pi R g(R, v)}{N_p(R_{\text{max}}) - N_p(R_{\text{min}})}, \quad (23)$$

where  $N_p(R) = \int_0^R 2\pi R' \Sigma(R') \, dR'$  is the predicted number of tracers expected within the projected radius  $R$ .

MAMPOSSt has recently been generalized to include handling of proper motion data (MAMPOSSt-PM, Mamon & Vitral, in prep.). The full local velocity distribution function is assumed to be the product of three Gaussians:

$$\begin{aligned} h(\mathbf{v}|R, r) &= \mathcal{G}(v_{\text{LOS}}, 0, \sigma_{\text{LOS}}^2(R, r)) \\ &\times \mathcal{G}(v_{\text{POSr}}, 0, \sigma_{\text{POSr}}^2(R, r)) \\ &\times \mathcal{G}(v_{\text{POST}}, 0, \sigma_{\text{POST}}^2(R, r)), \end{aligned} \quad (24)$$

where the plane-of-sky velocity variances are  $\sigma_{\text{POSr}}^2(R, r) = [1 - \beta(r) + \beta(r)R^2/r^2] \sigma_r^2$  and  $\sigma_{\text{POST}}^2(R, r) = [1 - \beta(r)] \sigma_r^2$  as given in the integrands of equations (13) and (14), respectively, and where  $\sigma_r^2$  is the solution of the Jeans equation taken from equations (6) and (7).

The likelihood is now given by:

$$-\ln \mathcal{L} = - \sum_i \ln q(R_i, \mathbf{v}_i | \boldsymbol{\theta}), \quad (25)$$

where  $i$  are the indices of the individual tracers,  $\mathbf{v}_i$  is the velocity vector for tracer particle  $i$ ,  $\boldsymbol{\theta}$  is the vector of parameters, while the probability density  $q$  is now

$$q(R, \mathbf{v}) = \frac{2\pi R g(R, \mathbf{v})}{N_p(R_{\text{max}}) - N_p(R_{\text{min}})}. \quad (26)$$

For both MAMPOSSt and MAMPOSSt-PM, measurement uncertainties are added in quadrature to the predicted line-of-sight and plane-of-sky velocity dispersions in the local velocity distribution function  $h$ .

The mass profile is assumed to follow Equation (3), for which the enclosed total mass (here assuming only dark matter contributes) is

$$\begin{aligned} M(r) &= \frac{4\pi}{3-\gamma} \rho_0 r^3 \left(\frac{r_D}{r}\right)^\gamma \\ &\times {}_2F_1\left[\frac{3-\gamma}{\alpha}, \frac{\beta-\gamma}{\alpha}, \frac{3-\gamma}{\alpha} + 1, -\left(\frac{r}{r_D}\right)^\alpha\right], \end{aligned} \quad (27)$$

where  $\alpha = \alpha_D$ ,  $\beta = \beta_D$ ,  $\gamma = \gamma_D$ , and  ${}_2F_1$  is the ordinary hypergeometric function. The MAMPOSSt analyses assume  $\alpha_D = 1$  and free inner and outer slopes. The tracer density profile is also described by Equation (3), with  $\beta_* = 5$ ,  $\alpha_* = 2$  and a free inner slope  $\gamma_*$ . Finally, the velocity anisotropy profile has the form of Equation (4) with  $n = 2$  and free  $\beta_0$ ,  $\beta_\infty$  and  $r_a$ .

The nine free parameters are thus the mass normalisation, the mass scale radius  $r_2$ , the DM inner and outer slopes  $\gamma_D, \beta_D$ , the stellar scale radius  $r_*$  and inner density slope  $\gamma_*$ , the inner and outer anisotropies  $\beta_0$  and  $\beta_\infty$ , and the anisotropy radius  $r_a$ .

The minimisation is performed with the CosmoMC (Lewis & Bridle 2002) MCMC sampler, using 6 chains with 90 000 elements, discarding the first 15 000 elements of each chain for the analysis. We adopted  $R_{\text{min}} = 0$  and  $R_{\text{max}} = 1.9$  kpc for our allowed range of projected radii. Flat priors were used on the log masses, log scales, and on

the indices and symmetrised anisotropies, with ranges listed in Table 2.

### 3.4 DF fitting with the AGAMA library

The AGAMA library (Vasiliev 2019) is a C++ library for constructing galaxy models using distribution functions (DF). A DF,  $f(\mathbf{w})$ , fully describes the distribution of a population of stars with phase-space coordinates,  $\mathbf{w} \equiv \{\mathbf{x}, \mathbf{v}\}$ . In a steady state,  $f$  only depends on the integrals of motion,  $\mathbf{I}(\mathbf{w})$ , supported by the potential. The log-likelihood of a model  $\mathcal{M}$  characterised by the potential  $\Phi(r)$  and the DF  $f(\mathbf{I})$  is given by the sum of log-probabilities of each star being drawn from this DF (if we assume that all stars could be observed, i.e., the selection function is unity):

$$\ln \mathcal{L} = \sum_{i=1}^{N_{\text{stars}}} \ln f[\mathbf{I}(\mathbf{w}_i; \Phi)]. \quad (28)$$

In the case of incomplete data we need to marginalise over the probability of each star by integrating over missing phase-space coordinates. In the case of errors, we need to convolve with the distribution of errors. For instance, if we have only the projected radius  $R$  and the line-of-sight velocity  $v_z$  that carries a Gaussian error  $\delta v_z$ , the above expression is modified to

$$\ln \mathcal{L} = \sum_{i=1}^{N_{\text{stars}}} \ln \left\{ \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} dv_x \int_{-\infty}^{\infty} dv_y \int_0^1 d\xi \times \right. \\ \left. f[\mathbf{I}(R, z, v_x, v_y, v_z + \sqrt{2} \delta v_z \text{erf}^{-1}(2\xi - 1); \Phi)] \right\}, \quad (29)$$

where  $\xi \sim \mathcal{U}(0, 1)$  marginalises over velocity error.

We evaluate these multidimensional integrals by using the Monte Carlo approach. For each star with index  $i$ , we construct an array of  $N_{\text{samp}} \gtrsim 10^3$  ‘‘sample points’’ with missing phase-space coordinates  $\mathbf{w}_{i,k}$  assigned from a suitable prior distribution  $P_i(\mathbf{w})$ . The marginalised value of DF for the  $i$ -th star is  $f_i = \sum_{k=1}^{N_{\text{samp}}} f[\mathbf{I}(\mathbf{w}_{i,k}; \Phi)]/P_i(\mathbf{w}_{i,k})$ . Some of these samples would have a zero probability (e.g., if the energy is positive), but as long as there remain samples in a valid region of phase space, the total probability is positive, and hence the log-likelihood is finite.

To eliminate the impact of Poisson noise on the relative odds of different models, we use the same set of sample points for all models (e.g. McMillan & Binney 2013), and to further improve the accuracy, we design the priors according to the importance sampling approach, i.e. we place samples more densely in regions that are expected to have a higher value of  $f$ , thereby approximately equalising their contribution to the total sum. Namely, to sample the missing  $z$  coordinate, we determine a smooth non-parametric estimate of the surface density profile of tracers, deproject it to obtain an estimate of the 3d density  $\rho(r)$ , and sample  $z$  from  $\rho(\sqrt{R_i^2 + z^2})$  at the position of each star  $R_i$ . For the missing velocity components, we estimate the velocity dispersion of the entire system from the measured values  $v_{z,i}$ , and sample  $v_{x,y}$  from a heavy-tailed bell-like distribution with the same dispersion. The non-uniform sampling is accounted for when computing the contribution  $P_i$  of each sample point to the marginalization integral; the estimates of the density pro-

file and velocity dispersion are only used during the initial resampling step, but not in subsequent fitting.

In a spherical system, the integrals of motion are the energy  $E$  and angular momentum  $L$  per unit mass. Alternatively, one may use actions as the integrals of motion; in this case,  $E$  is replaced with the radial action  $J_r$ , defined as:

$$J_r \equiv \frac{1}{\pi} \int_{r_{\text{min}}}^{r_{\text{max}}} v_r dr \\ = \frac{1}{\pi} \int_{r_{\text{min}}}^{r_{\text{max}}} dr \sqrt{2[E - \Phi(r)] - L^2/r^2}, \quad (30)$$

where  $r_{\text{min}}$  and  $r_{\text{max}}$  are the roots of the expression under the radical. For each choice of potential, we pre-compute a high-accuracy 2d interpolating spline for efficient evaluation of  $J_r(E, L)$ , thus using the action variables adds a minor overhead compared to the use of just the classical integrals  $E, L$ . The advantages of action variables are not apparent in the context of spherical models, but are more important in non-spherical cases.

For the density profile of the DM, responsible for the gravitational potential, we adopt the same five-parameter model in equation 3 that was used to produce the mocks. On the other hand, we use two different families of DF: one is expressed in terms of  $E$  and  $L$ , and the other – in terms of radial action  $J_r$  and the sum of two other actions  $L \equiv J_z + |J_\phi|$ ; both have six free parameters.

