To β or not to β: can higher order Jeans analysis break the mass–anisotropy degeneracy in simulated dwarfs?

A. Genina [,1][★] J. I. Read [,2] C. S. Frenk,[1] S. Cole,[1] A. Benítez-Llambay [,1] A. D. Ludlow [,3] J. F. Navarro,[4]† K. A. Oman [1] and A. Robertson [1]

[1]*Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*
[2]*Department of Physics, University of Surrey, Guildford GU2 7XH, UK*
[3]*International Centre for Radio Astronomy Research, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia*
[4]*Department of Physics and Astronomy, University of Victoria, Victoria, BC V8P 5C2, Canada*

ABSTRACT

We test a non-parametric higher order Jeans analysis method, GravSphere, on 32 simulated dwarf galaxies comparable to classical Local Group dwarfs like Fornax. The galaxies are selected from A Project Of Simulating The Local Environment (APOSTLE) suite of cosmological hydrodynamics simulations with cold dark matter (CDM) and self-interacting dark matter (SIDM) models, allowing us to investigate cusps and cores in density distributions. We find that, for CDM dwarfs, the recovered enclosed mass profiles have a bias of no more than 10 per cent, with a 50 per cent scatter in the inner regions and a 20 per cent scatter near the half-light radius, consistent with standard mass estimators. The density profiles are also recovered with a bias of no more than 10 per cent and a scatter of 30 per cent in the inner regions. For SIDM dwarfs, the mass and density profiles are recovered within our 95 per cent confidence intervals but are biased towards cuspy dark matter distributions. This is mainly due to a lack of sufficient constraints from the data. We explore the sources of scatter in the accuracy of the recovered profiles and suggest a χ^2 statistic to separate successful models from biased ones. Finally, we show that the uncertainties on the mass profiles obtained with GravSphere are smaller than those for comparable Jeans methods and that they can be further improved if stronger priors, motivated by cosmological simulations, are placed on the velocity anisotropy. We conclude that GravSphere is a promising Jeans-based approach for modelling dark matter distributions in dwarf galaxies.

Key words: galaxies: dwarf – galaxies: kinematics and dynamics – dark matter.

1 INTRODUCTION

Dark matter makes up ∼85 per cent of the matter density of the Universe (e.g. Planck Collaboration et al. 2018); yet its identity remains unknown. Dwarf satellite galaxies of the Milky Way are expected to be excellent sites for testing the properties of dark matter (Battaglia, Helmi & Breddels 2013). These objects have velocity dispersions indicative of a high dark matter content. If dark matter is a self-annihilating particle, the products of its annihilation may be detected with space- and ground-based instruments (Lake 1990), such as the Fermi Large Area Telescope (Albert et al. 2017) and the upcoming Cherenkov Telescope Array (Morselli & Consortium 2017). The detected spectra of these annihilation products would reveal the particle physics properties of dark matter; however, these analyses require the underlying dark matter distribution to be well constrained. Some of the classical dwarf spheroidals, like Sculptor and Fornax, host ∼10^6–10^8 M$_\odot$ in stars (McConnachie 2012) and are sufficiently close to our Galaxy that large samples of high-quality kinematic and photometric stellar data can be obtained, so inferences may be made of the underlying gravitational potential (e.g. Battaglia et al. 2008; Strigari et al. 2008; Walker et al. 2009; Charbonnier et al.

2011; Walker & Peñarrubia 2011; Bonnivard et al. 2015b; Read, Walker & Steger 2019; Strigari, Frenk & White 2018).

The exact dark matter density distribution in the central regions of dwarf galaxies is controversial (e.g. de Blok 2010). Galaxy rotation curves measured in some dwarf irregulars (Moore 1994; Oh et al. 2008; Adams et al. 2014; Oh et al. 2015; Zhu et al. 2016a; Read et al. 2017; Oman et al. 2019), analyses based on multiple tracer populations in dwarf spheroidals (Walker & Peñarrubia 2011; Amorisco & Evans 2012) and globular cluster survivability within dwarfs (Goerdt et al. 2006; Cole et al. 2012; Contenta et al. 2018; Orkney et al. 2019) have been used to argue in favour of dark matter 'cores', where the density remains constant in the central regions, with $\rho \propto r^0$ (Flores & Primack 1994; Moore 1994). On the other hand, dark matter-only *N*-body simulations in Λ-cold dark matter (ΛCDM) cosmologies have found 'cusps' in inner haloes, with density profiles scaling approximately as $\rho \propto r^{-1}$ (Navarro, Frenk & White 1996b, 1997). This became known as the 'core-cusp problem'. The problem has motivated an introduction of alternative dark matter models, such as self-interacting dark matter, where cores are created through dark matter self-scattering on a scale related to the interaction cross-section (Spergel & Steinhardt 2000; Elbert et al. 2015).

In recent years, the use of hydrodynamics to model baryonic processes has become more common in simulations. Outflows associated with supernova feedback have been shown to cause fluctuations in the gravitational potential, which can alter the inner structure of haloes.

★ E-mail: anna.genina@durham.ac.uk
† Senior CIfAR Fellow (JFN).

This may occur in a single violent burst (Navarro, Eke & Frenk 1996a), through repeated dark matter 'heating' over time (Read & Gilmore 2005; Pontzen & Governato 2012; Oñorbe et al. 2015; Read, Agertz & Collins 2016) or both (Benítez-Llambay et al. 2019). Cores in these models typically form on the scale of the half-light radius of dwarf galaxies.

The formation of cores in dwarfs that have undergone extended periods of star formation is a testable hypothesis. This idea was explored in Read et al. (2019). These authors find that dwarf spheroidals which have continued to form stars until recent times, like Fornax and the Local Group dwarf irregulars, have lower densities at 150 pc, ρ_{150}, than those predicted for isolated dwarfs using the halo mass–concentration relation in ΛCDM from Dutton & Macciò (2014). These lower densities could be explained by core formation through dark matter 'heating'. Dwarfs that have ceased star formation a long time ago have higher values of ρ_{150}, consistent with a cusp.

The dark matter density distribution in dwarf galaxies can be constrained through Jeans analysis applied to line-of-sight stellar velocities and projected positions. This relies on the spherical Jeans equation:

$$\frac{1}{\nu}\frac{d}{dr}\left(\nu\sigma_r^2\right) + 2\frac{\beta\sigma_r^2}{r} = -\frac{GM(<r)}{r^2}, \tag{1}$$

where ν is the tracer number density distribution, σ_r is their radial velocity dispersion, β their velocity anisotropy, $M(<r)$ is the enclosed mass, and G is Newton's gravitational constant. The velocity anisotropy, β, is defined as $\beta = 1 - \frac{\sigma_t^2}{2\sigma_r^2}$, where σ_t is the tangential velocity dispersion. The product $\nu(r)\sigma_r^2(r)$ is typically obtained through deprojection of $\Sigma(R)\sigma_P^2(R)$, where $\sigma_P(R)$ is the line-of-sight velocity dispersion and $\Sigma(R)$ is the projected tracer number density at a distance, R, both of which are observable quantities. If models are assumed for $\beta(r)$ and $M(<r)$, the equation can be solved for $\sigma(R)$ via sampling methods such as Markov Chain Monte Carlo (MCMC). This analysis assumes a non-rotating spherical system in a steady pseudo-equilibrium state. These assumptions are known to be violated by Local Group dwarfs, which exhibit ellipticity (McConnachie 2012), signs of rotation (Battaglia et al. 2008; del Pino et al. 2017), and are susceptible to tidal effects from their hosts (Read et al. 2006; Peñarrubia et al. 2009; Ural et al. 2015).

Typically, only the line-of-sight motions of the stellar tracers are known. This means that β is poorly constrained, such that Jeans analysis suffers from the M–β degeneracy. This degeneracy results in a wide range of models that satisfy a set of observational constraints, such that cored and cuspy dark matter profiles both provide acceptable fits to line-of-sight data (Strigari, Frenk & White 2010).

The breaking of $M - \beta$ has been widely explored in the literature. Several works have focused on the use of multiple tracer populations in dwarf spheroidals and chemodynamical models (Walker & Peñarrubia 2011; Amorisco & Evans 2012; Agnello & Evans 2012; Zhu et al. 2016b), as well as proper motions (Strigari, Bullock & Kaplinghat 2007; Strigari et al. 2018). A number of works have used Schwarzschild orbit superposition methods (Schwarzschild 1979; Jardel & Gebhardt 2012; Breddels & Helmi 2013; Kowalczyk, Łokas & Valluri 2017, 2018; Kowalczyk et al. 2019), which are able to take into account the asphericity of stellar systems and have a benefit of making no assumptions about the velocity anisotropy. These methods, however, typically require significant computing time.

Other works, based on the Jeans equation, have focused on exploiting the higher order velocity moments. Specifically, it has been shown that line-of-sight velocity distributions are non-Gaussian

in the absence of isotropy (Merritt 1987). This warrants the use of the fourth moment of the velocity distribution to place a constraint on the anisotropy parameter (Mamon, Biviano & Boué 2013). The use of the fourth velocity moments has been explored for the case of constant anisotropy by Łokas (2002), Łokas & Mamon (2003), Łokas, Mamon & Prada (2005), Łokas (2009) and was generalized for radially varying anisotropies by Richardson & Fairbairn (2013). More recently, Diakogiannis et al. (2017) have presented a non-parametric method of reconstruction of the line-of-sight velocity dispersion profiles. An extension of the method employs a machine learning approach for data reconstruction that proves useful in the absence of large samples of kinematic data (Diakogiannis et al. 2019).

Another method for breaking this degeneracy is through the fourth-order projected virial theorem, giving rise to two equalities (Merrifield & Kent 1990):

$$\text{VSP1} = \frac{2}{5}\int_0^\infty GM\nu(5 - 2\beta)\sigma_r^2 r\,dr = \int_0^\infty \Sigma\left\langle\sigma_P^4\right\rangle R\,dR \tag{2}$$

and

$$\text{VSP2} = \frac{4}{35}\int_0^\infty GM\nu(7 - 6\beta)\sigma_r^2 r^3\,dr = \int_0^\infty \Sigma\left\langle\sigma_P^4\right\rangle R^3\,dR, \tag{3}$$

where VSP1 and VSP2 are referred to as the virial shape parameters (i.e. VSPs). Here $\left\langle\sigma_P^4\right\rangle$ is the fourth moment of the line-of-sight velocities. The right-hand sides of equations (2) and (3) contain quantities that can be directly inferred from data and the left-hand sides contain the same parameters as equation (1). It is thus possible to place two additional constraints on the velocity anisotropy β (see e.g. Richardson & Fairbairn 2014, where VSPs were used to show that a dark matter cusp is favoured in the Sculptor dwarf galaxy). In practice, however, the finite quality of data may result in only a partial breaking of the $M - \beta$ degeneracy.

Read & Steger (2017) introduced the non-parametric Jeans method, GravSphere (used in Read et al. 2019), which employs the additional constraints from the VSPs in their MCMC analysis. GravSphere operates under the standard assumptions of the spherical Jeans equation (spherical symmetry, equilibrium, and no rotation). The method had been shown to recover successfully the dark matter density distributions in mock observations of idealized spherical, triaxial and tidally stripped simulated dwarfs from the Gaia Challenge set.[1] The cases for which the method works less well, such as aspherical systems, are evident through poor quality fits to the line-of-sight velocity dispersion. The method has been shown to recover accurately the densities at 150 pc from the centre – a key region where core formation is expected to reduce dark matter densities, compared to ΛCDM predictions.

In this work, we test a new open-source PYTHON implementation of the GravSphere method, PYGRAVSPHERE, on a sample of simulated dwarf galaxies from cosmological hydrodynamics simulations of Local Group-like environments.[2] The aim of our work is to establish whether Jeans analysis, under the assumption of spherical symmetry, is a suitable method for constraining the mass profiles of dark matter haloes in a fully cosmological setting and how much information is typically gained through the inclusion of VSPs. We explore the biases associated with mass and dark matter density profile recovery for individual dwarfs as well as the sample as a whole. We examine in detail the cases where GravSphere fails and identify the reasons for this as well as potential warning signs.

---

[1] http://astrowiki.ph.surrey.ac.uk/dokuwiki/

[2] We ran also these same tests in the original GravSphere code, yielding indistinguishable results from those presented here.

In Section 2, we describe the suite of simulations used, as well as the setup of GravSphere. Our analysis of GravSphere's performance on each galaxy and the comparison to more conventional Jeans methods can be found in Section 3. We further discuss the various sources of bias in the performance of this method and suggest a χ^2 statistic to weed out particularly biased models. We summarize our findings and conclude in Section 4.

2 SIMULATIONS

2.1 APOSTLE simulations

In this work, we select analogues of classical dwarf spheroidals from the APOSTLE suite of cosmological simulations. APOSTLE volumes are zoom-in simulations of Milky Way and Andromeda analogue pairs, selected from a dark matter-only volume. The pairs have been chosen to satisfy constraints for the Local Group, such as the total mass, separation and relative velocities. The details of halo selection may be found in Fattahi et al. (2016). Each Milky Way or Andromeda analogue hosts a population of dwarf galaxies. The volumes additionally include isolated dwarfs.

APOSTLE was run with the EAGLE model of galaxy formation (Schaye et al. 2015; Crain et al. 2015), which is based on the smoothed particle hydrodynamics (SPH) *N*-body code GADGET-3, an improved version of the GADGET-2 code (Springel 2005). The original APOSTLE suite consists of five high-resolution cosmological volumes, with dark matter mass resolution of $m_{DM} = 2.5$–5×10^4 $M_\odot$ and spatial resolution $\epsilon = 134$ pc. An extra cosmological volume was also run assuming a self-interacting dark matter (SIDM) interaction cross-section of $\sigma/m = 10$ cm^{2}g^{-1} (Santos-Santos et al. 2020; Lovell et al. 2020). We note that this is an extreme value of the cross-section, which was chosen to explore the formation of the largest cores in SIDM. The SIDM implementation within EAGLE was introduced in Robertson et al. (2018), based on the SIDM simulation method described in Robertson, Massey & Eke (2017).

In order to increase our sample of dwarfs with a dark matter core, we additionally used an SIDM version of the cosmological volume presented in Benítez-Llambay et al. (2019),[3] with an interaction cross-section of $\sigma/m = 10$ cm^{2}g^{-1} and galaxy formation prescriptions following Schaye et al. (2015) and Crain et al. (2015). This simulation was run with the same cosmological parameters as the APOSTLE simulations but does not feature a Local Group-like setting. The dark matter particle mass resolution is $m_{DM} = 4 \times 10^5$ $M_\odot$ and the softening is $\epsilon = 234$ pc.

Each stellar particle in our simulations represents a stellar population assumed to follow a Chabrier (2003) initial mass function. Gas particles in APOSTLE have initial masses in the range $5-10 \times 10^3$ $M_\odot$. APOSTLE resolves Sculptor-mass dwarf galaxies with $\sim 10^2$–10^3 stellar particles and Fornax-mass dwarf galaxies with $\sim 10^3$–10^4 particles. Further details of the APOSTLE simulations may be found in Sawala et al. (2016), Fattahi et al. (2016), and Campbell et al. (2017). The SIDM run of the volume presented in Benítez-Llambay et al. (2019) has an initial gas particle mass $m_{gas} = 6.6 \times 10^4$ $M_\odot$. Fornax-mass dwarfs are resolved with $\sim 10^2$–10^3 stellar particles, sufficient for the purposes of this work.

2.2 Numerical considerations

In order to establish whether GravSphere reproduces the mass profiles of the simulated dwarfs, we must first define what the 'true' mass profile is within the simulations.