The first DF family is constructed from a given tracer density profile, using the Cuddeford–Osipkov–Merritt inversion formula (Cuddeford 1991). In this case, we use Equation 3 for the tracer density profile, which has 4 free parameters:  $\alpha_*$ ,  $\beta_*$ ,  $\gamma_*$  and  $r_*$  (the fifth one,  $\nu_0$ , is fixed by the normalization constraint). Two other parameters  $\beta_0$ ,  $r_a$  define the anisotropy profile (Equation 4 with  $n = 2$  and  $\beta_\infty = 1$ ). This DF family includes the true DF of all variants of models, so we expect to be able to recover the true parameters, given enough data.

The second family is a generalisation of the action-based DF presented in Posti et al. (2015):

$$f(J_r, L) = A \frac{\left(1 + [J_0/\{pJ_r + (1-p)L\}]^\eta\right)^{\Gamma/\eta}}{\left(1 + [\{qJ_r + (1-q)L\}/J_0]^\eta\right)^{(B-\Gamma)/\eta}}. \quad (31)$$

Here the power-law indices  $\Gamma$  and  $B$  control the steepness of the density profile at small and large radii, correspondingly;  $J_0$  determines the transition between the two asymptotic regimes,  $\eta$  controls the sharpness of this transition, the mixing parameters  $p$  and  $q$  are responsible for the velocity anisotropy at small and large radii, and finally  $A$  is the overall normalisation constant, which ensures that the integral of  $f$  over the entire phase space is unity (it is computed numerically for each choice of six DF parameters and is not a free parameter by itself). In this case, there are no separate parameters for the density profile, because it is determined from the DF and the potential:  $\nu(r) = \iiint d^3\mathbf{v} f[\mathbf{I}(\mathbf{x}, \mathbf{v}; \Phi)]$ . However, we do not need the density *per se*, since we fit the joint DF of position and velocity. Since the true DF used to generate the mock data does not belong to this family of models, we may expect that best-fit parameters of the potential could be biased by the constrained form of the adopted DF.

To explore the parameter space, we use the EMCEE code

with 32 walkers evolved for several thousand steps. The evaluation of model likelihoods is performed using AGAMA<sup>6</sup>.

### 3.5 Summary of methods and assumptions

Table 2 summarises the different assumptions and parametrisations of the different methods.

## 4 RESULTS

In this section, we present results for each method described in §3 for the recovery of the total density and stellar velocity anisotropy at  $R = [0.25, 0.5, 1, 2, 4] R_{1/2}$ , where  $R_{1/2}$  is the projected half light radius, using only line-of-sight velocity data for 1,000 and 10,000 tracers. For brevity, we focus on the ‘Plum’ mocks that are representative of the results from the full suite; for completeness, we present results for the ‘NonPlum’ mocks in Appendix A. For all mocks, we applied  $2 \text{ km s}^{-1}$  velocity errors to be consistent with current spectroscopic accuracy for dwarf galaxy data (see §1).

### 4.1 Isotropic mocks

Figures 1 and 2 show the recovery of the density (top) and *symmetrised velocity anisotropy* (bottom; see equation 8) for the isotropic mocks for 1,000 (left) and 10,000 (right) tracers. For these isotropic mocks, all four methods perform similarly well, with the results improving with increased sampling. The density profile is already well-recovered with 1,000 tracers, with all methods recovering the input solution within their 95% confidence intervals over the radial range  $0.25 < R/R_{1/2} < 4$ .

However, the uncertainty on  $\tilde{\beta}$  remains substantial for 1,000 tracers. Increasing their number to 10,000, AGAMA and GRAVSPHERE obtain the tightest constraints on  $\tilde{\beta}$ , with AGAMA – that uses the full shape information in the distribution function – having the smallest uncertainties.

### 4.2 Radially anisotropic mocks

Moving to radially anisotropic mocks (Figures 3 and 4), we now see a larger difference between the different methods. Each method recovers some parts of the solution better than others, highlighting the value of modelling the data with multiple techniques. Overall, the accuracy of recovery of the potential is tightly linked to the ability of the method to recover the anisotropy profile of the tracers (c.f. Read & Steger 2017).

For 1,000 tracers, MAMPOSST and AGAMA  $f(E, L)$  recover  $\tilde{\beta}(r)$  and  $\rho(r)$  over the radial range  $0.25 < R/R_{1/2} < 4$  within their 95% confidence intervals, for both the PlumCoreOM and PlumCuspOM mocks. This may reflect the fact that in these methods, the assumed functional forms for the mass, light and  $\tilde{\beta}$  profiles encompass the true solution (in particular,  $\beta_\infty = 1$  in the AGAMA  $f(E, L)$  models). By contrast, the AGAMA  $f(\mathbf{J})$  models do not fully recover

the  $\tilde{\beta}$  profile, being more radially anisotropic at small  $r$  and less radially anisotropic at large  $r$ . This negatively impacts the accuracy of potential recovery in the PlumCuspOM case, although not in the PlumCoreOM case. GRAVSPHERE recovers the anisotropy profile within its 95% confidence intervals in the PlumCoreOM case, but performs more poorly in the PlumCuspOM case. DISCRETEJAM and GRAVSPHERE are both biased to low  $\rho(r)$  beyond  $R_{1/2}$ , while DISCRETEJAM is also biased to high  $\rho(r)$  inside  $R_{1/2}$  for the PlumCoreOM mock. This owes to a bias towards tangential anisotropy for the DISCRETEJAM method in both PlumCoreOM and PlumCuspOM mocks that does not diminish even for 10,000 tracer stars. This behaviour is also seen in the GRAVSPHERE code if the VSPs are not used in the fit. Read & Steger (2017) show that this bias owes to the mass-anisotropy degeneracy, combined with the true solution for this mock lying on the edge of the hypervolume of acceptable models. Indeed, they show that the correct model *is* recovered within the full MCMC chains, but is rare as compared to the large number of tangentially anisotropic models that fit the data similarly well.

Finally, moving to 10,000 tracers, bias starts to creep in or become more statistically significant in most of the methods. The AGAMA  $f(E, L)$  models still produce good fits within their 95% confidence intervals, while the  $f(\mathbf{J})$  models retain the biases in the anisotropy profile in both cases, and in  $\rho(r)$  in the PlumCuspOM case, reflecting the fact that the assumed distribution function differs from the one used to set up the mocks (c.f. the earlier discussion on this in §1). MAMPOSST also starts to show similarly biased behaviour, with the inner density underestimated at small radii for the PlumCuspOM and PlumCoreOM mocks, and bias now present in the recovery of  $\tilde{\beta}(r)$ . This likely owes to the assumed Gaussian form of the local distribution function being a good, but not perfect, approximation to its true shape. The DISCRETEJAM method – that is not able to break the mass-anisotropy degeneracy with line-of-sight data alone – retains the bias seen in its recovery of the 1,000 star mocks, but with smaller error bars. The GRAVSPHERE method – that is designed to make as few assumptions as possible about the form of the gravitational potential and the distribution function – approaches the true solution more closely, although some bias towards lower  $\rho(r)$  remains at large radii for the PlumCuspOM mock. This bias is, however, significantly reduced as compared to the 1,000 star solution.

Finally, we note that some of the differences in performance between the different methods owes also to their different priors (see Table 2). GRAVSPHERE has the loosest priors of all methods presented here and, correspondingly, the greatest challenge in recovering the mocks. By contrast, MAMPOSST assumes tighter priors that are symmetrised around the true solution, DISCRETEJAM uses priors for the anisotropy and DM scale radii that are tightly linked to the stellar density scale radius (which is itself very well recovered), and AGAMA  $f(E, L)$  assumes  $\beta_\infty = 1$ , as in the mocks. Moreover, DISCRETEJAM, MAMPOSST, and AGAMA each assume the same functional form for the mass profile as in the mock data (with MAMPOSST fixing the value of  $\alpha_D$ ), whereas GRAVSPHERE does not. As such, we cannot really speak of any one method being ‘superior’ to the others; rather, each has its strengths and weaknesses.

<sup>6</sup> The AGAMA library, together with the complete Python program performing the DF fitting for this study, is available at <http://agama.software>

**Table 2.** Assumptions and parameters of different methods

Method	GRAVSPHERE	DISCRETEJAM	MAMPOSSST	AGAMA	
Basis	Binned $v_{\text{LOS}}$ + VSPs	Gaussian $v_{\text{LOS}}$	Gaussian $v_{3\text{D}}$	$f(E, L)$	$f(J_r, L)$
Radial binning	yes	no	no	no	no
$\rho_{\text{D}}(r)$	5-piece power law	MGE fit to eq. (3)	eq. (3), $\alpha = 1$	eq. (3)	eq. (3)
$\nu_*(r)$	sum of Plummer models	MGE fit to eq. (3)	eq. (3), $\alpha = 2, \beta = 5$	eq. (3)	fitted by DF
$\beta_*(r)$	eq. (4)	eq. (4), $n = 2$	eq. (4), $n=2$	eq. (4), $n = 2, \beta_\infty = 1$	fitted by DF
# of free parameters	17	12	9	11	11
Ranges of free parameters:					
$\log_{10}(M_{200c}/M_\odot)$	unbounded	unbounded	true $\pm 1$		unbounded
$r_{\text{D}}/\text{kpc}$	–	$0.25 r_* \leftrightarrow 50 r_*$	true $\pm 0.5$ dex		$0.1 \leftrightarrow 100$
$\alpha_{\text{D}}$	–	$0.5 \leftrightarrow 2.5$	1		$0.5 \leftrightarrow 2.5$
$\beta_{\text{D}}$	–	$2 \leftrightarrow 5$	$2 \leftrightarrow 6$		$2.2 \leftrightarrow 6$
$\gamma_{\text{D}}$	–	$-0.5 \leftrightarrow \beta_{\text{D}}$	$-0.1 \leftrightarrow 2$		$-0.5 \leftrightarrow 1.9$
$\log_{10}(r_*/\text{kpc})$	–	$-2 \leftrightarrow 1$	true $\pm 0.3$	$-2 \leftrightarrow 1$	–
$\alpha_*$	–	$0.5 \leftrightarrow 2.5$	2	$0.5 \leftrightarrow 3$	–
$\beta_*$	–	$2 \leftrightarrow 7$	5	$3.5 \leftrightarrow 6$	–
$\gamma_*$	–	$-0.5 \leftrightarrow \beta_*$	$-0.05 \leftrightarrow 1.5$	$0 \leftrightarrow 1.9$	–
$\tilde{\beta}_0$	$-1 \leftrightarrow 1$	$-0.95 \leftrightarrow 1$	$-0.9 \leftrightarrow 1$	$-0.5 \leftrightarrow 1$	–
$\tilde{\beta}_\infty$	$-1 \leftrightarrow 1$	$-0.95 \leftrightarrow 1$	$-0.9 \leftrightarrow 1$	1	–
$\log_{10}(r_a/\text{kpc})$	$-2 \leftrightarrow 0$	$\log_{10}(r_*) \pm 0.3$	true $\pm 0.3$	$-2 \leftrightarrow 5$	–
$n$	$1 \leftrightarrow 10$	2	2	2	–
$\Gamma$	–	–	–	–	$0 \leftrightarrow 2.8$
B	–	–	–	–	$3.2 \leftrightarrow 12$
Minimization	EMCEE	EMCEE	COSMOMC		EMCEE
# of chains	1,000	100	6		32
Length of chains	5,000	20,000	90,000		few $\times 10^3$
CPU minutes ( $10^3$ tracers)	360	$\mathcal{O}(10^4)^\dagger$	$6 \times 15$ (the 6 in   )		$\mathcal{O}(10^3)$

<sup>†</sup>The run time speeds up if fewer Gaussians are used; a non-parametric approach would use many fewer Gaussians and could be significantly faster.

### 4.3 Tangentially anisotropic mocks

In Figure 5, we show our results for the tangentially anisotropic mocks, PlumCoreTan and PlumCuspTan assuming 1000 tracer stars. Notice that, similarly to the isotropic mocks, all methods recover  $\rho(r)$  within their 95% confidence intervals over the range  $0.25 < R/R_{1/2} < 4$  (top panels), though for the PlumCoreTan mock, DISCRETEJAM is biased (at 68% confidence) towards cuspy, tangentially anisotropic mocks and MAMPOSSST marginally so. This is similar, though less severe, to the bias seen for the PlumCoreOm mock and owes to the mass-anisotropy degeneracy (see discussion above).

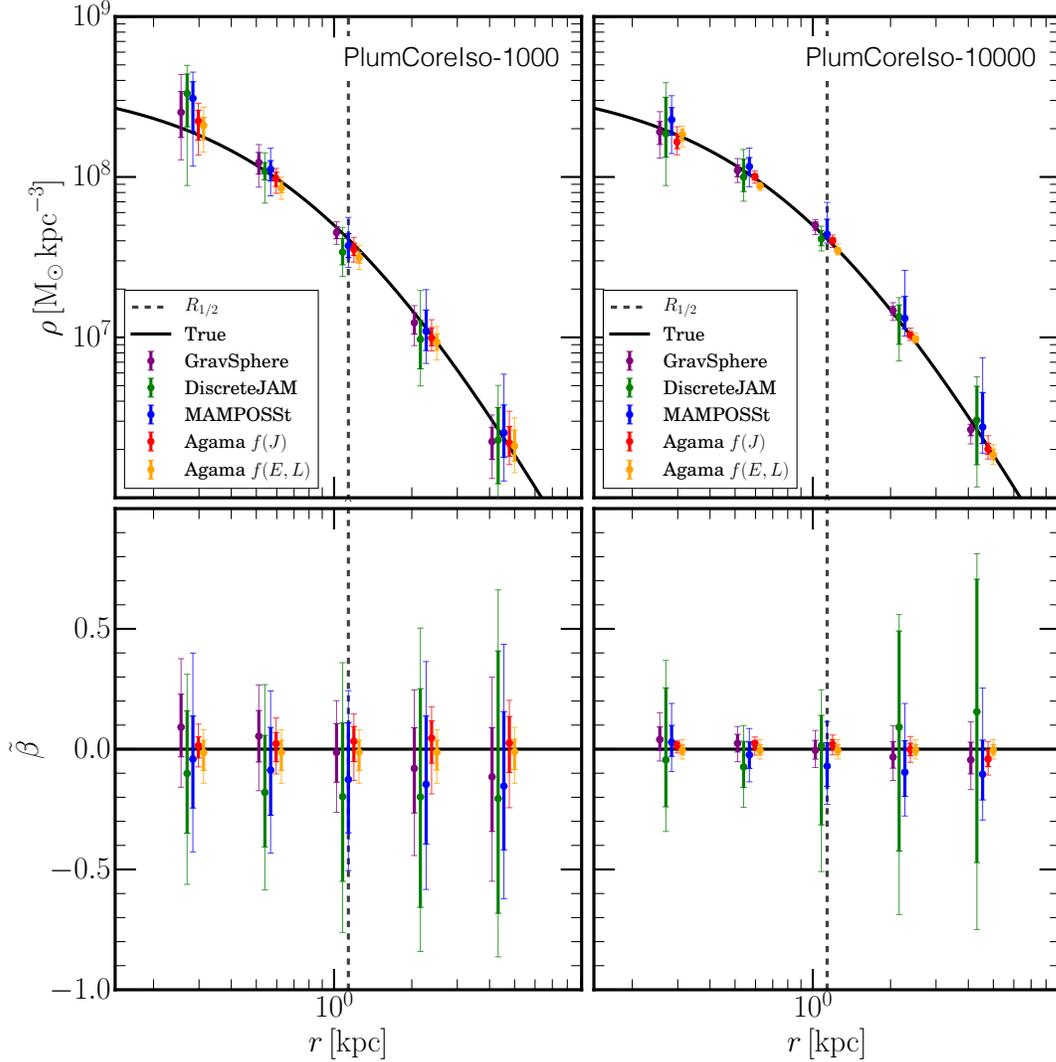
All methods successfully detect the tangential anisotropy, though for 1,000 tracers, this is only statistically significant near the half-light radius (vertical dashed line). Interestingly, both the AGAMA  $f(E, L)$  and the AGAMA  $f(\mathbf{J})$  methods become biased at greater than their 95% confidence intervals. AGAMA  $f(\mathbf{J})$  is overly isotropic in the inner regions and, for PlumCuspTan, overly tangential in its outer regions, while AGAMA  $f(E, L)$  is overly isotropic at all radii. These biases reflect the assumed parameterisation of the distribution function and the choice of priors. As discussed above and in §1, this will lead to bias if the phase space distribution function of the mock data is inconsistent with these assumptions, as is the case here. Indeed, the tangentially anisotropic mocks present a particular challenge for the AGAMA  $f(E, L)$  models, since the true value of  $\beta = -0.5$  lies at the boundary of the allowed range (for technical reasons, anisotropic DFs computed by the Cuddeford–Osipkov–Merritt method are restricted to have  $\tilde{\beta}_0 \geq -0.5$ ).