The mass profiles of dark matter haloes identified in pure *N*-body simulations are affected by collisional relaxation. The enclosed mass profiles of haloes are suppressed (relative to a higher resolution simulation) below a radius where the two-body relaxation time, t_{relax}, is comparable to the age of the Universe, t_0 (see e.g. Power et al. 2003; Ludlow, Schaye & Bower 2019b). For the typical number of dark matter particles in systems that are considered in this work, the Power et al. (2003) radius, where $t_{relax} \sim 0.6 t_0$, corresponds to ~ 60 per cent of the half-light radius (about 0.7 kpc for high-resolution APOSTLE CDM and SIDM dwarfs and ~ 2 kpc for the lower-resolution SIDM volume from Benítez-Llambay et al. 2019). Moreover, the APOSTLE simulations model dark matter and stars using particles of unequal mass, making them subject to energy equipartition, which artificially inflates galaxy sizes (Ludlow et al. 2019a, 2020). These effects are most problematic for systems with stellar half-mass radii smaller than ~ 0.055 of the mean interparticle separation (dark matter, stars, and gas), corresponding to ~ 0.5 kpc for the APOSTLE simulations and ~ 1.2 kpc for the SIDM volume of Benítez-Llambay et al. (2019). We note, however, that the relaxation times at these radii are still considerably longer than the dynamical times of the stars. Therefore, our simulated dwarfs may be considered to be in a steady state locally. The ability of GravSphere to recover masses in the innermost regions is of interest in this work; therefore we will present results below the convergence radius derived by Power et al. (2003) and Ludlow et al. (2019b), although we will interpret these with caution.[4]

The use of the gravitational softening in *N*-body simulations sets a limit on the central density, such that the innermost regions of simulated haloes exhibit a small artificial core on the scale of the gravitational softening. In comparing GravSphere to *N*-body simulations, we thus restrict ourselves to radii greater than 2.8ϵ, where ϵ is the Plummer-equivalent gravitational softening. For the APOSTLE high-resolution simulations $\epsilon = 134$ and 234 pc for the SIDM version of the volume introduced in Benítez-Llambay et al. (2019).[5] At the radius of 2.8ϵ the forces become exactly Newtonian.

2.3 Sample of dwarfs

For our sample of galaxies, we have selected simulated dwarfs with comparable properties to classical Milky Way satellites and at least 400 bound stellar particles. We restrict our sample to satellites only, defined as the objects within 300 kpc of the Milky Way and Andromeda analogues. As the GravSphere method relies on the standard assumptions of the spherical Jeans equation (namely, spherical symmetry, lack of rotation, and equilibrium), we further restrict our sample of galaxies to those with no bound gas and those that do not exhibit significant signs of rotation. We do, however, include aspherical dwarfs in our sample. These conditions are satisfied by classical Milky Way dwarfs, for which there are no H I detections, only weak rotation is observed and asphericity is often

---

[3]The SIDM version of this volume has not yet been published. The simulation data were obtained through private communications.

[4]The Power et al. (2003) criterion was derived for dark matter-only simulations and it is unclear how applicable this criterion is in the presence of a baryonic component.

[5]Although 2.8ϵ for this volume is near 0.5 kpc, the core sizes are typically much larger than that, such that we are still able to probe the interesting regions in these dwarfs.
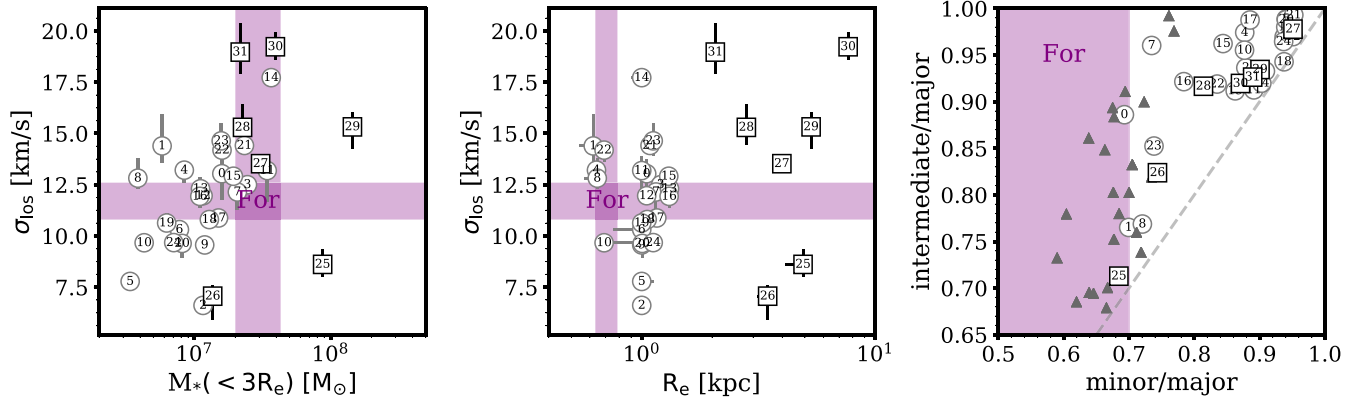
Figure 1. The sample of classical dwarf analogues. CDM dwarfs are shown with grey circles and SIDM dwarfs are shown with black squares. Left: line-of-sight velocity dispersion as a function of stellar mass within three times the projected half-light radius. The purple bands show the measurements for Fornax from McConnachie (2012) and de Boer et al. (2012). Middle: line-of-sight velocity dispersion as a function of the projected half-light radius. Right: ratio of intermediate to major axes as a function of the ratio of minor to major axes for our sample of dwarfs. The limits on the sphericity of Fornax dwarf galaxy (minor-to-major axial ratio $\simeq 0.7$), which has a 2D ellipticity $e \simeq 0.3$, are shown with a purple shaded band. The grey triangles show a sample of 24 isolated APOSTLE dwarfs with axis ratios comparable to the upper limit set on Fornax. The dashed one-to-one line highlights prolate galaxy shapes.

present (Battaglia et al. 2008; McConnachie 2012). We quantify rotation by the fraction of stellar particles which are rotating in the same direction as the total stellar angular momentum vector, f_{corot}. We select galaxies with $f_{corot} < 0.6$. Note that, in order to obtain a statistically significant sample of dwarfs in both CDM and SIDM dark matter models, we do not include restrictions on the orbits or star formation histories in the selection of our sample of simulated dwarfs.

Our selection criteria have significantly cut down the available sample of high-resolution SIDM dwarfs (which have large dark matter cores), prompting inclusion of simulated dwarfs from the lower resolution SIDM version of the volume presented in Benítez-Llambay et al. (2019). The relative lack of suitable dwarfs in SIDM is intriguing, however, the peculiarities of galaxy formation and evolution in alternative dark matter models are beyond the focus of this work (see Lovell et al. 2020 for details of star formation in the APOSTLE SIDM volume).

The resulting sample consists of dwarfs with properties that are generally similar to Fornax. This includes 25 CDM dwarfs and 7 SIDM dwarfs (two high resolution and five lower resolution). Structural and kinematic properties of the sample are plotted in Fig. 1. For each dwarf in Fig. 1, we show the distribution of projected half-light radii, R_e, from various line-of-sight projections, line-of-sight velocity dispersion, σ_{los}, and stellar mass within $3R_e$, $M_*(<3R_e)$. These properties are computed using 192 isotropically distributed lines of sight generated with the HEALPIX algorithm (Górski et al. 2005). HEALPIX provides a convenient way of generating an isotropically distributed set of lines of sight with a more uniform distribution than one generated randomly. To calculate R_e, to each projection we fit a three-component 2D Plummer profile (more details may be found in Section 2.4.1) and compute the radius which contains half the projected stellar mass. The means and standard deviations of these values are shown with circles and their error bars, with errors primarily reflecting the asphericity of each system. Similarly, we compute the mass-weighted mean velocity dispersion along each projection, taking into account the error from sample size (the standard error on the mean). The stellar masses within $3R_e$ are computed by summing the masses of stellar particles identified as bound by the SUBFIND algorithm (Springel et al. 2001; Dolag et al. 2009) and removing the contaminant stars that belong to the host

galaxy.[6] The contaminant stars are also removed in the calculation of other galaxy properties and the Jeans analysis. Purple bands in Fig. 1 show corresponding properties of the Fornax dwarf galaxy. The velocity dispersion and the half-light radius of Fornax were taken from McConnachie (2012). Stellar masses for Fornax span the range of values from literature (McConnachie 2012; de Boer et al. 2012).

The rightmost panel of Fig. 1 shows intermediate-to-major axis ratios as a function of minor-to-major axis ratios, computed for the stellar component. The axes have been derived from the eigenvalues of the reduced inertia tensor, computed for the stellar particles (see e.g. Bett et al. 2007). The purple dashed line is the upper limit set on the minor-to-major axial ratio of Fornax, determined by the measured projected ellipticity, $e = 1 - b/a \simeq 0.3$ (Battaglia et al. 2006). Our sample, which has an average axial ratio of $c/a \sim 0.9$, is considerably more spherical than Fornax ($c/a \simeq 0.7$). This is a consequence of our selection criteria, specifically the lack of rotation and the lack of gas for the selected sample of dwarfs. These are more likely to be features of dwarf galaxies that have undergone tidal effects, which tends to reduce asphericity (Barber et al. 2015).

In order to further investigate the effects of asphericity, we included a sample of 24 isolated CDM dwarfs with $c/a \sim 0.7$ (grey triangles in Fig. 1). The sample has been chosen to contain galaxies where gas does not dominate by mass within the 3D half-mass radius of the stars. This is the reason for the lack of isolated dwarfs available for the SIDM sample. In fact, we did find four isolated SIDM dwarfs that match these criteria, but they turned out to be very oblate. We excluded these dwarfs from our analysis. Since the GravSphere method explicitly accounts for the mass contributed by stars, but not by gas, when comparing GravSphere's performance for these dwarfs with the 'true' values, we compare to the combined mass in gas and dark matter.

In order to generate the photometric and kinematic data required by GravSphere, we obtained the stellar positions and velocities for particles in each subhalo, which were classified as bound by SUBFIND

---

[6]Due to the specifics of particle assignment to haloes in SUBFIND, some low-velocity host halo stars may end up attributed to a subhalo. We remove these stars by ensuring they were not 'bound' to any given subhalo in the previous simulation snapshot (Springel et al. 2001).

(Dolag et al. 2009). We removed all contaminant stars belonging to the host halo. Each stellar system was centred at the central peak of the density field, computed using the 'shrinking spheres' algorithm (Power et al. 2003). The positions and velocities were then projected along three different lines of sight – major, minor, or intermediate axes of the subhaloes. The stellar particles within $2R_{2D}$, where R_{2D} is the projected radius that contains half the stellar particle mass, were randomly sampled, providing 400–2500 particles for the kinematic sample of each dwarf. The photometric sample included 400–2500 particles and was chosen to be the same or bigger in size than the kinematic sample. The velocities within the kinematic sample were perturbed with Gaussian noise with a standard deviation of 2 km s^{-1}, representing typical measurement errors. The mass within $3R_{2D}$ has additionally been provided, $M(<3R_{2D})$, such that GravSphere may include the stellar mass contribution to the gravitational potential in the Jeans modelling.

2.4 The PYGRAVSPHERE code

In this work we present a PYTHON implementation of the GravSphere method, PYGRAVSPHERE. PYGRAVSPHERE is open source software.[7] As in the work of Read & Steger (2017), PYGRAVSPHERE is based on the affine-invariant ensemble sampler EMCEE (Foreman-Mackey et al. 2013). EMCEE differs from the classic Metropolis–Hasting algorithm in that each individual Markov Chain, or 'walker', communicates with the other 'walkers' at each step, thus allowing the chains to efficiently sample the posterior distribution. EMCEE has parallel functionality, which we exploit in this work. In the following, we outline the assumptions and parameters that enter into our EMCEE setup.

2.4.1 EMCEE parameters

As in Read & Steger (2017), to parametrize the dark matter distribution, PYGRAVSPHERE employs a broken power-law model with 5 spatial bins defined as logarithmically spaced fractions of the half-light radius,[8] R_e, with bins $r_j = [0.25, 0.5, 1, 2, 4]R_e$. Within each bin, the density follows a power law defined by slopes γ_j. The overall distribution is described by

$$\rho_{dm}(r) = \begin{cases} \rho_0 \left(\dfrac{r}{r_0} \right)^{-\gamma_0}, & r < r_0 \\ \rho_0 \left(\dfrac{r}{r_{j+1}} \right)^{-\gamma_{j+1}} \prod_{n=0}^{n<j+1} \left(\dfrac{r_{n+1}}{r_n} \right)^{-\gamma_{n+1}}, & r_j < r < r_{j+1} \end{cases} \tag{4}$$

where ρ_0 is the density at r_0. Beyond the outermost bin, the power law is extrapolated. Note that this radial extent typically covers the positions of available kinematic tracers.

For the light profile, PYGRAVSPHERE uses a sum of three Plummer (1911) components ($N_P = 3$):

$$\nu(r) = \sum_{j}^{N_P} \frac{3M_j}{4\pi a_j^3} \left(1 + \frac{r^2}{a_j^2} \right)^{-5/2}, \tag{5}$$

---

[7]https://github.com/AnnaGenina/pyGravSphere

[8]In this work, we use multiple definitions of the half-mass radius, which we list here for the purposes of clarification. R_{2D} is the projected radius which contains half of a dwarf's stellar mass and it is computed by direct summation. R_e is the projected radius containing half the stellar mass, derived from a three-component Plummer profile fit to the *sample* of stellar particle data. We frequently refer to R_e as the half-light radius.

where M_j and a_j are the relative weight and spatial extent of each component, respectively. This distribution is straightforward to project, yielding:

$$\Sigma(R) = \sum_{j}^{N_P} \frac{M_j}{\pi a_j^2} \left(1 + \frac{R^2}{a_j^2} \right)^{-2}. \tag{6}$$

The velocity anisotropy is parametrized following Baes & van Hese (2007):

$$\beta(r) = \beta_0 + (\beta_\infty - \beta_0) \frac{1}{1 + \left(\frac{r_t}{r} \right)^\eta}, \tag{7}$$

where β_0 is the central value of the anisotropy, β_∞ is the value at infinity, r_t is the radius of transition, and η is its steepness.

2.4.2 PYGRAVSPHERE data input

The photometric sample of stars is split into bins of $N_{phot}/\sqrt{N_{phot}}$ particles per bin, where N_{phot} is the size of the photometric sample. This choice allows for efficient spatial coverage and low Poisson error. We weight each particle by the relative number of stars it represents (i.e. we define the weight of each particle as $w_p = m_p N_{tot}/M(<3R_{2D})$, where N_{tot} is the total number of particles in the sample). For each photometric bin we calculate the Poisson errors and use the LMFIT algorithm (Newville et al. 2014) to obtain the best three-component Plummer fit (equation 6). This profile is then input into EMCEE.

The kinematic data are also split into bins of $N_{kin}/\sqrt{N_{kin}}$ particles per bin, where N_{kin} is the size of the kinematic sample. The error in each bin is computed by adding the Poisson and sampling errors in quadrature, where we again weight each particle by w_p. This procedure is described in detail in Read & Steger (2017). The effect of the number of particles per bin was explored in Read & Steger (2017), who found little impact on their results. We confirm this to be the case, provided the signal-to-noise ratio is not low due to 'overbinning' in the inner, dense, regions.

We use the same kinematic bins to compute the mean and errors of the two VSPs. Because VSP2 is sensitive to the behaviour of the velocity dispersion profile in the outer regions (due to the R^3 term), we fit a power law to the computed v_P^4 profile outside of the projected half-light radius and extrapolate it following Read, Walker & Steger (2018).

2.4.3 EMCEE set-up and priors

The priors on each of the parameters in the default PYGRAVSPHERE set-up are shown in Table 1. Parameters $\log_{10}M_j$ and a_j of the best three-component Plummer fit are allowed to vary within 50 per cent of their linear best-fitting values, as determined by LMFIT, while the stellar masses $\log_{10}M_*$ were varied within 25 per cent of the $M(<3R_{2D})$ value.