### 4.4 Adding proper motion data

In Figure 6, we show the performance of all methods when adding proper motion data. For this, we consider 1,000 tracers for the PlumCuspOM and PlumCoreOM mocks assuming  $2 \text{ km s}^{-1}$  errors on all three components of the velocity (the other mock data produced comparable results). All three Jeans models constrain the line of sight, radial and tangential dispersion profiles as in Read & Steger (2017), while the AGAMA DF method uses the 5D phase-space data. As can be seen in Figure 6, most methods now recover both  $\rho(r)$  and  $\beta(r)$  within their quoted 95% confidence intervals, although some biases still remain. GRAVSPHERE underestimates  $\rho(r)$  at large radii, while DISCRETEJAM overestimates  $\rho(r)$  at small radii for the PlumCoreOM mock. AGAMA  $f(\mathbf{J})$  and GRAVSPHERE produce less radially anisotropic  $\tilde{\beta}$  profiles at large radii for the PlumCuspOM mock, and AGAMA  $f(\mathbf{J})$  models are slightly radially biased in the centre. MAMPOSSST obtains a good recovery of  $\rho(r)$  and the correct shape for  $\tilde{\beta}(r)$ , but has  $\tilde{\beta}(r)$  biased low at small radii in PlumCuspOM. Nonetheless, with data of this quality all methods are able to distinguish the cusped and cored mocks at high confidence.

## 5 DISCUSSION & CONCLUSIONS

We have shown that all four mass modelling methods considered in this paper, using just line-of-sight velocity data, are able to recover both the density and velocity anisotropy as a function of radius within their 95% confidence intervals over the radial range  $0.25 < R/R_{1/2} < 4$ , for spher-

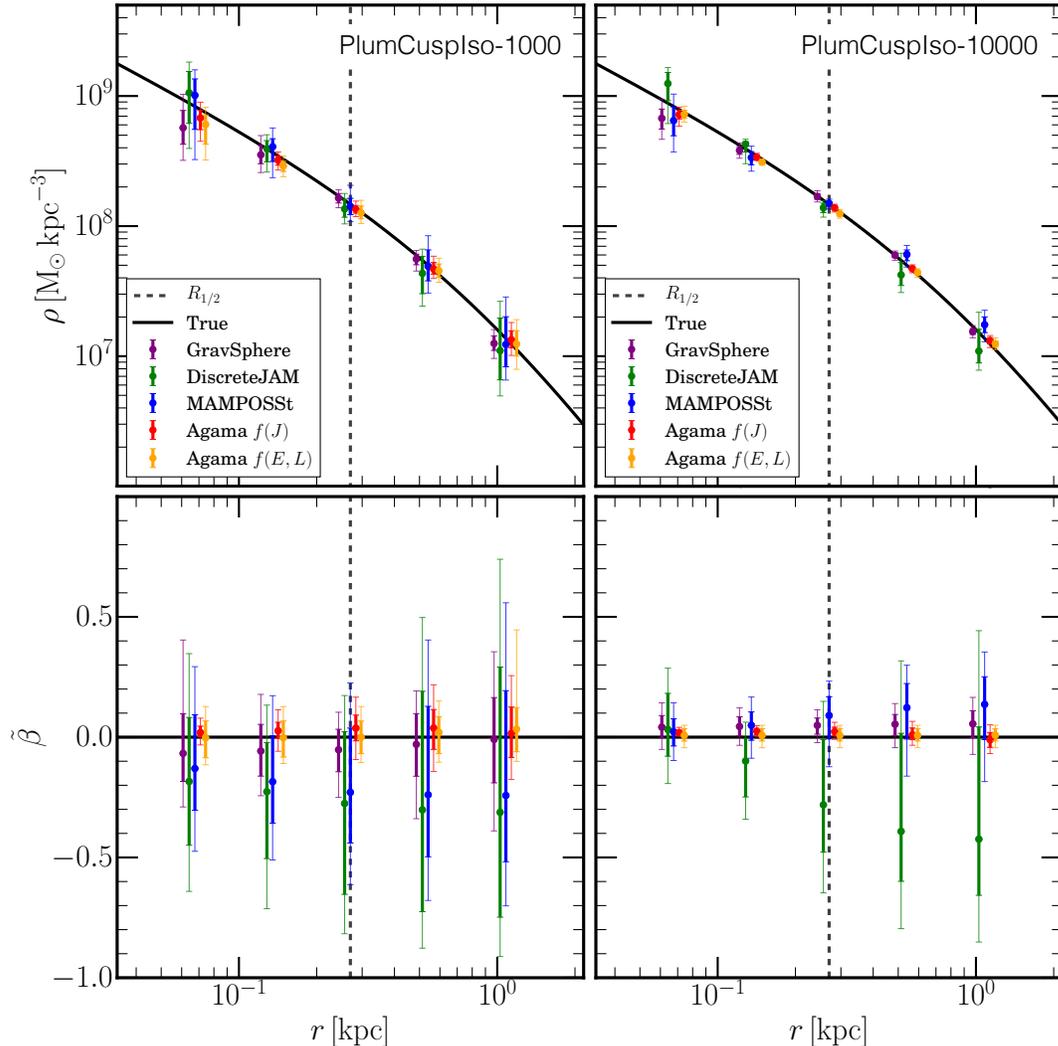


**Figure 1.** Recovery of the density (top) and symmetrised velocity anisotropy (bottom; see equation 8) for the PlumCoreIso mocks, with line-of-sight velocities for 1,000 (left) and 10,000 (right) tracer stars. The data points show the recovery at  $R = [0.25, 0.5, 1, 2, 4] R_{1/2}$ , where  $R_{1/2}$  is the projected half light radius (vertical dashed line). The thick/thin error bars mark the 68%/95% confidence intervals, respectively. The different colours show results for GRAVSPHERE (purple), DISCRETEJAM (green), MAMPOSSt (blue), AGAMA  $f(\mathbf{J})$  (red) and  $f(E, L)$  (orange), as marked in the legend. Each of the methods is slightly offset left-right from one another to aid clarity (where points are offset, we show the density profile recovery at exactly this offset point). The true mock density and velocity anisotropy profiles are shown by the black lines.

ically symmetric mock stellar systems, provided that the mocks are isotropic or tangentially anisotropic. However, strong radial anisotropy at large radii presents a more challenging test. Only methods that utilise some information about the shape of the distribution function are able to recover the density and velocity anisotropy profiles of these mocks. GRAVSPHERE achieves this by using the two fourth order ‘Virial Shape Parameters’ (Merrifield & Kent 1990); AGAMA achieves it by directly fitting a distribution function; and MAMPOSSt achieves it by projecting an assumed local Gaussian velocity distribution function along the line of sight. To illustrate this, in Figure 7 we show the radial profiles of line-of-sight velocity dispersion and kurtosis  $\kappa$ , as well as the line-of-sight velocity distribution (LOSVD) in the outer parts, for the PlumCoreOM model. Notice that the

profiles are substantially non-Gaussian (fat-tailed) across the entire range of radii, indicated by an elevated value of  $\kappa$ . The DF-based AGAMA models are able to capture this behaviour, while the projected Jeans equations used in the DISCRETEJAM method, which uses only the second moment of the distribution function, assuming a Gaussian LOSVD, become biased. Also, MAMPOSSt predicts too low and high values of kurtosis at small and large radii, respectively, explaining its respective underestimate and overestimate of  $\beta$  at small and large radii (right panels of Fig. 3).

Adding internal proper motion data for 1,000 tracer stars, most methods recovered both  $\rho(r)$  and  $\beta(r)$  within their 95% confidence intervals over the radial range  $0.25 < R/R_{1/2} < 4$ , albeit with some small biases at the largest and smallest radii. (GRAVSPHERE, in particular, underesti-



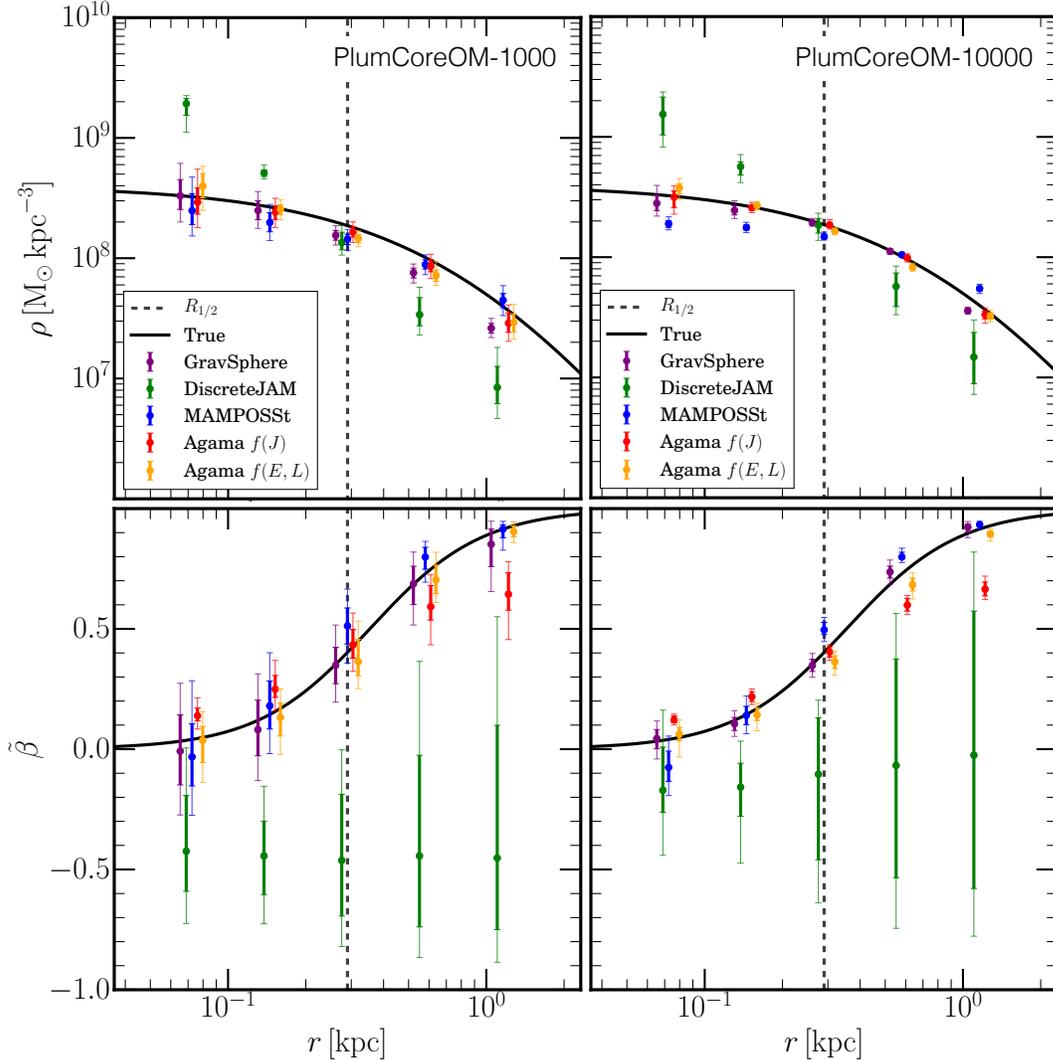
**Figure 2.** As Figure 1, but for the PlumCuspIso mock.

mated the density for  $R > 2R_{1/2}$  on these tests.) These findings are in good agreement with previous work (e.g. [Strigari et al. 2007](#); [Read & Steger 2017](#)), demonstrating the value of obtaining proper motion data for significant numbers of individual stars in nearby dwarf galaxies. This is just beginning to become possible now with the advent of combined *Hubble Space Telescope* and *Gaia* proper motions ([Massari et al. 2018, 2020](#); [Vitril & Mamon 2020](#)), but really large datasets will have to wait for the upcoming *Nancy Grace Roman Space Telescope* ([Sanderson et al. 2017](#)) and proposed space astrometry missions like *Theia* ([Theia Collaboration et al. 2017](#)).

We conclude that breaking the mass-anisotropy degeneracy is crucial for obtaining an unbiased measurement of the density profile and velocity anisotropy of spherical stellar systems. Previous work has focused on achieving this using multiple tracer populations with different scale lengths (e.g. [Battaglia et al. 2008](#); [Walker & Peñarrubia 2011](#); [Amorisco & Evans 2012](#); [Read & Steger 2017](#)). While this has been shown to work well, it can only be applied to systems that have distinct populations. In this paper, we have focused

instead on modelling single tracer population data. In this case, with only line-of-sight velocity data, we have shown that the degeneracy can be broken by using higher order ‘Virial Shape Parameters’, by directly fitting a global distribution function, or by self-consistently projecting an assumed local velocity distribution function along the line of sight. We have also uncovered interesting differences even between rather similar-looking methodologies, highlighting the utility of performing detailed tests on mock data and exploring different mass modelling approaches, as we have done here.

Finally, we note that all of the tests in this paper assumed perfect data with small (2 km/s) velocity errors. Already this was challenging for some methods. Real stellar systems will have additional uncertainties from binary star and foreground contamination (e.g. [Walker & Peñarrubia 2011](#); [Read et al. 2018](#)), disequilibrium due to tides (e.g. [Ural et al. 2015](#); [Kowalczyk et al. 2013](#); [Read et al. 2018](#); [Genina et al. 2020](#)), rotation (e.g. [Watkins et al. 2013](#); [Zhu et al. 2016](#)), departures from spherical symmetry (e.g. [Mamon et al. 2013](#); [Kowalczyk et al. 2013](#); [Read & Steger 2017](#); [Gen-](#)



**Figure 3.** As Figure 1, but for the PlumCoreOM mock.

ina et al. 2020; Hayashi et al. 2020), and inconsistency between the tracer density and kinematic measurements (e.g. Read 2014). Some of these problems can be ameliorated with better data, for example, repeat measurements of individual stars to explore binary contamination (e.g. Koposov et al. 2011; Spencer et al. 2018), metallicity measurements to improve foreground separation, and having a sufficient number of unbiased stars with kinematics that the photometric and kinematic samples are identical. However, some problems like tides will remain, even with exquisite data (e.g. Kowalczyk et al. 2013; Ural et al. 2015). We will explore these issues further in future work.

## 6 ACKNOWLEDGEMENTS

We would like to thank the anonymous referee for a constructive and thorough report that improved the paper. JIR would like to acknowledge support from STFC consolidated grant ST/M000990/1 and the MERAC foundation. GAM thanks the French Programme National de Cos-

mologie et Galaxies (PNCG) of CNRS/INSU with INP and IN2P3, co-funded by CEA and CNES, for financial support of a workshop at IAP during which this work was finalized. EV and PD acknowledge support from the European Research Council (ERC) Horizon 2020 programme under grants 321067 (EV and PD) and 308024 (EV). LLW acknowledges support provided by NASA through *HST* grants AR-14322 and AR-15055, from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555, and from the ERC Horizon 2020 programme under grant agreement No 724857 (Consolidator Grant ArcheoDyn).

This research has made use of Astropy<sup>7</sup>, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013), PYNBODY<sup>8</sup> for the simulation

<sup>7</sup> <http://www.astropy.org>

<sup>8</sup> <https://github.com/pynbody/pynbody>

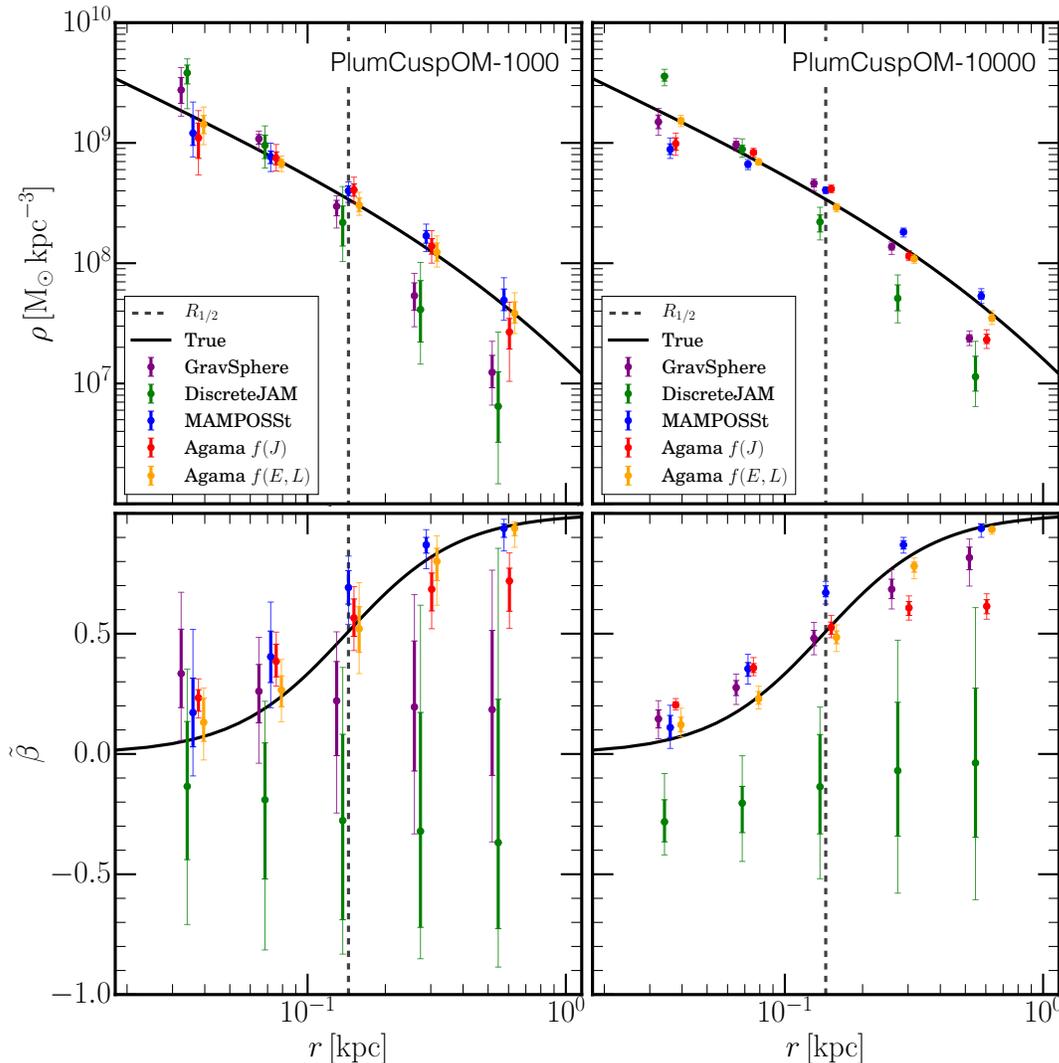


Figure 4. As Figure 1, but for the PlumCuspOM mock.

analysis (Pontzen et al. 2013), EMCEE<sup>9</sup> (Foreman-Mackey et al. 2013), and CosmoMC<sup>10</sup> (Lewis & Bridle 2002) for MCMC sampling, as well as NASA’s Astrophysics Data System.