To ensure that the walkers sample the whole hypervolume of parameter space, defined by the parameter constraints, the starting positions of walkers are ideally generated to follow a uniform distribution. This is difficult to achieve when the values of the dark matter slope, γ_j, in each radial bin, r_j, are constrained to monotonically increase. We find that typically ~5 per cent of the initially generated walker positions fall within these defined bounds. If the chains are allowed to run for long enough, the walkers that are 'stuck' in forbidden regions of the parameter space may eventually make their way to the allowed regions. This process is, however, dependent on the efficiency of the active walkers in probing the

Table 1. Default GravSphere priors.

Property	Parameter	Prior
Dark matter	$\log_{10}\rho_0/\mathrm{M}_\odot\,\mathrm{kpc}^{-3}$	[5, 10]
	$\gamma_{0,1,2,3,4}$	[0, 3], $\Delta\gamma_{\max} = 1$
Anisotropy	$\tilde{\beta}_0$	[−1, 1]
	$\tilde{\beta}_\infty$	[−1, 1]
	$\log_{10}r_t/\mathrm{kpc}$	$\log_{10}[R_e/2, 2R_e]$
	η	[1, 3]
Tracers	$\log_{10}M_j/\mathrm{M}_\odot$	$\log_{10}[M_{\mathrm{bf},j}/2\,,\,3/2M_{\mathrm{bf},j}]$
	a_j/kpc	$[a_{\mathrm{bf},j}/2,\,3/2a_{\mathrm{bf},j}]$
Baryons	$\log_{10}M_*/\mathrm{M}_\odot$	$\log_{10}[0.75M(<3R_{\mathrm{2D}}),$
		$1.25M(<3R_{\mathrm{2D}})]$

posterior distribution. In cases where the posterior is multimodal, for example, some regions of the parameter space will not be probed due to the nature of the ensemble sampler, where the walkers communicate with each other, unlike in the classical Metropolis–Hastings algorithm.

We thus use the following procedure to generate the initial walker positions. For each walker, we generate the free parameters following a uniform distribution. We then throw away the walkers that do not satisfy our γ_j constraints. For the discarded walker positions we generate new ones, accepting those that satisfy the constraints and rejecting the others. This procedure is repeated recursively until each walker has a randomly generated initial position that satisfies the constraints. The effective priors and the advantages of the method over the one implemented in the original works using GravSphere (e.g. Read & Steger 2017; Read et al. 2018) are discussed in Appendix A.

We use 1000 walkers to probe the posterior distribution. Each walker is run for 10^4 steps as a conservative 'burn-in' measure, and then for a further 10^4 steps. The results presented in Appendix A suggest that our walkers are converged after $\sim 6 \times 10^3$ steps, so the above choices are rather extreme.

Since the anisotropy parameter, β, can take on values between 1 and $-\infty$, one would benefit from transforming this into a finite range. As in Read & Steger (2017), we use the symmetrized anisotropy parameter,

$$\tilde{\beta} = \frac{\beta}{2-\beta}, \quad \text{with} \quad -1 < \tilde{\beta} < 1. \tag{8}$$

This symmetrized form of the anisotropy allows the EMCEE walkers to probe the entire range of possible anisotropy values. In practice, we apply the constraints $\tilde{\beta}_0 > -0.95$ and $\tilde{\beta}_\infty > -0.95$, as for more negative values the calculation becomes numerically unstable.

PYGRAVSPHERE solves the Jeans equation for the projected velocity dispersion profile, $\sigma_P(R)$. It additionally fits the projected number density distribution, $\Sigma(R)$, and the two VSPs. We define the 'log-likelihood function' as the chi-squared sum of these four components:

$$\ln \mathcal{L} = -\frac{1}{2} \left(\chi^2_{\sigma_{\mathrm{LOS}}} + \chi^2_{\Sigma} + \chi^2_{\mathrm{VSP1}} + \chi^2_{\mathrm{VSP2}} \right). \tag{9}$$

We note that equation (9) involves quantities that are, to some extent, correlated. We must, therefore, consider whether the form of the likelihood function in equation (9) is justified. It has been pointed out in Łokas & Mamon (2003) and Łokas (2009) that the correlation between the second and the fourth velocity moments is typically weak. Aside from the fourth velocity moment, the virial

shape parameters are also related to the projected surface density $\Sigma(R)$. Given that the photometric samples for classical dwarfs are typically large, $\Sigma(R)$ is measured to sufficiently high accuracy, such that its uncertainties are negligible compared to those of the fourth velocity moment. Moreover, Read & Steger (2017) find that the estimates of the second velocity moment and the virial shape parameters are typically normally distributed. This, together with the weak correlation between the second and fourth velocity moments, suggests that equation (9) is a good approximation to the true likelihood function.

3 RESULTS

In this section, we present the performance of the GravSphere method, under a default setup, on each of the dwarfs in our sample. We further quantify its global performance, for the entire sample, and compare to standard Jeans analysis approaches. We identify the causes of bias and scatter in the recovered enclosed mass profiles.

3.1 Individual dwarfs

Fig. 2 shows the recovery by GravSphere of the cumulative mass, density, and the velocity anisotropy profiles of all dwarfs in our sample. In this figure, we only display results for projections along the vector to the centre of the host galaxy. The 'true' mass, density, and anisotropy profiles are displayed in blue and the GravSphere results in grey shaded bands. To generate the GravSphere profiles, we took 10^5 random samples from the output MCMC chains and for each radial position computed the median and the 68th and the 95th percentiles. The choice of 10^5 samples is sufficient to produce representative posteriors, but is otherwise arbitrary.

Mass profiles of simulated dwarfs were computed by summing dark matter particle masses radially from the centre of each dwarf, where we define the centre as the centre of mass of the stellar component. Densities were computed in 31 logarithmically spaced bins in radius, from $\log_{10}r/\mathrm{kpc} = -2$ up to the furthest bound dark matter particle. In order to compute the stellar velocity anisotropy profiles of simulated dwarfs, we bin the stars in each dwarf into 50 logarithmically spaced bins, starting from the position of the star that is closest to the centre of mass and ending at the outermost star. We then reduce the number of bins and widen the bin edges such that each bin has at least 50 stars within. We construct 1σ error bars by taking 1000 random samples of 25 stars with replacement and computing the standard deviation of the velocity anisotropy in each bin.

In this analysis, we focus on the key region within the half-light radius of each galaxy (solid vertical blue lines in Fig. 2) and above 2.8ϵ (dashed blue lines). GravSphere appears to be unbiased on average. For the CDM sample, the dwarf mass, density, and anisotropy profiles are typically contained within the 68 per cent confidence limits, although the uncertainties can be large. Notable exceptions in the CDM sample are Galaxies 1 and 8. These are known aspherical objects (see Fig. 1). Similar systematics can be seen for the remainder of the aspherical sample of dwarfs. It is clear that in SIDM dwarfs the enclosed mass is *always* overestimated, although even in these galaxies the true profiles are contained within the 95 per cent confidence regions. We will return to this issue in Section 3.5.

We also note that the symmetrized anisotropy profiles in our simulations are generally consistent with being constant and isotropic ($\tilde{\beta} = 0$). Deviations from this are seen in galaxies that are aspherical, and, as we will see in Section 3.7.3, in those which are affected by tides.
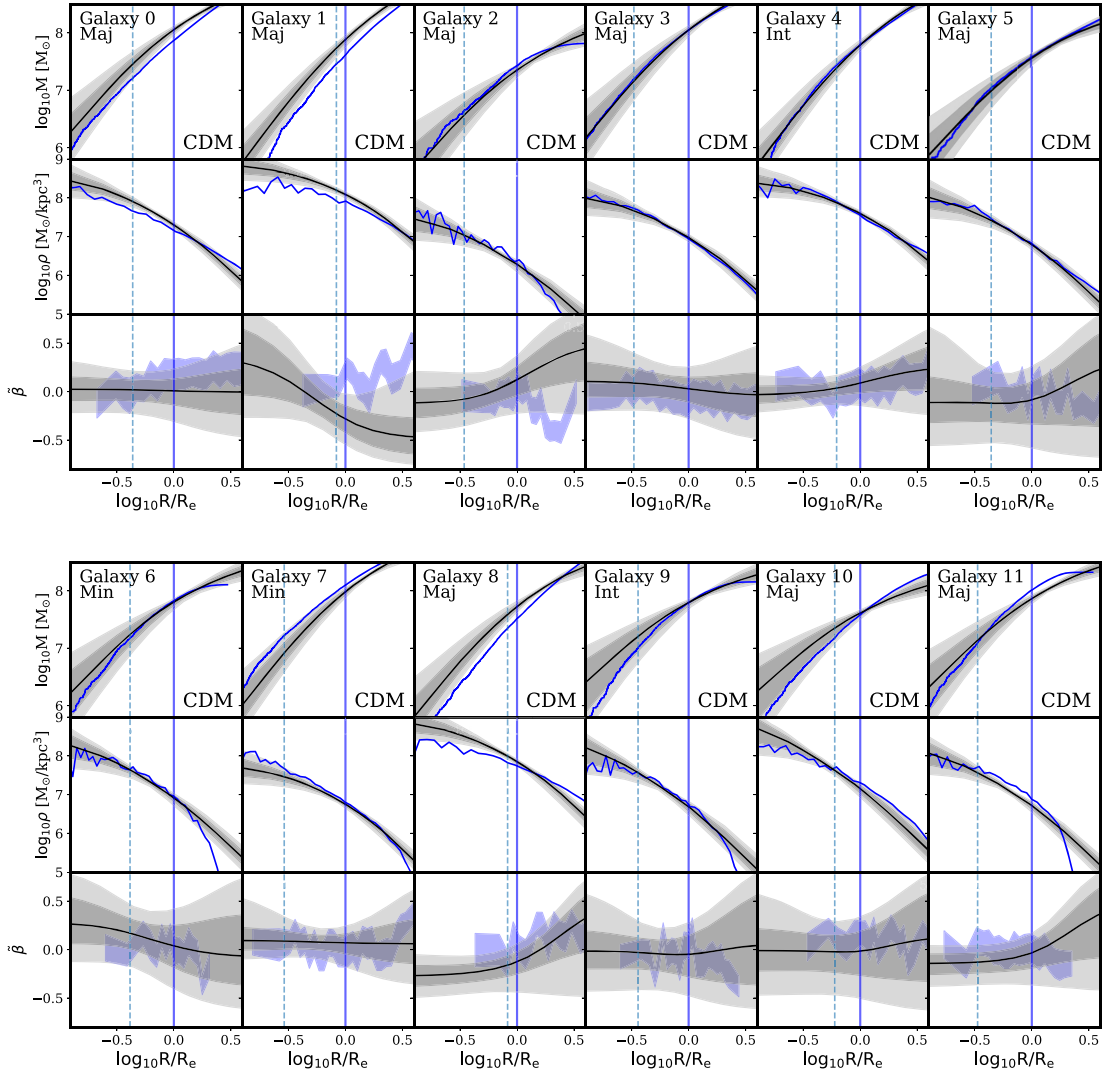
Figure 2. Enclosed dynamical mass profiles (top), density profiles (middle), and symmetrized anisotropy profiles recovered for a sample of 32 simulated dwarf galaxies using the GravSphere method. Only one projection (from the centre of the host galaxy) is shown for each dwarf. The profiles are shown as a function of 3D radius, normalized by the projected half-light radius, R_e. The black lines and the dark and light grey shaded bands display the median, 68 and 95 per cent confidence limits, respectively. The 'true' mass and density profiles, measured directly from the simulation, are shown in blue. The solid vertical blue line shows the location of the projected half-light radius and the dashed blue line shows 2.8ϵ ($\approx$380 pc for CDM dwarfs and SIDM dwarfs (25,26) and $\approx$655 kpc for SIDM dwarfs 27–31), which is close to the convergence radius for these systems. The 'true' velocity anisotropy, as measured directly from the stellar particles, is shown with a shaded blue band. For each galaxy, in the top left corner, we display the unique galaxy number and the principal axis along which the galaxy was projected to produce this figure (which is the principal axis most closely aligned with the vector to the host galaxy). In the bottom right corner we show whether the galaxy is from CDM or SIDM cosmology. Galaxies 25–31 are SIDM dwarfs. Note that for a number of these dwarfs the spatial resolution, 2.8ϵ (vertical dashed line), is below $0.125R_e$ and below the limits of the figure.

3.2 Comparison to standard estimators

We now compare the performance of GravSphere in recovering enclosed masses to three mass estimators from the literature. The mass estimators take the form:

$$M(< \mu R_e) = \frac{\lambda R_e \langle \sigma_P^2 \rangle}{G}, \qquad (10)$$

where μ and λ are constants.

The estimator provided by Wolf et al. (2010) gives the mass enclosed within the deprojected half-light radius R_3, with $R_3 \approx 4/3R_e$; the Walker et al. (2009) estimator gives the enclosed mass at the projected half-light radius, R_e. Another estimator has been derived by Campbell et al. (2017) for the mass within $1.44R_{2D}$, where R_{2D} is

the projected radius containing half the stellar mass. The latter has been calibrated on dwarfs from the APOSTLE suite of simulations that we use here. We omit the estimator derived by Errani, Peñarrubia & Walker (2018), which we find produces similar results to those of Campbell et al. (2017).

As input for the estimators, we use the half-light radius, R_e, obtained from the best-fitting three-component Plummer profile, as this is expected to provide more accurate results than the circular radius containing half the stellar mass, R_{2D} (González-Samaniego et al. 2017). The latter definition was used in the calibration of the Campbell et al. (2017) estimator, however, we refrain from using this definition as the spatial extent of the sample of stellar particles in each dwarf is already cut-off at $2R_{2D}$. For Wolf et al. (2010) and Walker et al. (2009) estimators we calculate the mean velocity dispersion
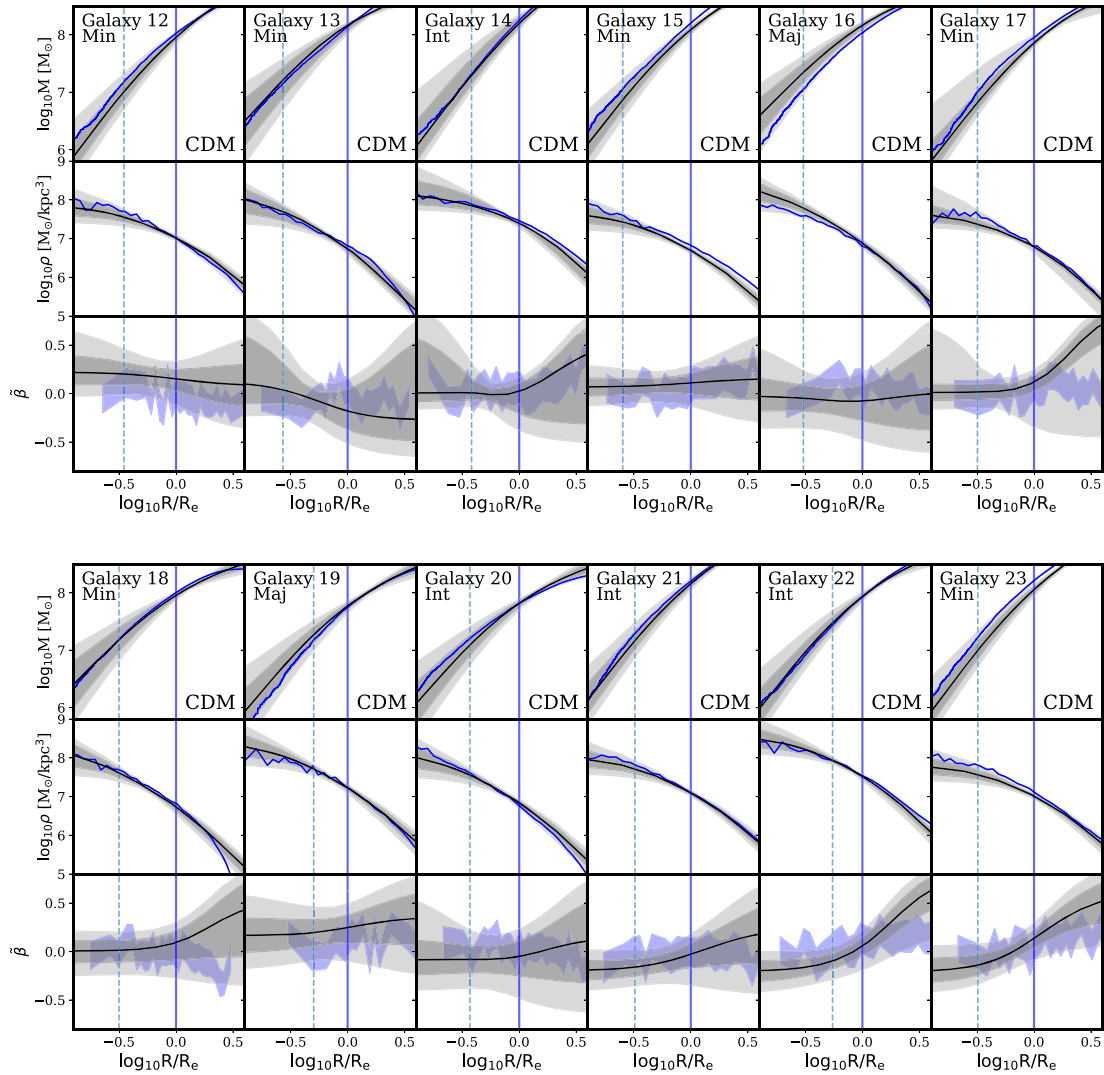
Figure 2 – *continued*

and the associated errors using the technique of Walker et al. (2006), where we incorporate weighting by the number of stars per stellar particle, w_p, in the likelihood function. For the Campbell et al. (2017) estimator we compute the mass-weighted mean velocity dispersion below $1.04R_e$, as prescribed.

In the left-hand panel of Fig. 3, we show the bias in mass profiles returned by GravSphere (mass recovered by GravSphere divided by the true mass) and the 68 and 95 per cent confidence levels, as a function of normalized radius, R/R_e. We use the following procedure to compute the global radial bias and the associated confidence levels. For each galaxy and for each distance, R, we obtain a cumulative mass distribution from 10^5 random samples of the MCMC posteriors. We then use this cumulative distribution for Monte Carlo sampling of accuracies, combining the samples from all galaxies at each distance. This allows us to take into account the asymmetry in the GravSphere confidence limits, as seen in Fig. 2.

The lower axis limit on R/R_e in Fig. 3 has been chosen to be the average value of $2.8\epsilon/R_e$ for our sample and the upper axis limit was chosen so as to contain the smallest dark matter halo in the sample. The bias and associated errors for each galaxy are only included in the making of this figure for distances $R > 2.8\epsilon$.

In purple, we show the CDM sample and in black the SIDM sample. The symbols with error bars show the performance of dynamical mass estimators from Walker et al. (2009), Wolf et al. (2010), and Campbell et al. (2017, circles for CDM and squares for SIDM). The Wolf et al. (2010) and Walker et al. (2009) estimators are accurate to better than 10 per cent, however we observe a bias in the Campbell et al. (2017) estimator. This is likely due to the aforementioned difference in the definition of the half-light radius. In all cases, the true mass is contained within the uncertainty of the estimators.

From the left-hand panel of Fig. 3 it is clear that, for CDM dwarfs, GravSphere performs just as well as the Wolf et al. (2010) and Walker et al. (2009) estimators, and with similar scatter. The inferred masses are, on average, very accurate across the entire radial range, with the scatter becoming more significant in the innermost regions as well as the outer regions. For the large values of R/R_e this is caused by the lack of kinematic tracers. Moreover, our priors permit only a narrow range of slopes ($0 < \gamma_j < 3$), whilst our sample is expected to have undergone tidal effects, resulting in an outer slope $\gamma \approx 4$ (Peñarrubia et al. 2009). Nevertheless, the estimate is accurate, on average, out to $3R_e$, with a root mean square (rms) fractional
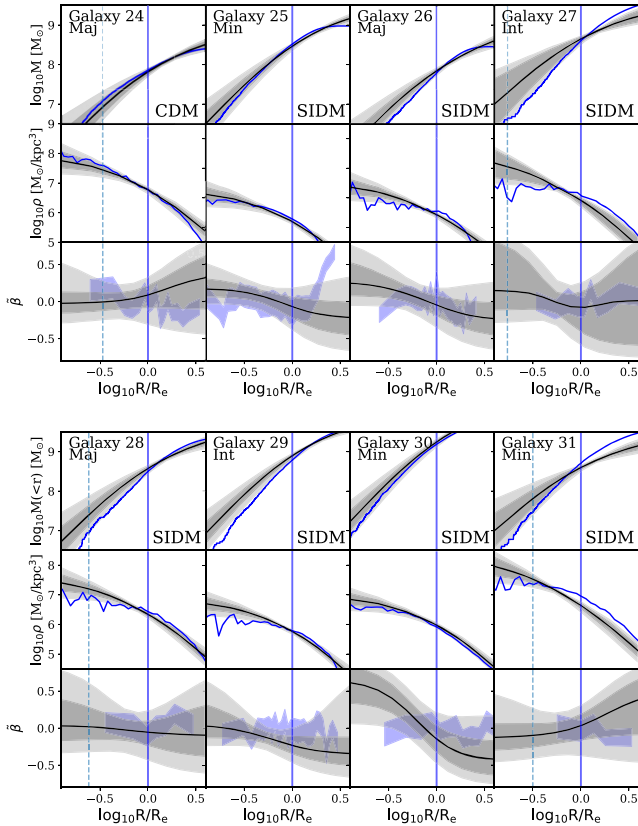
Figure 2 – *continued*

error[9] of only 5 per cent. It is clear that the masses for SIDM dwarfs are significantly overestimated in the inner regions (on average, up to 80 per cent), with very large scatter. The mass is accurate near the half-light radius and beyond that it is underestimated by ∼20 per cent.

3.3 The core-cusp problem

What do these results mean in the context of the core-cusp problem? We now consider two ways to infer cores or cusps in dark matter haloes: via their characteristic inner density (as in Read et al. 2019) and via an accurate inference of profile shape.

3.3.1 Cores versus cusps via characteristic densities

In the right-hand panel of Fig. 3, we show the recovery by GravSphere of the dark matter density profiles of our sample of dwarfs. For CDM dwarfs, the density profiles are accurate across the entire radial range (rms = 0.07) and the scatter is only ∼30 per cent in the inner regions. In fig. 5 of Read et al. (2019) it can be seen that, for a dwarf of Fornax-like pre-infall halo mass, the core and cusp-like densities, ρ_{150}, differ by a factor of at least 3.5. The spatial resolution of our simulations does not allow us to probe radii below 380 pc; however, if GravSphere provides a similar level of bias and

scatter for ρ_{150} (corresponding to $\log_{10} R/R_e \approx -0.7$ for a Fornax-size dwarf), it is certainly possible to differentiate between the core- and cusp-like densities, provided there is complete core formation below the half-light radius and no reduction of central dark matter density due to tides (Read et al. 2016). For SIDM dwarfs, the density is overestimated in the inner regions by up to 50 per cent and underestimated above the half-light radius by ∼30 per cent, reflecting the pattern with enclosed mass. This suggests that in the case of SIDM dwarfs with cores, GravSphere is biased towards cusps in its standard configuration.

3.3.2 Cores versus cusps via profile shape

Let us now approximate the density profiles in the inner regions by a single power law

$$\rho(r) = \rho(R_e) \left(\frac{r}{R_e} \right)^{-\gamma}, \tag{11}$$

where $\rho(R_e)$ is the density at R_e and γ is the slope of the power law, with $\gamma = 0$ corresponding to a core and $\gamma = 1$ to a cusp. If we assume that a core forms below the radius R_e, where otherwise the density within R_e follows $\rho \propto r^{-1}$, then the density ratio is

$$\frac{\rho_{\text{cusp}}}{\rho_{\text{core}}} = \left(\frac{r}{R_e} \right)^{-1}, \tag{12}$$

and the mass ratio is

$$\frac{M_{\text{cusp}}}{M_{\text{core}}} = \frac{3}{2} \left(\frac{r}{R_e} \right)^{-1}. \tag{13}$$

We display these relations, and their inverse, with dashed (an incorrectly inferred cusp) and solid (an incorrectly inferred core) grey lines in Fig. 3. The relations are displayed out to the radius $R = R_e$ for the case where a core forms below the half-light radius.[10] Beyond this radius, the relations are expected to converge to 1. The relations should be taken more as visual guides for the radial dependence of the bias that we expect in order to incorrectly infer a core or a cusp. Note that for a core that forms at some fraction of the half-light radius, we would only need to scale the relation in equation (12) by a corresponding factor (or shift the relation up and down in log-space). Unfortunately, cores on the scale of $0.5 R_e$ are too small to be probed by our simulations for the dwarf galaxy masses we consider here.

For the case of CDM dwarfs, it is clear that within the 68 per cent confidence regions the mass and density profiles returned by GravSphere are fully consistent with cusps. Cores lie outside the 95 per cent confidence regions. For SIDM dwarfs there is a clear bias towards more cuspy profiles. Fully cored profiles below the half-light radius are contained within the 68 per cent regions, but so are cusps that are only slightly shallower than $\rho \propto r^{-1}$.

3.4 Comparison to other methods and parametrizations

We now compare the performance of GravSphere to that of more conventional Jeans approaches. We will focus on the enclosed mass profiles, rather than densities. Masses are more robust than their differentials, they are simply measured in simulations and are a fundamental property in Jeans analysis.

---

[9]We define rms = $\sqrt{\frac{1}{N} \sum_{i=1}^{i=N} (M_{\text{calc}}/M_{\text{true}} - 1)^2}$, where M_{calc} is the mass obtained through GravSphere, or another method, N is the number of radial intervals at which M_{calc} is computed and M_{true} is the true enclosed mass at these intervals, found directly from the simulation. We use a maximum of 30 intervals to construct Fig. 3.

[10]We have verified that the relations in equations (12) and (13) are good approximations for a dwarf with a Fornax-like pre-infall halo mass (Read et al. 2019) with full core formation below the half-light radius and an NFW profile otherwise (Read et al. 2016).
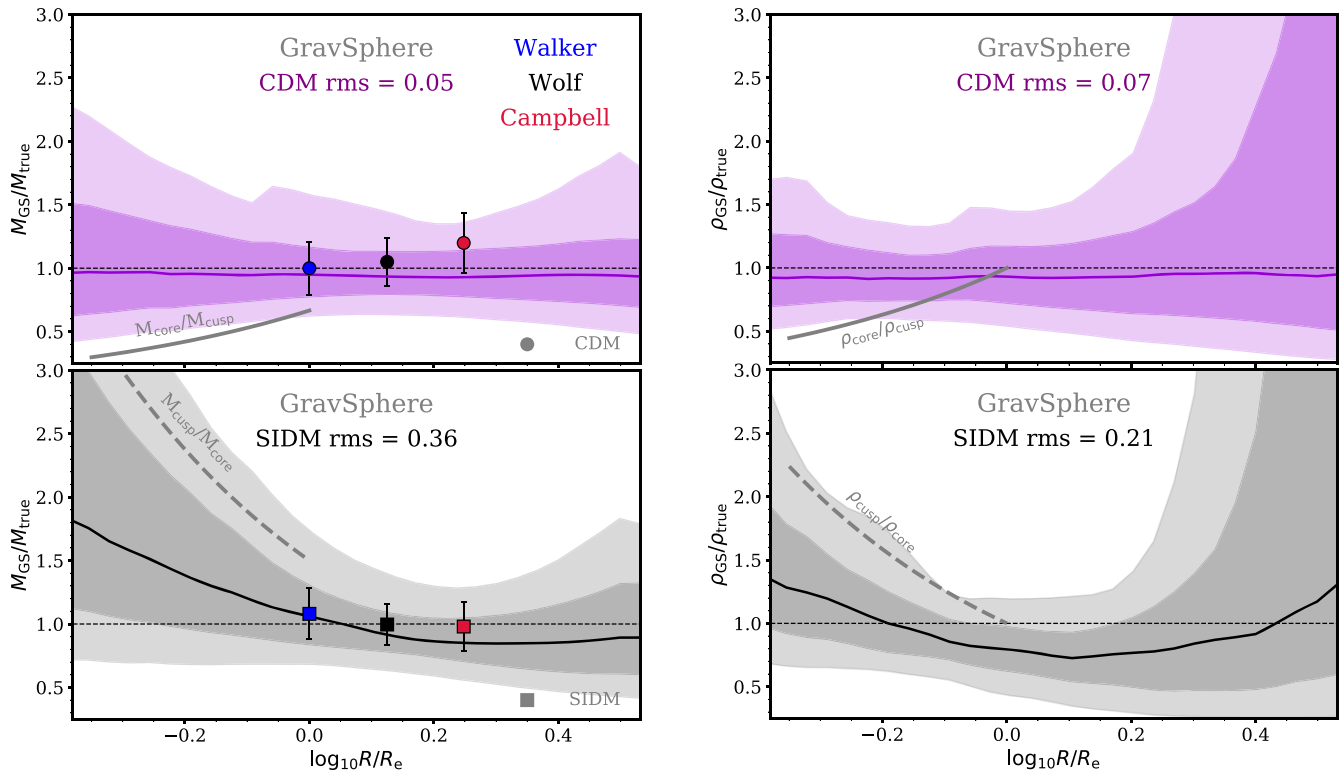
Figure 3. Left: ratio of the recovered to true enclosed mass as a function of radial distance in units of the half-light radius, R_e. The purple line and purple shaded bands show the median, 68 and 95 per cent confidence limits for the CDM galaxy sample. The black line and the corresponding shaded bands show the sample of SIDM dwarfs. The symbols and their error bars show the accuracy of standard estimators from Walker et al. (2009), Wolf et al. (2010), and Campbell et al. (2017), identified by their colours. Circles are for the CDM dwarfs and squares for SIDM. The 'rms' values show the rms offset of the median value from 1 along the entire radial range displayed. The grey dashed line shows the radially dependent bias in the mass profiles that is required to infer an NFW cusp when in reality there is a core. The solid grey line shows the bias required to infer a dark matter core when in reality there is a cusp. Right: ratio of the recovered to true density as a function of radial distance in units of the half-light radius. The grey dashed line shows the bias in the recovered density profiles required to infer an NFW cusp when in reality there is a core. The solid grey line shows the corresponding bias for an incorrect inference of a core when the true profiles are cuspy.

Table 2. Zhao (1996)+γ, r_s MCMC priors, and the cuts applied in post-processing.

Parameter	Prior	Constraint
$\log_{10}\rho_s/\mathrm{M}_\odot\mathrm{kpc}^{-3}$	[5, 13]	
$\log_{10}r_s/\mathrm{kpc}$	[−3, 1]	$\log_{10}r_s \geq \log_{10}R_e$
α	[0.5, 3]	
β	[3, 7]	
γ	[0, 1.5]	$\gamma \leq 1$

3.4.1 Comparison to Zhao (1996) profile

We compare the performance of GravSphere to the method outlined by Bonnivard et al. (2015a), who used the profile proposed by Zhao (1996) to parametrize the dark matter distribution:

$$\rho_{\rm dm} = \frac{\rho_0}{\left(\frac{r}{r_s}\right)^{-\gamma}\left(1+\left(\frac{r}{r_s}\right)^{\alpha}\right)^{\frac{\beta-\gamma}{\alpha}}} \qquad (14)$$

where r_s and ρ_0 are the scale radius and scale density, γ is the inner slope, β the outer slope, and α governs the steepness of transition between γ and β.