## APPENDIX A: RESULTS FOR THE REMAINDER OF THE SPHERICAL MOCKS

For completeness, in this Appendix we show the spherically averaged density profile for the remainder of the spherical mocks with 1000 tracer stars, the ‘NonPlum’ models (Figure A1).

As can be seen, the results are broadly consistent with those for the similar ‘Plum’ mocks, with DISCRETEJAM showing a bias towards cuspy, tangentially anisotropic models, particularly for NonPlumCoreOm, while MAMPOSST

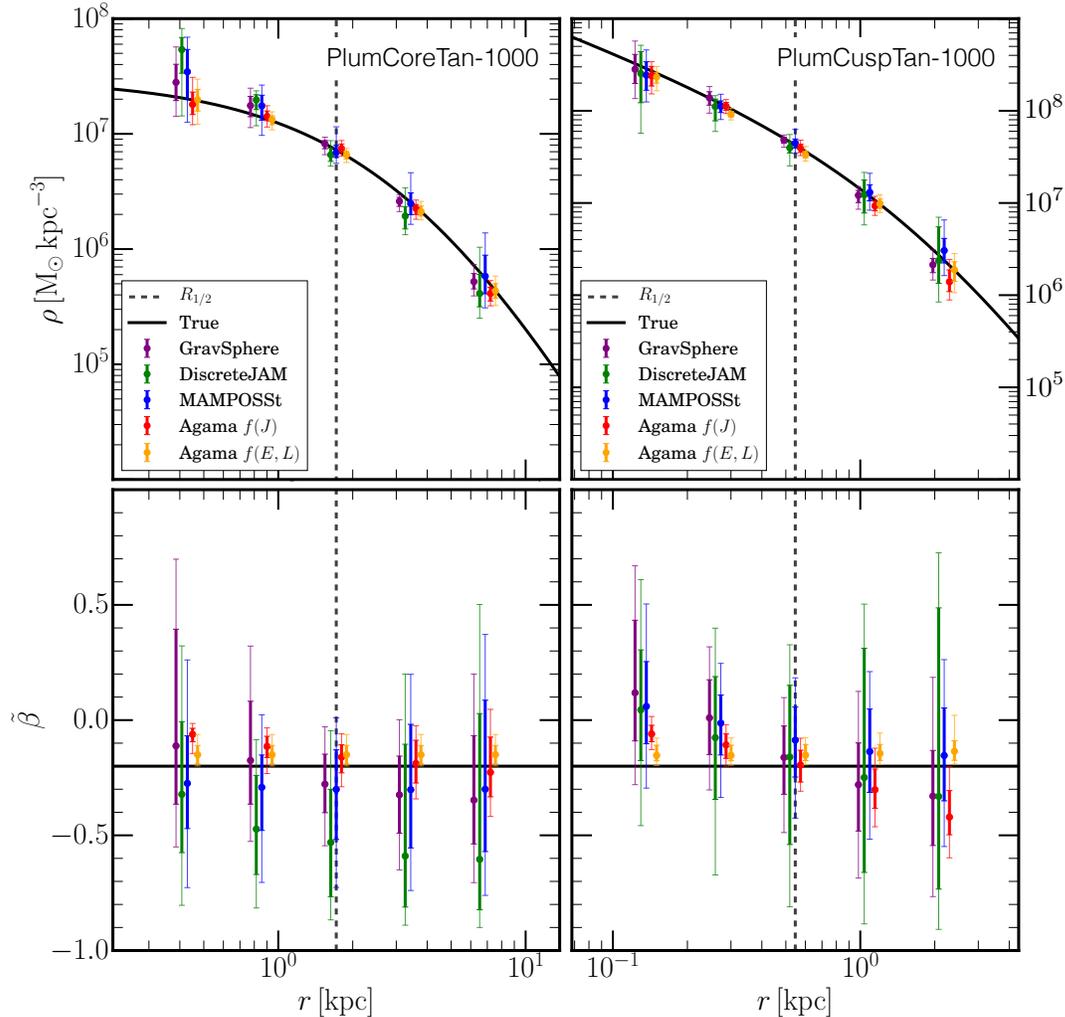
shows an even stronger cuspy bias for NonPlumCoreIso. As discussed in §4 and §5, this owes to the mass-anisotropy degeneracy.

Similarly to the PlumCuspOm mock, GRAVSPHERE and DISCRETEJAM fall off too steeply at large radii for NonPlumCuspOm, however MAMPOSST and the AGAMA models also show a similar, albeit less severe, bias.

Interestingly, all of the models except DISCRETEJAM show a slight bias towards cores in the innermost region ( $0.25 R_{1/2}$ ) for the NonPlumCuspIso mock. DISCRETEJAM does not show this bias, but this is because it shows a bias instead towards tangentially anisotropic models. (Tangentially anisotropic models require a steeper density profile to match the data that, in this case, drives a better match with the input model.) MAMPOSST (and DISCRETEJAM to a lesser extent) finds instead an even steeper inner density profile than the NFW model used to build this mock. By contrast, for the NonPlumCuspIso mock, both GRAVSPHERE and AGAMA are consistent with the input isotropic distribution function within their 95% confidence intervals, while MAMPOSST produces tangential orbits at all radii