The priors for this run, which we refer to as Zhao + γ, r_s, are given in Table 2. The method requires two post-processing cuts: one on $\gamma \leq 1$, which was shown to reduce the overall scatter, and another

on $r_s \geq R_e$, which weeds out unphysical models from the fit. The results are shown in the upper left panel of Fig. 4.

It can be seen that this method underestimates the enclosed mass by ~10 per cent in the inner regions for CDM dwarfs and is unbiased in the outskirts. In SIDM dwarfs, however, the bias in the centre is less severe than with GravSphere. In the inner regions, for CDM and SIDM dwarfs, the scatter has not changed significantly, but outside the half-light radius and in the outer parts the width of the scatter increases by over 50 per cent. This is likely due to the priors (Table 2) allowing a much wider range of outer slopes β than permitted by the default GravSphere priors (Table 1). The exclusion of virial shape parameters likely also plays a role. We explore this below.

3.4.2 GravSphere, excluding the VSPs

What is gained by including the virial shape parameters? We repeat our GravSphere run, this time excluding the two VSPs from equation (9). The results are shown in the top right of Fig. 4. It can be seen that the accuracy in the inner regions has suffered from excluding the VSPs in CDM dwarfs. In SIDM, however, the bias is slightly reduced. As we shall see below, this reduction is consistent with excluding only the second virial shape parameter. The bottom panel shows the ratio of the upper and lower errors (84th and 16th percentiles) compared to the default GravSphere run (including the VSPs). It is evident that the scatter has increased beyond the half-light
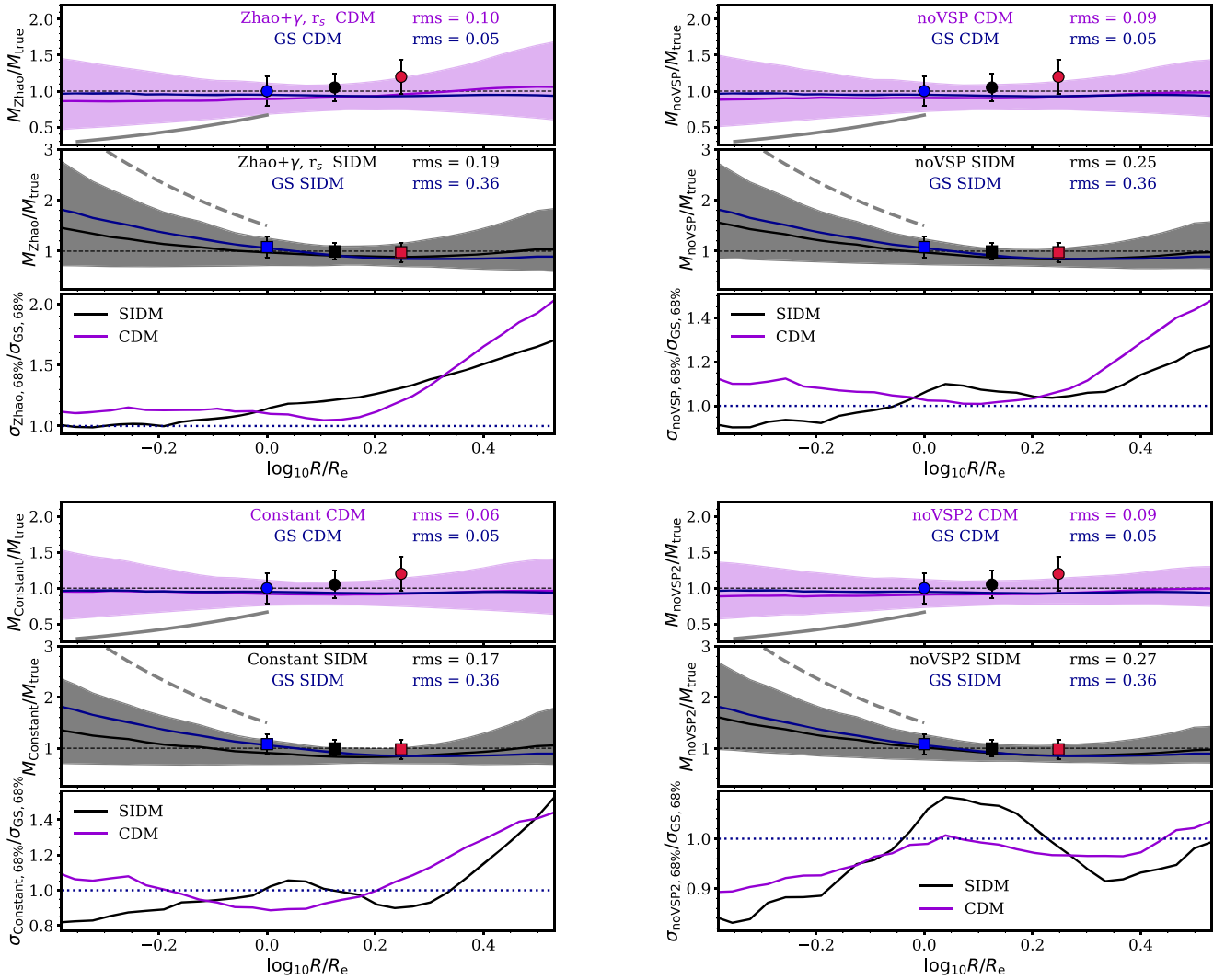
Figure 4. Top left: bias in enclosed mass profiles found assuming a Zhao (1996) profile and using priors as outlined in Table 2, together with post-processing cuts ($\gamma < 1$ and $r_s > R_e$). The median for the original GravSphere run is shown in blue. The medians for the CDM and SIDM samples are shown with purple (top subplot) and black (middle subplot) lines, respectively. The shaded bands of the same colour show the 68 per cent confidence limits on the bias, $M/M_{\rm true}$, for CDM (top) and SIDM (middle) samples. The bottom subplot shows the magnitude ratio of the 68 per cent confidence intervals between the Zhao (1996) and default GravSphere results. The colour symbols are the Walker et al. (2009, blue), Wolf et al. (2010, black) and Campbell et al. (2017, red) estimators. The grey dashed line corresponds to the bias required for an incorrect inference of an NFW cusp when in reality there is a core, and the solid grey line is the bias required for an inference of a core when in reality there is an NFW cusp. Top right: bias found with GravSphere assumptions, but excluding the VSPs (noVSP). Bottom left: bias found assuming constant anisotropy and no VSPs. Bottom right: bias found with the exclusion of the second virial shape parameter.

radius. This suggests that GravSphere runs that exclude the VSPs result in a wider range of allowed models. Evidently, the inclusion of VSPs plays a key role in minimizing the scatter in allowed anisotropy models, particularly in the outer regions. These results mimic those produced with the priors in Table 2, although the increase in scatter in the outer regions is smaller. This suggests that another significant source of this scatter is, in fact, the wider range of allowed slopes β.

3.4.3 Constant anisotropy and no VSPs

We now explore the performance of GravSphere under the assumption of constant anisotropy and no VSPs. From Fig. 2, it is clear that the vast majority of our simulated dwarfs have nearly constant stellar velocity anisotropy profiles. It is therefore possible that forcing the anisotropy profile to be constant with distance may encourage

the MCMC algorithm to select better models. The comparison with GravSphere is shown in the bottom left of Fig. 4. Similar accuracy to GravSphere is achieved (rms = 0.06) across the entire radial range. The errors for CDM dwarfs are similar to GravSphere in the innermost regions, but outside R_e the errors are larger. This could be partly due to lack of flexibility as compared to the Baes & van Hese (2007) profile, but this is also overall consistent with the noVSP run, suggesting a lack of constraint in the outer profile due to the exclusion of VSPs.

For SIDM dwarfs the bias has reduced remarkably (now below 20 per cent in the inner regions). The scatter has also reduced by ~25 per cent compared to GravSphere. This would suggest that anisotropy profiles are not very well recovered with GravSphere and are a significant source of bias for SIDM dwarfs. In Fig. 2, we can see that the true anisotropy profiles are generally enclosed within the
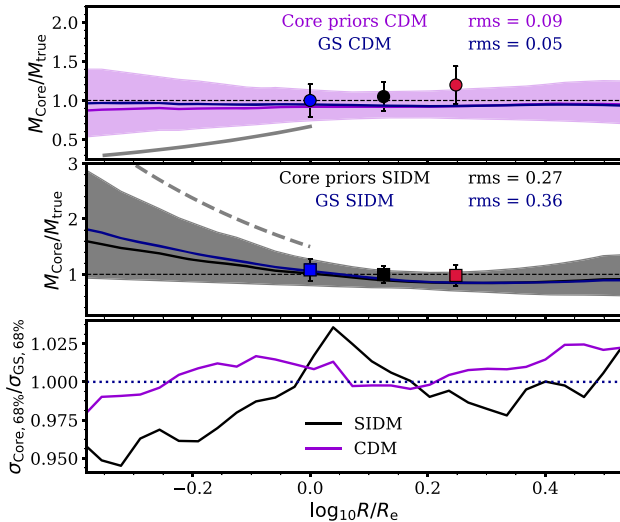
Figure 5. Bias found using wider priors on the slopes γ_j that allow a 'hole' in the central regions. All values of $\gamma_j < 0$ output by the MCMC are then equated to 0. The coloured symbols and grey lines are as in Fig. 4.

confidence limits. These limits are, however, quite large and so the $M - \beta$ degeneracy is not fully broken. The breaking of the degeneracy is forced when imposing a constant form for the anisotropy, resulting in the reduction in bias. Note also that the bias associated with a CDM cusp (dashed grey line in the bottom left of Fig. 4) is now far above the 68 per cent limits, compared to the case of GravSphere.

We note that, based on the sample of galaxies presented in this paper, our simulations suggest that dwarfs may have anisotropy profiles that are well described by a constant value. For this sample, the assumption of constant β is sufficient to accurately recover mass profiles in the innermost regions of dwarfs and this does not require the use of the VSPs.

3.4.4 Removing VSP2

In their recent work, Kaplinghat, Valli & Yu (2019) have opted to exclude the second virial shape parameter, which is extremely sensitive to the behaviour of the fourth velocity moment in the outer regions of the dwarfs (due to the R^3 weighting), where the velocity distribution is typically poorly constrained. In GravSphere, the uncertainty in the $\langle v_P^4 \rangle$ profile is encapsulated within the errors, though these errors are indeed very large. In this subsection, we investigate the effects of excluding the second virial shape parameter from our analysis.

The results can be seen in the lower right panel of Fig. 4. We can see that, for CDM dwarfs, the mass profiles are now underestimated in the inner regions by nearly 10 per cent, while the scatter is reduced. For SIDM dwarfs, the accuracy has marginally improved and the scatter has also reduced. We conclude that the inclusion of VSP2 in Jeans analysis tends to slightly increase the scatter (due to its large errors), but the accuracy of the mass profiles is improved at the expense of this scatter.

3.5 Bias towards cusps in SIDM haloes

We now return to the issue of the overestimation of the enclosed mass profiles in SIDM dwarfs. We consider the effects of possible

offsets between the centre of mass of the stars and the dark matter as well as the effect of our priors on the recovered mass profiles.

3.5.1 Galaxy–halo offsets in SIDM

For a profile with a core, it is particularly difficult to establish the location of the density centre. If the centre is offset from the true dynamical centre this may introduce a bias in the recovered mass profile. Although it is unlikely that this bias would always cause an overestimation of the true mass, we have investigated the differences between the centre of mass of the stars and the dark matter for each dwarf in our sample. We found that offsets are present, with a typical magnitude of 80 ± 35 pc for the CDM sample. For SIDM haloes, the offsets are particularly extreme, up to 1 kpc in size. We repeated our analysis, comparing the mass profiles recovered by GravSphere to those computed directly from the simulation, now centring the 'true' profiles at the centre of mass of the dark matter. We found no discernible differences from the results presented in Fig. 3. Centring on the centre of potential instead also did not change our results. We conclude that galaxy–halo offsets have no significant effect on the results presented in this work. This is because these offsets are rather small compared to the half-light radii of SIDM dwarfs.

3.5.2 Using priors that favour a core

The systematic overestimation of enclosed mass at small radii has been encountered previously in the work of Read et al. (2018). The priors in Table 1 may bias the results towards more cusp-like values when the data are not sufficiently constraining, as a core (with $\gamma = 0$) lies on the boundary of the allowed range of slopes (see Fig. A1). In Read et al. (2018), a different set of priors was introduced, which extends the parameter space of the dark matter density slopes γ_j such that EMCEE walkers are more likely to probe regions of space that are compatible with a core. This is achieved by allowing the slopes to range between $-2 < \gamma_j < 3$. Note that $\gamma_j < 0$ corresponds to a 'hole' in the central regions of the dwarf. While not implausible in principle, in this work we assume that, in these extreme cases, the dark matter distribution has a core. When computing the confidence limits of the enclosed mass and density distributions in each dwarf, we thus fix the γ_j position of any walkers venturing into the space where $\gamma_j < 0$ to $\gamma_j = 0$.

The results of this may be seen in Fig. 5. For SIDM dwarfs, we can see that the bias along the entire radial range has been reduced to $\sim$25 per cent and the scatter is reduced by $\sim$10 per cent in the inner regions. We also display the results for individual SIDM dwarfs in Fig. A3, where a reduction in bias compared to Fig. 2.4.3 is evident. For CDM dwarfs, the new priors slightly bias the recovered mass profiles towards cores, although the true profiles still lie within the 68 per cent confidence levels. This is consistent with the findings of Read et al. (2018).

We conclude that the lack of exploration of the models with $\gamma_j = 0$ by GravSphere is clearly an issue, however, the data still lack sufficient constraining power to prefer cores over cusps in SIDM.

3.6 Using all available stars

In this section, we briefly explore how much information is gained by including all stars within $2R_e$ of each dwarf (ranging from 10^3 to 10^4 stars for CDM), as opposed to a Fornax-like sample of $500-2500$ stars. We note, however, that for the majority of SIDM dwarfs the available samples are less than 1000 stellar particles.
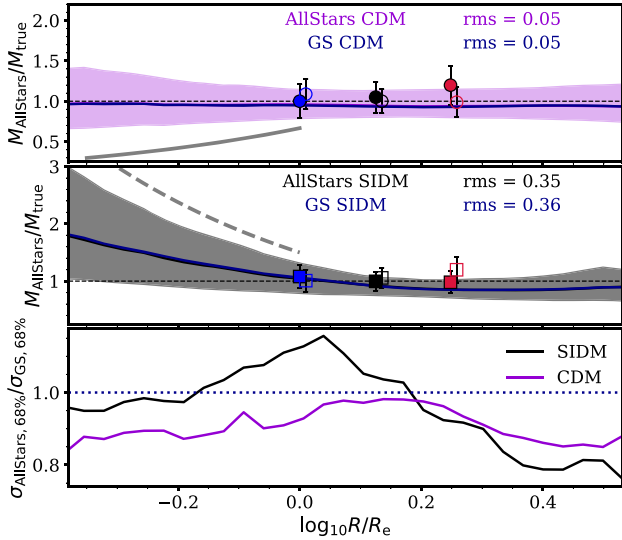
Figure 6. Bias found under standard GravSphere assumptions but using all available stars in each dwarf. The symbols are as in Fig. 4. The additional empty symbols representing the accuracy of mass estimators were evaluated using all available stars within $2R_e$.

The results of this can be seen in Fig. 6, where it is clear that the accuracy in the inner regions has slightly improved and the size of the errors is reduced by ∼10 per cent at smaller radii and closer to ∼20 per cent at large radii. Larger samples of data from future spectroscopic surveys will undoubtedly reduce the uncertainties associated with recovered dark matter mass profiles; however, this improvement is expected to be rather small when using line-of-sight data only. Moreover, GravSphere is able to achieve the same level of bias with present data samples. Further reduction in uncertainty is more likely to come from exploiting kinematically distinct stellar populations or proper motion data (Read & Steger 2017).

3.7 Sources of bias and scatter in GravSphere

We now explore possible origins of scatter in Fig. 3. Projection effects, asphericity, and tides are of particular interest (Genina et al. 2018).

3.7.1 Line-of-sight effects

How does asphericity affect the accuracy of the enclosed mass profiles recovered by GravSphere? In the top panel of Fig. 7, we split our sample of dwarfs in bins of minor-to-major axial ratio $s = c/a$. It can be seen that, on average, dwarfs of all asphericities in CDM have mass profiles recovered to better than 10 per cent; however it is also clear that the scatter is much larger for more aspherical objects ($s < 0.9$). Our sample has been chosen to contain galaxies projected along their three principal axes. We can therefore split our sample into three categories: dwarfs seen along the minor, major and intermediate axes, and examine the accuracy of GravSphere in recovering mass profiles in each case. For this, we select dwarfs with $s < 0.8$ in CDM (six galaxies) and $s < 0.85$ in SIDM (three galaxies).[11]

---

[11] We have considered projection effects for the more spherical dwarfs in the sample and found that they are irrelevant in those cases.
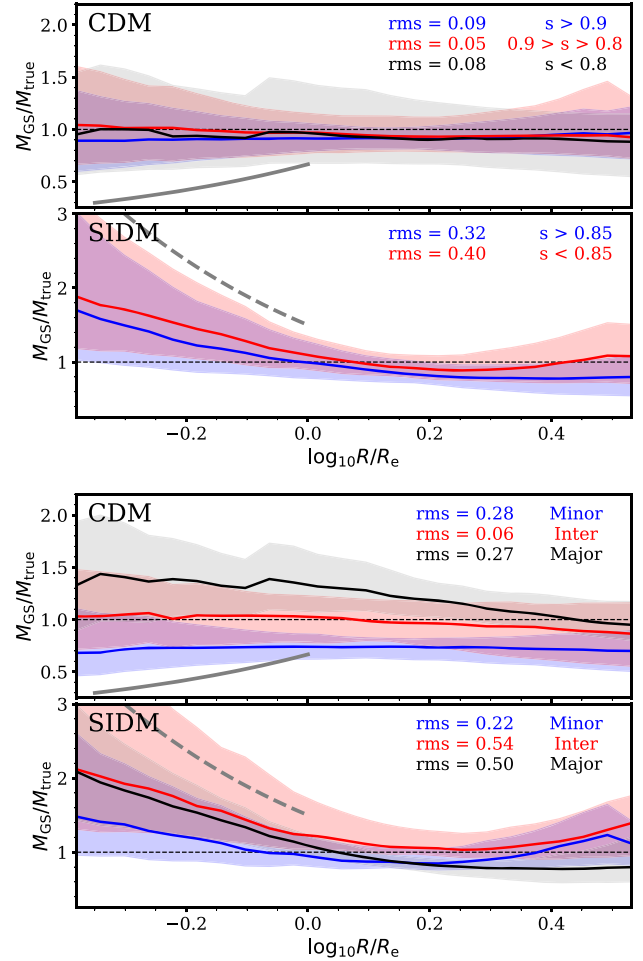
Figure 7. Top: bias in mass profiles returned by GravSphere for dwarfs grouped by their minor-to-major axial ratio c/a. The top panel shows the CDM sample and the bottom panel shows the SIDM sample. Bottom: line-of-sight effects on the accuracy of the mass profile recovery by GravSphere. The sample of galaxies is split into those viewed along the minor (blue line and bands), intermediate (red) and major (black) axis. Only dwarfs with $s < 0.8$ for CDM and $s < 0.85$ for SIDM are shown. The dotted black line in each plot represents unbiased results. The solid grey line corresponds to the bias expected in order to incorrectly infer a core within the half-light radius of a cuspy dark matter halo. The dashed grey line corresponds to the bias expected to incorrectly infer an NFW cusp, when in reality there is a core.

The results are shown in the bottom panel of Fig. 7. For CDM dwarfs, there is a clear distinction between mass profiles obtained when viewing along the minor (blue), intermediate (red) and major (black) axis. Along the intermediate axis, the mass profiles are generally unbiased, but they are underestimated along the minor axis and overestimated along the major axis. On average, this over- and underestimation is of magnitude ∼30 per cent. For prolate systems, this is consistent with the variation of line-of-sight velocity dispersion when viewing along the three principal axes.

The picture is somewhat different for SIDM dwarfs, where the mass profiles are overestimated more significantly when viewed along the intermediate axes. We note that this sample contains only three dwarfs and their three projections. We found that two of these have dark substructure present within their haloes, resulting in inflated velocity dispersion along the intermediate axis of these dwarfs.
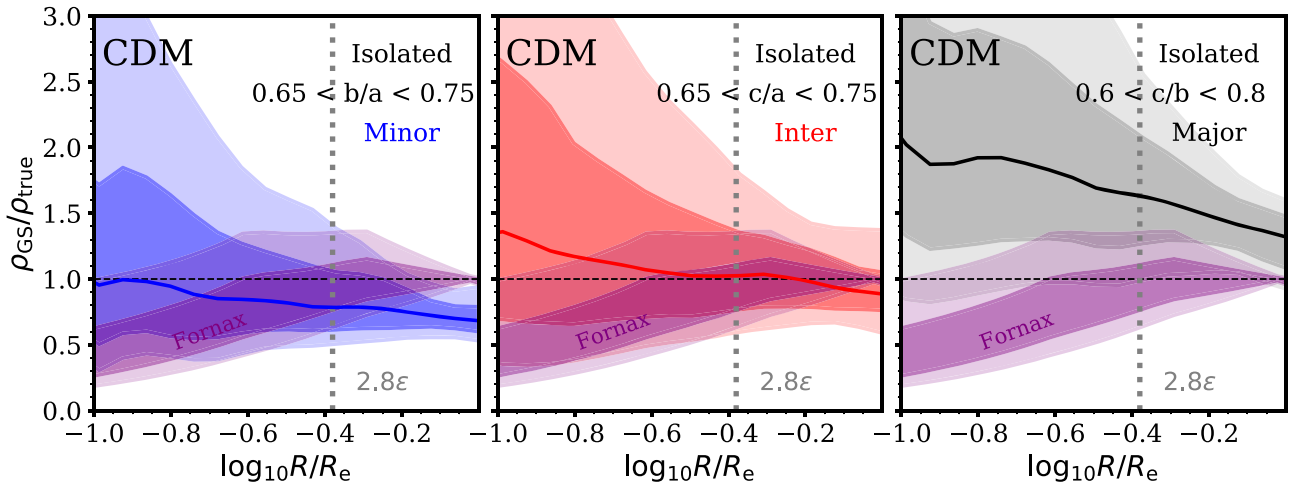
Figure 8. The bias in density profiles inferred by GravSphere for 24 isolated CDM dwarfs that match the criteria of having an on-the-sky ellipticity $e \sim 0.3$ when viewed along their minor (left, blue), intermediate (middle, red), and major (right, black) axes. The dark and light bands show the 68 and 95 per cent confidence levels, respectively. The dotted grey line shows the average value of $2.8\epsilon/R_e$. The purple line and bands show the inferred density profile for Fornax from Read et al. (2019), divided by a cuspy profile, $\rho \propto r^{-1}$, which is normalized to the recovered density of Fornax at R_e.

We have also considered the effects of dark matter halo asphericity. We found that the shapes of dark matter haloes are typically consistent with the stellar component, apart from a number of dwarfs with aspherical stellar distributions, for which the dark matter component was generally *less aspherical* than the stars. We conclude that the systematics are driven by the asphericity of the stellar component.

3.7.2 The core-cusp problem in Fornax

Results presented in Read et al. (2019) suggest that Fornax may have an inner core, based on the low inferred dark matter density and the inner slope, $\gamma_0 = 0.3^{+0.28}_{-0.21}$, below $0.25R_e$ (see also Goerdt et al. 2006; Cole et al. 2012; Pascale et al. 2019). Earlier, we have established that, for aspherical stellar distributions, line-of-sight effects may bias the recovered mass and density profiles. What do these results mean for the Fornax dwarf galaxy?

Fornax has a measured on-the-sky ellipticity $e \simeq 0.3$, matched by only a few satellites in our sample. We therefore focus instead on a sample of 24 isolated, dispersion-supported dwarfs from the APOS-TLE simulations in CDM. For consistency with Fornax, these dwarfs were chosen to have an ellipticity $e \sim 0.3$ when viewed along at least one of their principal axes. Otherwise, we marginalize over various shapes of the stellar and dark matter components. The axis ratios for these galaxies can be seen on the right panel of Fig. 1 (grey triangles).

Fig. 8 shows the bias in density profiles recovered with GravSphere for dwarfs viewed along the minor, intermediate, and major axes (and $e \sim 0.3$ in each case). The results are shown for radii between $0.1R_e$ and R_e, a key region where core formation would be apparent. Note that we have now gone below 2.8ϵ (grey dotted line), close to the softening length of our simulations. We must be wary of the discreteness noise contribution to the scatter at these radii. The purple bands show the 68 and 95 per cent confidence levels of the Fornax density profile recovered with GravSphere (Read et al. 2019), divided by the profile in equation (11), with ρ_0 equal to the density recovered by GravSphere for Fornax at R_e.

It is important to point out that, due to the mass resolution in APOSTLE, the innermost regions of our simulated dwarfs are insufficiently sampled, compared to the real Fornax data, which results in systematically larger errors in the recovered density profiles. Indeed,

the typical span of the 68 per cent confidence intervals below $0.25R_e$ in Fig. 2 is 0.5–0.6 in $\log_{10}\rho$, while for Fornax it is below 0.4, with the size of the errors being approximately the same for simulations and Fornax above 2.8ϵ. In order to compensate partially for this, in Fig. 8 we scale the errors for each dwarf by the ratio between the fractional errors (the span of the 68 per cent confidence intervals divided by the median) in a given APOSTLE dwarf and Fornax. This scaling, however, is not able to alleviate the large growth in uncertainty below $R = 2.8\epsilon$ for the mock data. Tackling this will require higher resolution simulations that are beyond the scope of this work.

A particularly striking result in Fig. 8 is that, above $0.1R_e$, the inferred density distribution in Fornax appears to be consistent within its 95 per cent confidence intervals with a cuspy dark matter halo viewed along any of the three principal axes. However, Read et al. (2019) disfavour this interpretation – at least in the context of a ΛCDM cosmology. This is because it requires an uncomfortably low pre-infall halo mass[12] for Fornax ($M_{200} \simeq 5 \times 10^9 \, \mathrm{M}_\odot$) as compared to expectations from abundance matching ($M_{200} \simeq 2 \times 10^{10} \, \mathrm{M}_\odot$; Read & Erkal 2019). It also requires a smaller halo concentration than is characteristic of such a halo mass in ΛCDM (Dutton & Macciò 2014). It remains to be seen whether a cuspy Fornax in a low pre-infall mass and concentration halo can self-consistently explain the relatively low density of Fornax in the context of ΛCDM cosmology. We will revisit this in future work, but note it here as a potential caveat to the conclusion of Read et al. (2019) that favours a core in Fornax in a more massive pre-infall halo.

Finally, Fig. 8 shows that the bias in GravSphere tends to remain fairly constant below the half-light radius. This is an encouraging result, since, in the context of the core-cusp problem, we must be particularly wary of any radially dependent systematics. If higher resolution simulations find similarly little radial dependence in the systematic errors down to $0.25R_e$ and below, then this would suggest that the dominant source of uncertainty, at present, in estimating the inner dark matter density profile of the Fornax dwarf lies in the random sampling error (i.e. the number of stars with kinematic data),

---

[12] M_{200} is the enclosed mass at r_{200}, the radius below which the mean density is 200 times the critical density of the Universe.
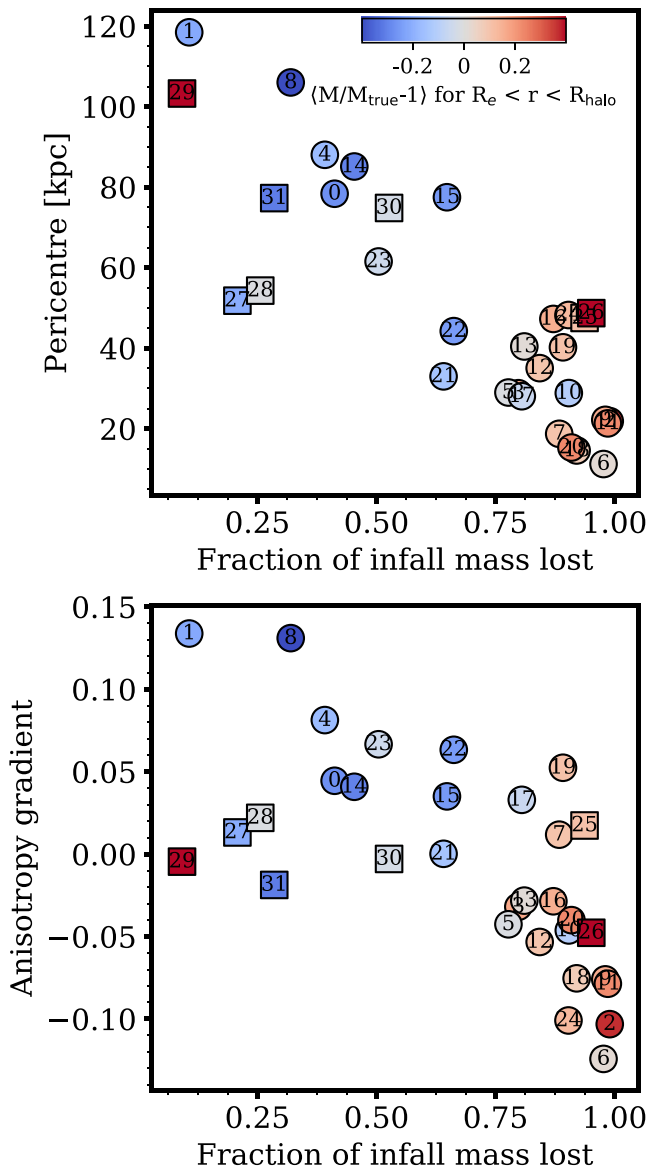
Figure 9. Top: the pericentre of the dwarf galaxy orbit as a function of the fraction of dark matter mass that has been lost since infall. The points are coloured by the mean bias of the recovered mass profile above the half-light radius and below the spatial extent of the dark matter halo. This is averaged for the three principal axes. Red indicates overestimation and blue indicates underestimation. Circles represent CDM dwarfs and squares the SIDM dwarfs. Bottom: the radial change in anisotropy, quantified by an anisotropy gradient measured below and above the half-light radius, as a function of the fraction of dark matter mass lost through stripping.

rather than the systematic error due to asphericity and the projection of Fornax on the sky.

3.7.3 The effect of tides

Dwarfs in our sample were selected to be satellites and these are susceptible to tidal stripping by their host galaxy. In this section, we explore whether the mass profiles recovered by GravSphere could be affected by tides.