<sup>9</sup> <https://github.com/dfm/emcee>

<sup>10</sup> <https://cosmologist.info/cosmomc/>

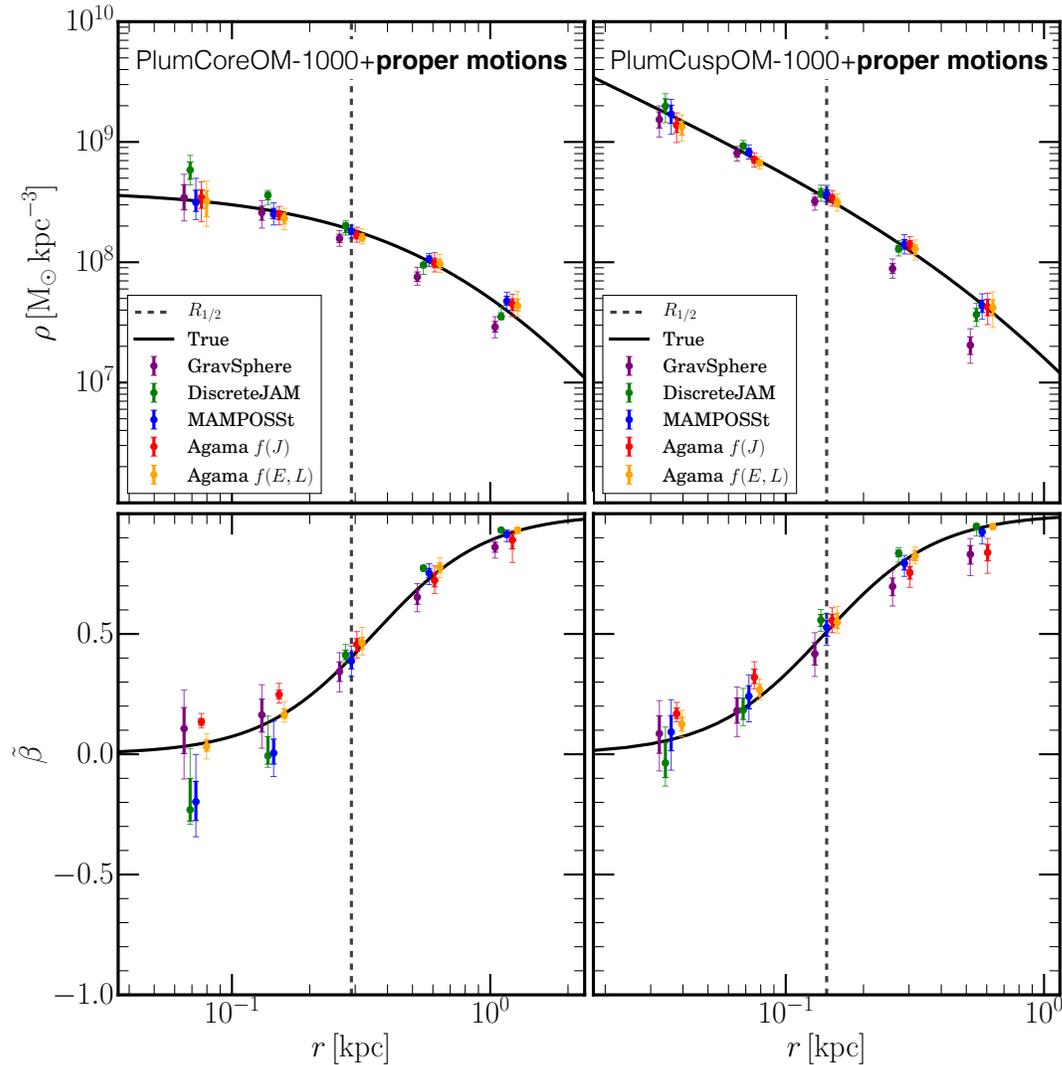


**Figure 5.** As Figure 1, but for mocks PlumCoreTan (left) and PlumCuspTan (right) with 1000 tracer stars.

(not shown for brevity). As such, the bias towards cores in the NonPlumCuspIso mock does not owe to the mass-anisotropy degeneracy. We speculate that it owes instead to the cuspy light profile in this mock yielding a poorer radial sampling of the stellar kinematics. (For a fixed number of tracers, cuspier light profiles will have relatively fewer sample points at large radii.) This warrants further investigation in future work. However, we note that it is unlikely to impact most real stellar systems, since these typically have a much shallower inner tracer density profile than the rather extreme ‘NonPlum’ mocks explored here (see e.g. [Read et al. 2019](#), for nearby dwarf galaxies).

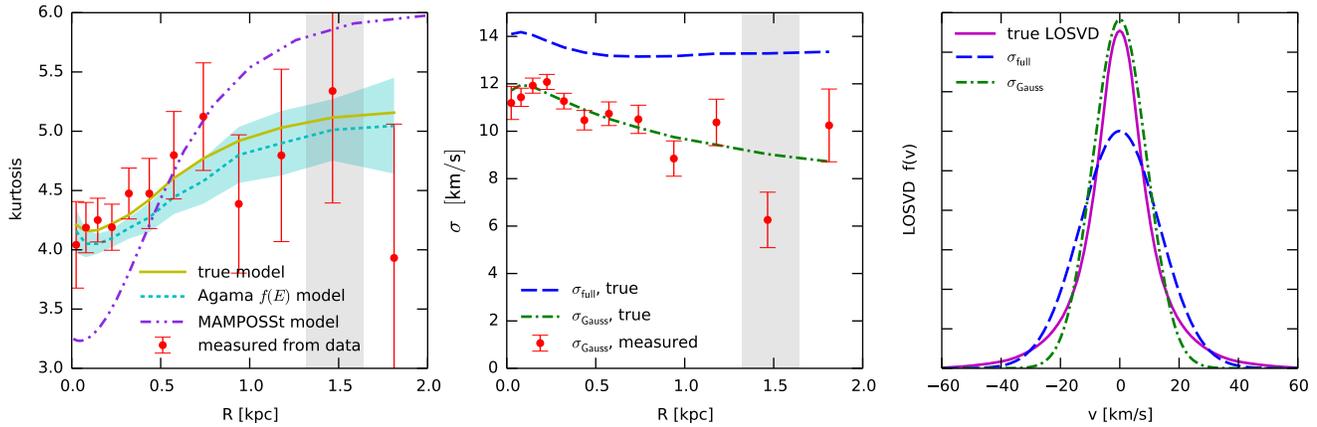
## REFERENCES

- Amorisco N. C., Evans N. W., 2012, *MNRAS*, 419, 184  
 An J. H., Evans N. W., 2006, *ApJ*, 642, 752  
 Astropy Collaboration et al., 2013, *A&A*, 558, A33  
 Battaglia G., Helmi A., Tolstoy E., Irwin M., Hill V., Jablonka P., 2008, *ApJ*, 681, L13  
 Bellini A. et al., 2014, *ApJ*, 797, 115  
 Binney J., Mamon G. A., 1982, *MNRAS*, 200, 361  
 Binney J., Tremaine S., 2008, *Galactic Dynamics: Second Edition*. Princeton Univ. Press  
 Bonnavard V. et al., 2015, *MNRAS*, 453, 849  
 Cappellari M., 2002, *MNRAS*, 333, 400  
 Cappellari M., 2008, *MNRAS*, 390, 71  
 Cappellari M., 2015, *ArXiv:1504.05533*  
 Cappellari M. et al., 2012, *Nature*, 484, 485  
 Chiappo A., Cohen-Tanugi J., Conrad J., Strigari L. E., Anderson B., Sánchez-Conde M. A., 2017, *MNRAS*, 466, 669  
 Cuddeford P., 1991, *MNRAS*, 253, 414  
 Dehnen W., 2009, *MNRAS*, 395, 1079  
 Diakogiannis F. I., Lewis G. F., Ibata R. A., Guglielmo M., Wilkinson M. I., Power C., 2019, *MNRAS*, 482, 3356  
 Eggen O. J., Lynden-Bell D., Sandage A. R., 1962, *ApJ*, 136, 748  
 Emsellem E., Monnet G., Bacon R., 1994, *A&A*, 285, 723  
 Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306  
 Garbari S., Read J. I., Lake G., 2011, *MNRAS*, 416, 2318  
 Genina A., Fairbairn M., 2016, *MNRAS*, 463, 3630



**Figure 6.** As Figure 1, but for mocks PlumCoreOM (left) and PlumCuspOM (right) including also proper motion data (see text for details).

- Genina A. et al., 2020, MNRAS  
 Geringer-Sameth A., Koushiappas S. M., Walker M. G., 2015, Phys. Rev. D, 91, 083535  
 Gregory A. L., Collins M. L. M., Read J. I., Irwin M. J., Ibata R. A., Martin N. F., McConnachie A. W., Weisz D. R., 2019, MNRAS, 485, 2010  
 Hayashi K., Chiba M., Ishiyama T., 2020, arXiv e-prints, arXiv:2007.13780  
 Hénault-Brunet V., Gieles M., Sollima A., Watkins L. L., Zocchi A., Claydon I., Pancino E., Baumgardt H., 2019, MNRAS, 483, 1400  
 Hubble E. P., 1926, ApJ, 64, 321  
 Jeans J. H., 1922, MNRAS, 82, 122  
 Kamann S. et al., 2018, MNRAS, 473, 5591  
 Koposov S. E. et al., 2011, ApJ, 736, 146  
 Kowalczyk K., Lokas E. L., Kazantzidis S., Mayer L., 2013, MNRAS, 431, 2796  
 Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511  
 Limousin M., Morand i A., Sereno M., Meneghetti M., Ectori S., Bartelmann M., Verdugo T., 2013, Space Sci. Rev., 177, 155  
 Mamon G. A., Biviano A., Boué G., 2013, MNRAS, 429, 3079  
 Mamon G. A., Cava A., Biviano A., Moretti A., Poggianti B., Bettoni D., 2019, A&A, 631, A131  
 Mamon G. A., Lokas E. L., 2005, MNRAS, 363, 705  
 Massari D., Breddels M. A., Helmi A., Posti L., Brown A. G. A., Tolstoy E., 2018, Nature Astronomy, 2, 156  
 Massari D., Helmi A., Mucciarelli A., Sales L. V., Spina L., Tolstoy E., 2020, A&A, 633, A36  
 McConnachie A. W., Côté P., 2010, ApJ, 722, L209  
 McMillan P. J., Binney J. J., 2013, MNRAS, 433, 1411  
 Merrifield M. R., Kent S. M., 1990, AJ, 99, 1548  
 Merritt D., 1985, AJ, 90, 1027  
 Merritt D., 1987, ApJ, 313, 121  
 Napolitano N. R., Pota V., Romanowsky A. J., Forbes D. A., Brodie J. P., Foster C., 2014, MNRAS, 439, 659  
 Noyola E., Gebhardt K., Bergmann M., 2008, ApJ, 676, 1008  
 Osipkov L. P., 1979, Pisma Astronomicheskii Zhurnal, 5,



**Figure 7.** Illustration of non-Gaussianity of LOSVD in a radially-anisotropic model (PlumCoreOM).

*Left panel:* radial profile of the line-of-sight kurtosis  $\kappa \equiv \overline{v^4}/\sigma_{\text{full}}^4$ , where  $\sigma_{\text{full}}^2 \equiv \overline{v^2}$  is the line-of-sight velocity dispersion (full second moment of the velocity distribution). For a Gaussian LOSVD,  $\kappa = 3$ ; a higher value indicates a fat-tailed distribution. Solid yellow line is the true  $\kappa$  in the given model; red points with error bars are the binned measured values in the  $10^4$ -star mock dataset; dotted cyan line with a shaded region is the mean and the 68% interval of values in the AGAMA  $f(E, L)$  models; dash-double-dotted purple line is the MAMPOSS model prediction (neglecting velocity errors). It is clear that the LOSVD is significantly non-Gaussian with fat tails across the entire range of radii. The DF-based models are able to capture this behaviour, while the MAMPOSS method reproduces the trend qualitatively but not in detail, underlining the limitations of using the Gaussian assumption even for the local (not projected) velocity distribution. Gray-shaded region indicates the range of radii used for the LOSVD in the rightmost panel.