In the top panel of Fig. 9, we plot the pericentre of the satellite orbits as a function of the dark matter mass lost since infall. We

find the pericentres by interpolating the position of each dwarf with respect to its host with a cubic spline. This method may underestimate the pericentres (see Richings et al. 2020); however, for all of our dwarfs, we see little difference between pericentres found through the linear and cubic splines. Moreover, the majority of the dwarfs have infall times of ~8 Gyr, such that typically 2–3 orbital periods are available for pericentre calculation, with snapshots having shorter temporal spacings at early times. We define the infall time as the snapshot at which the subhalo has its maximum dark matter mass. As expected, dwarfs with smaller pericentres tend to lose larger fractions of their mass.

In the bottom panel of Fig. 9, we show anisotropy gradients as a function of lost dark matter mass. We measure the anisotropy gradients between two points: the mean stellar particle radii below and above the projected half-light radius. A clear trend is evident, whereby tides cause more tangential anisotropy in the outer parts of galaxies. This is due to the preferential stripping of the stars moving on radial orbits (Henon 1970; Keenan & Innanen 1975; Kravtsov, Gnedin & Klypin 2004; Read et al. 2006; D'Onghia et al. 2010). We note the apparent simplicity of the two relations in Fig. 9, excluding perhaps the SIDM dwarfs, the orbits of which are more isotropic and the anisotropy is approximately constant.

We colour the points in Fig. 9 by the mean bias, $\langle M/M_{\text{true}} - 1 \rangle$, measured for all radii outside R_e and below the spatial extent of the halo. This is averaged for the three lines of sight along the principal axes of the dwarfs. We note that the bias is dominated by the line of sight closest to that which points from the centre of the host galaxy to the dwarf. This is indeed where we expect to see the largest effects on mass modelling due to the ongoing process of tidal stripping (Klimentowski et al. 2007).

It can be seen in Fig. 9 that for dwarfs that have undergone stronger tidal effects the mass profile is overestimated. This is unsurprising given that tides will result in the steepening of the outer slope (Peñarrubia et al. 2009), beyond the $0 < \gamma_j < 3$ permitted by our priors. Note, for example, the increased accuracy in the outer mass profile when using the Zhao (1996) dark matter parametrization (top left of Fig. 4), where such steep slopes are allowed. An exception from this trend is SIDM Galaxy 29, where the mass profile is overestimated in the outer parts, yet tidal effects seem to be less significant. None the less, the steepening of the density profile of this dwarf beyond $\gamma = 3$ is evident in Fig. 2.

We note the average underestimation of the outer mass profiles in dwarfs with weaker tidal effects. It is unclear how significant this result is, given that a number of these dwarfs (Galaxies 0,1,8,23) are amongst the most aspherical objects in our sample and thus are subject to other sources of bias.

Most importantly, we point out that we found no significant trend for the accuracy of the enclosed mass profile below R_e in our sample of dwarfs with tidal effects. This suggests that Jeans analysis is a valid method of mass modelling for dwarfs susceptible to tidal interactions, provided the impact of stars that are in the process of being tidally stripped is minimized. Here, we achieve this by limiting our sample of stellar particles to those within $2R_{2D}$, as well as only using the stellar particles considered as 'bound' by the subhalo finder. The study of the performance of GravSphere on a realistically contaminated sample of stars is certainly warranted and has been addressed, with a smaller sample of galaxies, by Read & Steger (2017).

3.8 Identifying failing models

Is it possible to identify the cases where GravSphere produces a biased result? In Read & Steger (2017), it was shown that for triaxial

haloes it is possible to tell that the model has been unsuccessful through the value of the χ^2. Is this true for our sample of dwarfs?

In Fig. 10, we display the mean number of standard deviations to the true mass profile from the profile recovered by GravSphere, computed within the half-light radius of each dwarf, as a function of the total normalized χ^2 value (i.e. we divide the χ^2 for the surface density profile and the line-of-sight velocity dispersion profile by the total number of photometric and kinematic bins, respectively). The points are coloured by the number of standard deviations from the 'truth', with red indicating an overestimate and blue indicating an underestimate. The grey shaded regions, highlighting the 1σ and 2σ intervals form the true mass profile, are labelled by the fraction of our galaxies falling within there regions. We thus confirm that in just under 60 per cent of the dwarfs in our sample, GravSphere returns the true mass profile within 1σ and in just under 90 per cent within 2σ, with the worst results encountered for SIDM dwarfs and the most aspherical objects in our sample.

Strictly speaking, we would expect to classify models with $\chi^2 > 4$ as poor fits. Indeed, we see that for the values of $\chi^2 \gtrsim 3.5$ (black dashed line) a stronger bias is observed. In general, there is a significant scatter the quality of results for dwarfs with $\chi^2 \gtrsim 3.5$. Among these galaxies, we see the particularly aspherical dwarfs and some SIDM galaxies (which are biased towards cusps), but not all. Evidently, this cut in χ^2 may be used for the selection of models that are more likely to be unbiased, such that the scatter in the accuracy of mass profiles returned by GravSphere is minimized.

4 CONCLUSIONS

Dwarf spheroidals are some of the best objects in which to study dark matter due to their proximity and high mass-to-light ratios. With the increasing availability of high-quality spectroscopic, photometric and proper motion data, studies of mass modelling methods and their limitations using realistic *N*-body simulations are certainly becoming more important in our efforts to narrow down the identity of dark matter and its behaviour on small scales. Here, we presented such a study for GravSphere, a higher order non-parametric Jeans analysis method (Read & Steger 2017).

First, we selected a sample of 32 dwarf galaxies from a suite of cosmological hydrodynamic simulations in ΛCDM and SIDM cosmologies. These simulated galaxies were chosen to resemble classical Local Group dwarfs-like Fornax. We then applied GravSphere, with its standard set of priors, to each of these dwarfs. We present the following findings:

(i) Within the key region inside the projected half-light radius, where dark matter cores form in some simulations (Navarro et al. 1996a; Pontzen & Governato 2012; Oñorbe et al. 2015; Tollet et al. 2016; Fitts et al. 2017; Read et al. 2016; Benítez-Llambay et al. 2019), the enclosed mass distributions are recovered within the 68 per cent confidence limits for ∼60 per cent of the dwarfs in our sample and within the 95 per cent confidence limits for ∼90 per cent of the dwarfs.

(ii) For our sample of CDM dwarfs, GravSphere returns unbiased mass profiles (rms = 0.05) along the radial range of (0.4–3)R_e, but with ∼50 per cent scatter in the innermost regions and ∼25 per cent scatter at the projected and deprojected half-light radii; this is comparable to standard mass estimators (see the left-hand panel of Fig. 3). In comparison to other Jeans methods, GravSphere achieves a more consistent performance across the radial range considered and, typically, has smaller scatter in the recovered mass profiles (see Fig. 4).
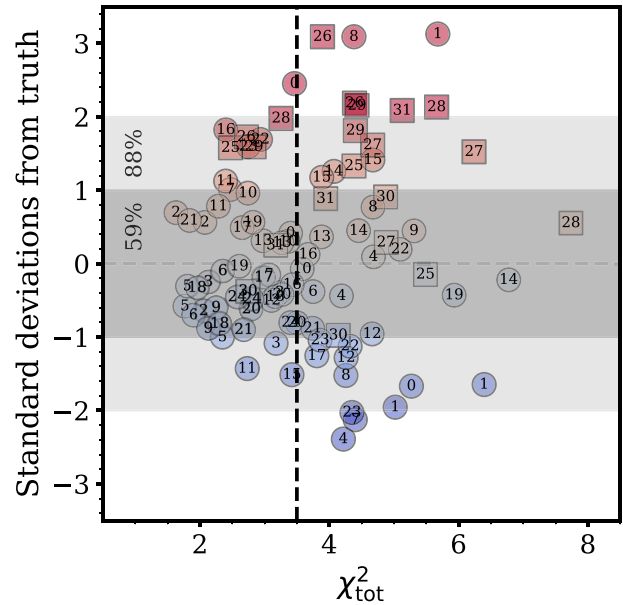

Figure 10. Number of standard deviations between the GravSphere result and the true mass profile, computed below the half-light radius and above 2.8ϵ, as a function of total, reduced χ^2. CDM dwarfs are shown with circles and SIDM dwarfs with squares. Red highlights an overestimation of true mass and blue an underestimation (*y*-axis values). The grey dashed line indicates unbiased results. The black dashed line shows our suggestion for a 'cut' in χ^2 to separate 'successful' and likely biased models. The shaded regions display the 1σ and 2σ regions and are labelled by the percentage of our sample, where the true mass profile is contained within these regions.

(iii) The density profiles for our CDM sample recovered by GravSphere are also accurate to better than 10 per cent (rms = 0.07) and exhibit a scatter of 30 per cent (see the right-hand panel of Fig. 3). Within the 68 per cent confidence levels, this is sufficient to reject cores that form on the scale of the half-light radius (when, in reality, there is a cusp). Due to the spatial and mass resolution of our simulations, we were only able to test GravSphere in regions outside 380 pc; however, if the uncertainty in the density profile does not increase for regions near 100 pc, it should be possible to separate core- and cusp-like densities for Fornax-like dwarfs that have undergone complete core formation on the scale of the half-light radius, as described in Read et al. (2016), provided the central density has not been reduced by tides and the pre-infall halo mass is well constrained.

(iv) For the sample of SIDM dwarfs with interaction cross-section of $\langle\sigma/m\rangle = 10\,\mathrm{cm}^2\mathrm{g}^{-1}$, which have cores on the scale of the half-light radius, we find that GravSphere is biased towards cuspy models. The density is overestimated by ∼20 per cent in the central regions. We show that this bias is relieved when imposing the correct form of anisotropy ($\beta = \beta_0$), suggesting that the data are not sufficiently constraining to break the $M - \beta$ degeneracy for these dwarfs (see the bottom left panel of Fig. 4). Moreover, our priors do not allow a full exploration of parameter space when there is a core in the central density distribution. Widening the priors to allow 'holes' in central regions leads GravSphere to a reduction in the bias for the SIDM sample (see Fig. 5), although at the expense of slightly biasing the CDM sample towards cores, with a 10 per cent underestimation in the inner regions.

(v) We explored the benefits gained by having larger stellar samples for our dwarfs (see Fig. 6). We found that using all available stellar particles within $2R_e$ mildly improves the accuracy of the mass profiles and reduces the scatter by ∼15 per cent for the CDM sample and ∼5 per cent for the SIDM sample (where we are limited in the number of available stellar particles due to mass resolution).

(vi) Our simulations suggest that Fornax-like dwarfs may have anisotropy profiles consistent with a constant value, $\beta = \beta_0$. In fact, for dwarfs in CDM, assuming a constant β and no VSPs results in a similar bias and scatter in the inner regions as GravSphere (with an increase in the size of the errors in the outer parts). For SIDM dwarfs, the assumption of constant β significantly reduces the bias and scatter at small radii compared to GravSphere, without the need to change the priors on the dark matter density slopes to allow for more models with cores (see the bottom left panel of Fig. 4).

We have explored the reasons for the scatter in GravSphere's performance. We found the following:

(i) The scatter in the accuracy of the recovered mass profiles is largest for objects that are particularly aspherical (see top panel of Fig. 7). In our CDM sample of dwarfs with the minor-to-major axial ratio, $c/a \sim 0.7$, galaxies viewed along their intermediate axis typically have their masses recovered accurately (rms = 0.05) compared to objects viewed along the minor axis (where the mass is underestimated by ∼30 per cent) or the major axis (where the mass is typically overestimated by ∼30 per cent). If Fornax has a sphericity of $c/a \simeq 0.7$ and is viewed along the intermediate axis, GravSphere is expected to accurately recover the mass (and the density) profile out to the half-light radius.

(ii) We have explored the effect of tides on the performance of GravSphere. We found no significant effect on the recovery of the mass profiles below the half-light radius. However, we did find that the mass profiles of systems which are more significantly affected by tides are typically overestimated in the outer regions (see Fig. 9), primarily due to the imposed priors on the outer density slope. This suggests that Jeans analysis is still valid for systems affected by tidal interactions, provided the impact of stars in the outer regions, which could be in the process of becoming unbound, is minimized.

(iii) We have investigated whether models which are biased (e.g. due to the underlying asphericity of the system) manifest themselves through poorer fits to the data. We found that these models typically have a higher total value of the χ^2 (see Fig. 10). We suggest a $\chi^2_{\rm tot} = 3.5$ cut to weed out biased models and reduce the uncertainty in the recovered profiles.

In conclusion, GravSphere is certainly a promising method for modelling dark matter distributions in dwarf galaxies. It remains to be seen whether it can maintain its lack of bias in the innermost regions of dwarfs (below $R_e/2$), which would be possible with higher resolution simulations. In this work, we focused on studying the effects of the violation of the assumptions of the spherical Jeans equation on GravSphere's performance. This study could, in the future, be extended with an inclusion of the effects of stellar binaries and contamination from the Galactic halo to provide a more realistic description of the expected systematics. This, together with modelling of multiple tracer populations and proper motions, has been explored on more idealized systems in Read & Steger (2017). We aim to extend this study with *N*-body and hydrodynamics simulations of realistic dwarf spheroidals in future work.

DATA AVAILABILITY

The data analysed in this article can be made available upon reasonable request to the corresponding author. PYGRAVSPHERE is open source software and is available on https://github.com/AnnaGenina/pyGravSphere.

REFERENCES

Adams J. J. et al., 2014, ApJ, 789, 63
Agnello A., Evans N. W., 2012, ApJ, 754, L39
Albert A. et al., 2017, ApJ, 834, 110
Amorisco N. C., Evans N. W., 2012, MNRAS, 419, 184
Baes M., van Hese E., 2007, A&A, 471, 419
Barber C., Starkenburg E., Navarro J. F., McConnachie A. W., 2015, MNRAS, 447, 1112
Battaglia G. et al., 2006, A&A, 459, 423
Battaglia G., Helmi A., Tolstoy E., Irwin M., Hill V., Jablonka P., 2008, ApJ, 681, L13
Battaglia G., Helmi A., Breddels M., 2013, New Astron. Rev., 57, 52
Benítez-Llambay A., Frenk C. S., Ludlow A. D., Navarro J. F., 2019, MNRAS, 488, 2387
Bett P., Eke V., Frenk C. S., Jenkins A., Helly J., Navarro J., 2007, MNRAS, 376, 215
Bonnivard V., Combet C., Maurin D., Walker M. G., 2015a, MNRAS, 446, 3002
Bonnivard V. et al., 2015b, MNRAS, 453, 849
Breddels M. A., Helmi A., 2013, A&A, 558, A35
Campbell D. J. R. et al., 2017, MNRAS, 469, 2335
Chabrier G., 2003, PASP, 115, 763
Charbonnier A. et al., 2011, MNRAS, 418, 1526
Cole D. R., Dehnen W., Read J. I., Wilkinson M. I., 2012, MNRAS, 426, 601
Contenta F. et al., 2018, MNRAS, 476, 3124
Crain R. A. et al., 2015, MNRAS, 450, 1937
D'Onghia E., Vogelsberger M., Faucher-Giguere C.-A., Hernquist L., 2010, ApJ, 725, 353
de Blok W. J. G., 2010, Adv. Astron., 2010, 789293
de Boer T. J. L. et al., 2012, A&A, 544, A73
del Pino A., Aparicio A., Hidalgo S. L., Łokas E. L., 2017, MNRAS, 465, 3708
Diakogiannis F. I., Lewis G. F., Ibata R. A., Guglielmo M., Kafle P. R., Wilkinson M. I., Power C., 2017, MNRAS, 470, 2034

Diakogiannis F. I., Lewis G. F., Ibata R. A., Guglielmo M., Wilkinson M. I., Power C., 2019, MNRAS, 482, 3356

Dolag K., Borgani S., Murante G., Springel V., 2009, MNRAS, 399, 497

Dutton A. A., Macciò A. V., 2014, MNRAS, 441, 3359

Elbert O. D., Bullock J. S., Garrison-Kimmel S., Rocha M., Oñorbe J., Peter A. H. G., 2015, MNRAS, 453, 29

Errani R., Peñarrubia J., Walker M. G., 2018, MNRAS, 481, 5073

Fattahi A. et al., 2016, MNRAS, 457, 844

Fitts A. et al., 2017, MNRAS, 471, 3547

Flores R. A., Primack J. R., 1994, ApJ, 427, L1

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306

Genina A. et al., 2018, MNRAS, 474, 1398

Goerdt T., Moore B., Read J. I., Stadel J., Zemp M., 2006, MNRAS, 368, 1073

González-Samaniego A., Bullock J. S., Boylan-Kolchin M., Fitts A., Elbert O. D., Hopkins P. F., Kereš D., Faucher-Giguère C.-A., 2017, MNRAS, 472, 4786

Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759

Henon M., 1970, A&A, 9, 24

Jardel J. R., Gebhardt K., 2012, ApJ, 746, 89

Kaplinghat M., Valli M., Yu H.-B., 2019, MNRAS, 490, 231

Keenan D. W., Innanen K. A., 1975, AJ, 80, 290

Klimentowski J., Łokas E. L., Kazantzidis S., Prada F., Mayer L., Mamon G. A., 2007, MNRAS, 378, 353

Kowalczyk K., Łokas E. L., Valluri M., 2017, MNRAS, 470, 3959

Kowalczyk K., Łokas E. L., Valluri M., 2018, MNRAS, 476, 2918

Kowalczyk K., del Pino A., Łokas E. L., Valluri M., 2019, MNRAS, 482, 5241

Kravtsov A. V., Gnedin O. Y., Klypin A. A., 2004, ApJ, 609, 482

Lake G., 1990, ApJ, 356, L43

Łokas E. L., 2002, MNRAS, 333, 697

Łokas E. L., 2009, MNRAS, 394, L102

Łokas E. L., Mamon G. A., 2003, MNRAS, 343, 401

Łokas E. L., Mamon G. A., Prada F., 2005, MNRAS, 363, 918

Lovell M. R., Hellwing W., Ludlow A., Zavala J., Robertson A., Fattahi A., Frenk C. S., Hardwick J., 2020, preprint (arXiv:e-prints)

Ludlow A. D., Schaye J., Schaller M., Richings J., 2019a, MNRAS, 488, L123

Ludlow A. D., Schaye J., Bower R., 2019b, MNRAS, 488, 3663

Ludlow A. D., Schaye J., Schaller M., Bower R., 2020, MNRAS, 493, 2926

McConnachie A. W., 2012, AJ, 144, 4

Mamon G. A., Biviano A., Boué G., 2013, MNRAS, 429, 3079

Merrifield M. R., Kent S. M., 1990, AJ, 99, 1548

Merritt D., 1987, ApJ, 313, 121

Moore B., 1994, Nature, 370, 629

Morselli A., Consortium C., 2017, Proceedings of the 35th International Cosmic Ray Conference (ICRC2017), Proc. Sci., Vol. 301, p. 921

Navarro J. F., Eke V. R., Frenk C. S., 1996a, MNRAS, 283, L72

Navarro J. F., Frenk C. S., White S. D. M., 1996b, ApJ, 462, 563

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

Newville M., Stensitzki T., Allen D. B., Ingargiola A., 2014, LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python (Version 0.8.0), Zenodo

Oh S.-H. et al., 2015, AJ, 149, 180

Oh S.-H., de Blok W. J. G., Walter F., Brinks E., Kennicutt Robert C. J., 2008, AJ, 136, 2761

Oman K. A., Marasco A., Navarro J. F., Frenk C. S., Schaye J., Benítez-Llambay A., 2019, MNRAS, 482, 821

Oñorbe J., Boylan-Kolchin M., Bullock J. S., Hopkins P. F., Kereš D., Faucher-Giguère C.-A., Quataert E., Murray N., 2015, MNRAS, 454, 2092

Orkney M. D. A., Read J. I., Petts J. A., Gieles M., 2019, MNRAS, 488, 2977

Pascale R., Binney J., Nipoti C., Posti L., 2019, MNRAS, 488, 2423

Peñarrubia J., Navarro J. F., McConnachie A. W., Martin N. F., 2009, ApJ, 698, 222

Planck Collaboration VI, 2018, preprint (arXiv:1807.06209)

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

Pontzen A., Governato F., 2012, MNRAS, 421, 3464

Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T., 2003, MNRAS, 338, 14

Read J. I., Erkal D., 2019, MNRAS, 487, 5799

Read J. I., Gilmore G., 2005, MNRAS, 356, 107

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

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

Read J. I., Agertz O., Collins M. L. M., 2016, MNRAS, 459, 2573

Read J. I., Iorio G., Agertz O., Fraternali F., 2017, MNRAS, 467, 2019

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

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

Richardson T., Fairbairn M., 2013, MNRAS, 432, 3361

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

Richings J. et al., 2020, MNRAS, 492, 5780

Robertson A., Massey R., Eke V., 2017, MNRAS, 465, 569

Robertson A. et al., 2018, MNRAS, 476, L20

Santos-Santos I. M. E. et al., 2020, MNRAS, 495, 58

Sawala T. et al., 2016, MNRAS, 457, 1931

Schaye J. et al., 2015, MNRAS, 446, 521

Schwarzschild M., 1979, ApJ, 232, 236

Spergel D. N., Steinhardt P. J., 2000, Phys. Rev. Lett., 84, 3760

Springel V., 2005, MNRAS, 364, 1105

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726

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

Strigari L. E., Bullock J. S., Kaplinghat M., Simon J. D., Geha M., Willman B., Walker M. G., 2008, Nature, 454, 1096

Strigari L. E., Frenk C. S., White S. D. M., 2010, MNRAS, 408, 2364

Strigari L. E., Frenk C. S., White S. D. M., 2018, ApJ, 860, 56

Tollet E. et al., 2016, MNRAS, 456, 3542

Ural U., Wilkinson M. I., Read J. I., Walker M. G., 2015, Nat. Commun., 6, 7599

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

Walker M. G., Mateo M., Olszewski E. W., Bernstein R., Wang X., Woodroofe M., 2006, AJ, 131, 2114

Walker M. G., Mateo M., Olszewski E. W., Peñarrubia J., Wyn Evans N., Gilmore G., 2009, ApJ, 704, 1274

Wolf J., Martinez G. D., Bullock J. S., Kaplinghat M., Geha M., Muñoz R. R., Simon J. D., Avedo F. F., 2010, MNRAS, 406, 1220

Zhao H., 1996, MNRAS, 278, 488

Zhu Q., Marinacci F., Maji M., Li Y., Springel V., Hernquist L., 2016a, MNRAS, 458, 1559

Zhu L., van de Ven G., Watkins L. L., Posti L., 2016b, MNRAS, 463, 1117

APPENDIX A: CONVERGENCE CRITERIA AND GENERATING INITIAL POSITIONS

In this section, we describe the effect of our choice of the initial positions of EMCEE walkers on the convergence of GravSphere's results.

A1 Effective priors

As mentioned in the main text, we generate the initial positions through the selection of walkers that satisfy the condition of radial increase in power-law slopes, γ_j, and the constraint on smoothness, $\Delta\gamma = 1$.

The top panel of Fig. A1 shows the effective priors on each slope. It is clear that these priors are not uniform as the selection of the slopes is not independent; however, the width of these distributions allows for a variety of density profiles. The inset shows the priors for the density at 150 pc, ρ_{150}. Despite the non-uniform nature of the γ_0 prior, when combined with a uniform prior on the scale density,
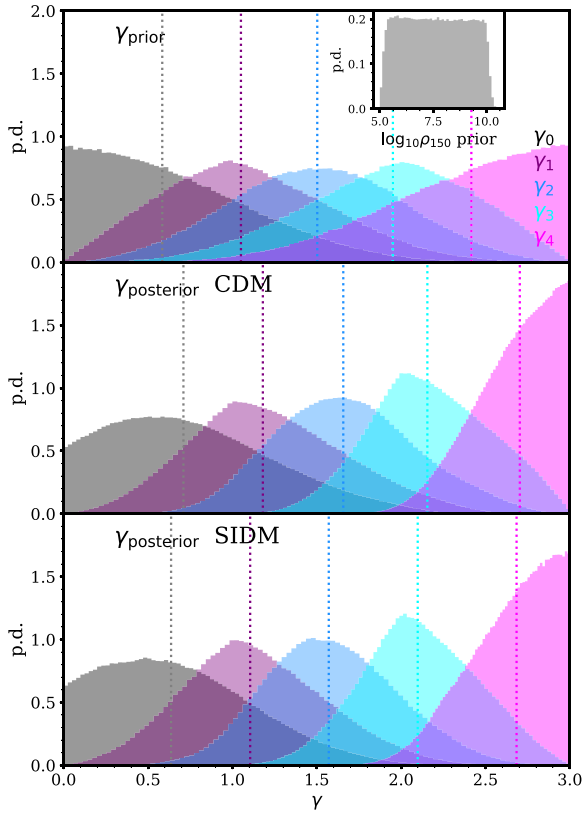
Figure A1. Top: effective priors on the logarithmic slopes γ_j. Each slope is identified by its colour. The inset shows priors on the density at 150 pc, assuming a 1 kpc half-light radius. Middle: posterior distributions, collected from all CDM dwarfs within our sample, weighted equally. The dotted lines are the medians of the distributions. Bottom: posterior distributions, collected from all SIDM dwarfs within our sample, weighted equally. The dotted lines are the medians of the distributions.

ρ_0, the resulting ρ_{150} prior is effectively uniform and is not biased towards more core or cusp-like values (Read et al. 2019).

The middle panel of Fig. A1 suggests that the posteriors on γ are not completely determined by the priors. The prior and posterior distributions are offset, as seen from their median values, and their shapes are noticeably different. The γ_4 posterior is clearly pushing against the prior boundary, suggesting that a wider prior on this parameter is desirable.

In the bottom panel, we see similar posterior distributions for SIDM dwarfs. It is clear that the regions with $\gamma_0 = 0$ are not prioritized by the EMCEE walkers, resulting in an inference of more cuspy dark matter density profiles. It can be seen, however, that lower values of γ_j are preferred, compared to the CDM sample.

A2 Walker convergence compared to previous implementations

Previous implementations of GravSphere have used initial positions of the walkers for the broken power-law slopes γ_j that are completely uniform. This results in a large number of EMCEE walkers starting off
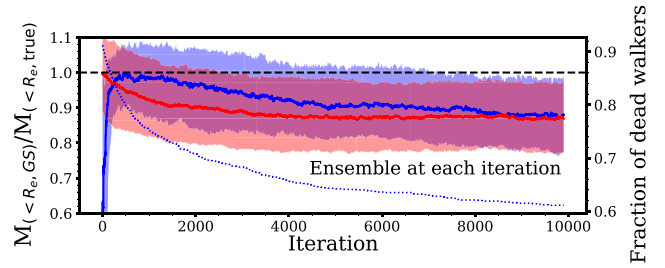

Figure A2. The convergence of mass within the half-light radius, expressed as a ratio of this mass to the true value, for two different prior selection methods. Priors that are generated completely uniformly are shown in solid blue, and those selected in a uniform, yet conditional, fashion (such that all priors satisfy the monotonic increase in the values of γ_j) are shown in red. The shaded bands represent 68 per cent confidence limits. The blue dotted line shows the fraction of 'dead walkers' (those stuck in the infinitely negative log-likelihood space), when using fully uniform priors.

in regions of infinitely negative log-likelihood. This is because these walkers do not satisfy the constraints for monotonically increasing values of γ_j. We will refer to these as 'dead walkers'. Eventually, some of these climb out and explore the posterior distribution, but not all, and this can take many iterations. In Read & Steger (2017), the chains were run for 5000 iterations, with the last 2500 used for analysis. In Fig. A2, we compare this method to one employed in this work.

We pick Galaxy 4 as our representative example and we select the mass within the half-light radius, $M(< R_e)$, as a quantity for which we wish to establish convergence. Fig. A2 shows the median value of the bias, $M(< R_e)/M_{\rm true}$, and the 68 per cent confidence levels for *each* walker iteration using the original GravSphere's method for generating initial positions (blue) and the initial positions generated using the method described in this work (red). The blue dotted line shows the fraction of dead walkers remaining after each iteration when using the original GravSphere method (right vertical axis).

It can be seen that our new method reaches convergence after $\sim 5 \times 10^3$ iterations, whereas the original method requires ~ 500 iterations to get out of the low log-likelihood regions and, in fact, does not reach the converged distribution until after $\sim 9 \times 10^3$ MCMC iterations. The chains start off with over 90 per cent dead walkers. This percentage drops to ~ 60 per cent near 10^4 iterations and can be seen to decrease slowly. We conclude that our new method of initial position selection allows for faster chain convergence and efficient walker exploitation.

A3 Extended priors that favour cores

In Fig. A3, we display the mass, density, and anisotropy profile recovery with GravSphere for our sample of SIDM dwarfs when the priors on the power-law density slopes γ_j are allowed to vary between $\gamma_j = [-2, 3]$, allowing 'holes' in the density distribution. In the post-processing, we fix all slopes $\gamma_j < 0$ to $\gamma_j = 0$. This effectively increases the sampling by EMCEE walkers of the parameter space regions where the dark matter distribution is cored.
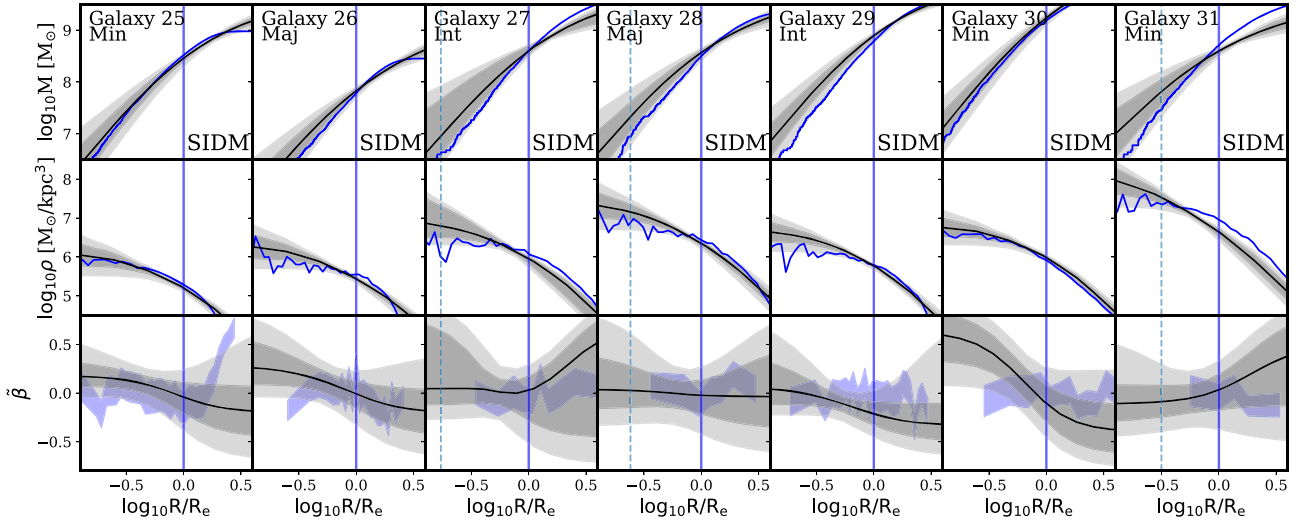
Figure A3. As Fig. 2c, but now using an extended set of priors on the density slopes γ_j, allowing 'holes' in central regions of dwarfs. Note that for a number of these dwarfs the spatial resolution, 2.8ϵ (vertical dashed line), is below $0.125R_e$ and below the limits of the figure.

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