*Centre panel:* radial profile of the full line-of-sight velocity dispersion  $\sigma_{\text{full}}$  (dashed blue line) and the width of the best-fit Gaussian  $\sigma_{\text{Gauss}}$  (dot-dashed green line). The difference between the two curves indicates the non-Gaussian shape of the LOSVD: the width of the best-fit Gaussian is comparable to the width of the main peak of the LOSVD, but is significantly smaller than the true second moment of the LOSVD, which is heavily influenced by its tails. Red points with error bars are the binned measured values of the Gaussian approximation to the LOSVD in the  $10^4$ -star mock dataset.

*Right panel:* LOSVD in the penultimate bin (the range of radii indicated by the shaded gray region in the other panels). Solid magenta curve is the true LOSVD; dashed blue curve is a Gaussian profile with the width given by the full second moment of the LOSVD  $\sigma_{\text{full}}$  (the quantity that enters the Jeans equations); dot-dashed green curve is the best-fit Gaussian with width  $\sigma_{\text{Gauss}}$ . Neither is a good approximation to the actual LOSVD, which has prominent fat tails.

77, english translation in Soviet Astron. Lett. 5, 42

Plummer H. C., 1911, MNRAS, 71, 460

Pontzen A., Roškar R., Stinson G., Woods R., 2013, pynbody: N-Body/SPH analysis for python. Astrophysics Source Code Library

Posti L., Binney J., Nipoti C., Ciotti L., 2015, MNRAS, 447, 3060

Read J. I., 2014, Journal of Physics G Nuclear Physics, 41, 063101

Read J. I., Steger P., 2017, MNRAS, 471, 4541

Read J. I., Walker M. G., Steger P., 2018, MNRAS, 481, 860

Read J. I., Walker M. G., Steger P., 2019, MNRAS, 484, 1401

Read J. I., Wilkinson M. I., Evans N. W., Gilmore G., Kleya J. T., 2006a, MNRAS, 367, 387

Read J. I., Wilkinson M. I., Evans N. W., Gilmore G., Kleya J. T., 2006b, MNRAS, 366, 429

Richardson T., Fairbairn M., 2014, MNRAS, 441, 1584

Robotham A. et al., 2010, PASA, 27, 76

Rojas-Niño A., Read J. I., Aguilar L., Delorme M., 2016, MNRAS, 459, 3349

Sanchis T., Lokas E. L., Mamon G. A., 2004, MNRAS, 347, 1198

Sanders J. L., Evans N. W., 2017, MNRAS, 472, 2670

Sanderson R. E. et al., 2017, arXiv e-prints, arXiv:1712.05420

Spencer M. E., Mateo M., Olszewski E. W., Walker M. G., McConnachie A. W., Kirby E. N., 2018, AJ, 156, 257

Strigari L. E., Bullock J. S., Kaplinghat M., 2007, ApJ, 657, L1

Syer D., Tremaine S., 1996, MNRAS, 282, 223

Theia Collaboration et al., 2017, arXiv e-prints, arXiv:1707.01348

Ural U., Wilkinson M. I., Read J. I., Walker M. G., 2015, Nature Communications, 6, 7599

van der Marel R. P., 1994, MNRAS, 270, 271

van der Marel R. P., Anderson J., 2010, ApJ, 710, 1063

Vasiliev E., 2019, MNRAS, 482, 1525

Verolme E. K. et al., 2002, MNRAS, 335, 517

Vitral E., Mamon G. A., 2020, arXiv e-prints, arXiv:2010.05532

Walker M. G., Mateo M., Olszewski E. W., 2009, AJ, 137, 3100

Walker M. G., Peñarrubia J., 2011, ApJ, 742, 20

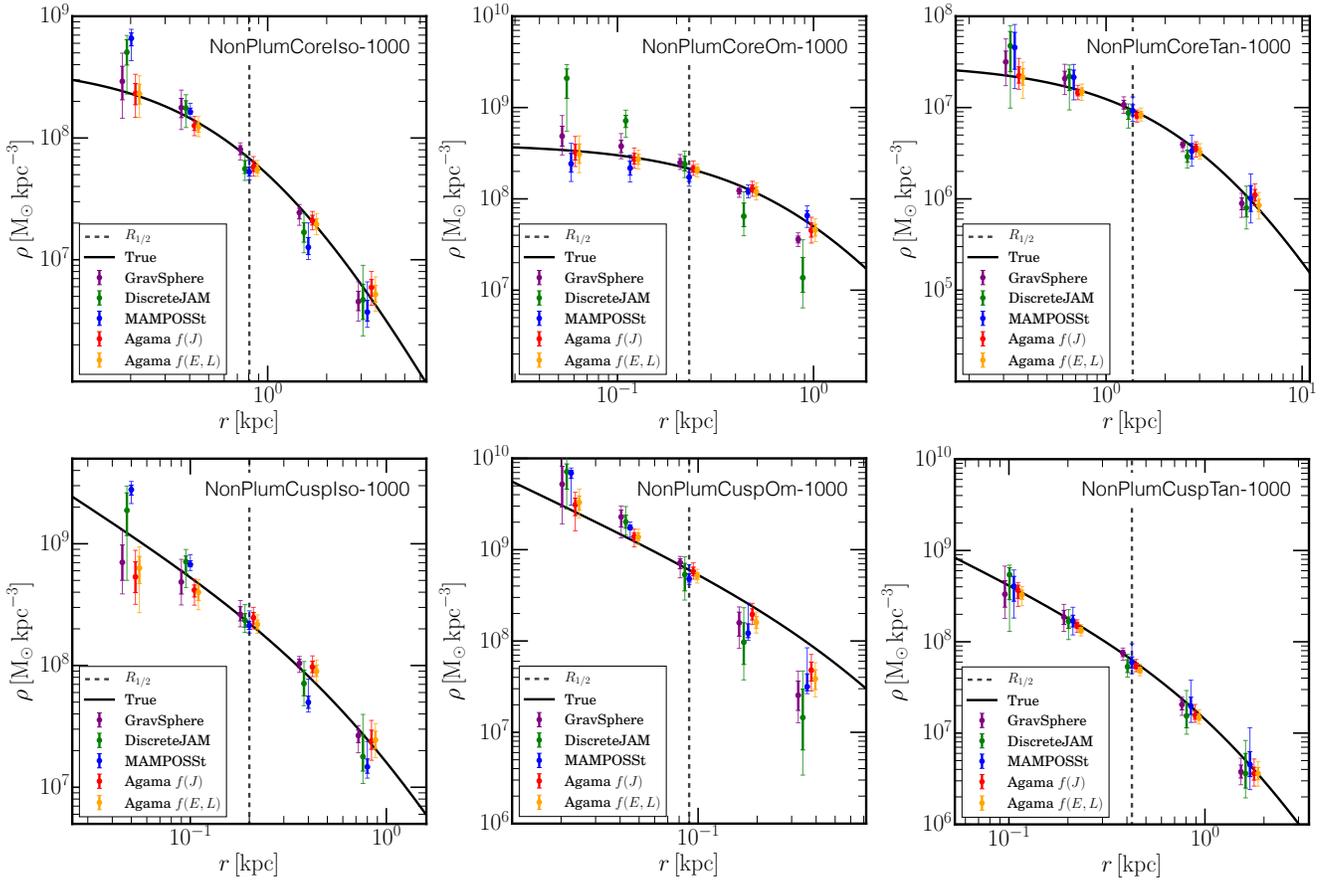
Watkins L. L., van de Ven G., den Brok M., van den Bosch R. C. E., 2013, MNRAS, 436, 2598

Watkins L. L., van der Marel R. P., Bellini A., Anderson J., 2015a, ApJ, 812, 149

Watkins L. L., van der Marel R. P., Bellini A., Anderson J., 2015b, ApJ, 803, 29

White R. E., Shawl S. J., 1987, ApJ, 317, 246

Wojtak R., Lokas E. L., Mamon G. A., Gottlöber S., 2009, MNRAS, 399, 812



**Figure A1.** As Figure 1 but for the ‘NonPlum’ spherical mocks, as marked on the panels.

- York D. G. et al., 2000, AJ, 120, 1579  
 Zhao H., 1996, MNRAS, 278, 488  
 Zhu L., van de Ven G., Watkins L. L., Posti L., 2016, MNRAS, 463, 1117  
 Zocchi A., Gieles M., Hénault-Brunet V., 2017, MNRAS, 468, 4429