




SMART: a new implementation of Schwarzschild’s Orbit Superposition technique for triaxial galaxies and its application to an N -body merger simulation

B. Neureiter ^{1,★}, J. Thomas^{1,2}, R. Saglia^{1,2}, R. Bender^{1,2}, F. Finozzi, A. Krukau^{3,4}, T. Naab⁵,
A. Rantala ⁵ and M. Frigo ^{5,6}

¹Max-Planck-Institut für Extraterrestrische Physik, Giessenbach-Str. 1, D-85748 Garching, Germany

²Universitäts-Sternwarte München, Scheinerstrasse 1, D-81679 München, Germany

³Leibniz-Rechenzentrum (LRZ), Boltzmannstrasse 1, D-85748 Garching, Germany

⁴Excellence Cluster Universe, Boltzmannstrasse 2r, D-85748 Garching, Germany

⁵Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85748 Garching, Germany

⁶Excellence Cluster ORIGINS, Boltzmannstrasse 2, D-85748 Garching, Germany

Accepted 2020 September 16. Received 2020 September 15; in original form 2020 April 11

ABSTRACT

We present SMART, a new 3D implementation of the Schwarzschild Method and its application to a triaxial N -body merger simulation. SMART fits full line-of-sight velocity distributions to determine the viewing angles, black hole, stellar and dark matter (DM) masses, and the stellar orbit distribution of galaxies. Our model uses a 5D orbital starting space to ensure a representative set of stellar trajectories adaptable to the integrals-of-motion space and it is designed to deal with non-parametric stellar and DM densities. SMART’s efficiency is demonstrated by application to a realistic N -body merger simulation including supermassive black holes that we model from five different projections. When providing the true viewing angles, 3D stellar luminosity profile and normalized DM halo, we can (i) reproduce the intrinsic velocity moments and anisotropy profile with a precision of ~ 1 per cent and (ii) recover the black hole mass, stellar mass-to-light ratio and DM normalization to better than a few per cent accuracy. This precision is smaller than the currently discussed differences between initial-stellar-mass functions and scatter in black hole scaling relations. Further tests with toy models suggest that the recovery of the anisotropy in triaxial galaxies is almost unique when the potential is known and full LOSVDs are fitted. We show that orbit models even allow the reconstruction of full intrinsic velocity distributions, which contain more information than the classical anisotropy parameter. Surprisingly, the orbit library for the analysed N -body simulation’s gravitational potential contains orbits with net rotation around the intermediate axis that is stable over some Gyrs.

Key words: methods: numerical – galaxies: elliptical and lenticular, cD – galaxies: evolution – galaxies: structure – galaxies: supermassive black holes – stars: kinematics and dynamics.

1 INTRODUCTION

Early-type galaxies (ETGs) have long been believed to emerge from collisions between other smaller progenitor galaxies (first proposed by Toomre & Toomre 1972), but nowadays it is clear that their formation history is more complex (e.g. Oser et al. 2010). Their structural and kinematic properties divide them into (1) fainter (absolute magnitude $M_B > -20.5$) and coreless fast rotators that are nearly axisymmetric and have discy-distorted isophotes and (2) brighter and more massive slow rotators with flat cores, which are moderately triaxial and have boxy-distorted isophotes (Faber et al. 1987; Bender 1988a; Bender et al. 1989; Kormendy & Bender 1996; Cappellari et al. 2007; Emsellem et al. 2007). For the formation of fainter elliptical galaxies, dissipational processes are believed to be important (e.g. Bender, Burstein & Faber 1992; Barnes & Hernquist

1996; Genzel et al. 2001; Tacconi et al. 2005; Cappellari et al. 2007; Hopkins, Cox & Hernquist 2008; Johansson, Naab & Burkert 2009), whereas the latest evolutionary phases in the formation of massive ellipticals are dominated by collisionless processes (e.g. Naab, Jesseit & Burkert 2006; Cappellari 2016; Naab & Ostriker 2017; Moster, Naab & White 2019).

In general, these merging events modify the potential structure and populate a rich diversity of stellar orbits (Röttgers, Naab & Oser 2014). The intrinsic shape and orbital structure of such galaxies are not directly observable. Instead, sophisticated dynamical models are needed to process kinematic and photometric observational data to extract all the information about the orbital structure and internal composition of the galaxy.

Dynamical models are based on the collisionless Boltzmann equation that governs the motion of stars in elliptical galaxies. Dynamical models that go beyond the recovery of velocity moments and aim at reconstructing the entire galaxy structure, additionally take advantage of the Jeans theorem (e.g. Binney & Tremaine

* E-mail: bneu@mpe.mpg.de

2008). This implies that the distribution function, which is the most general description of a system of stars, is constant along individual trajectories in phase space. In this regard, Schwarzschild (1979) pioneered an orbit superposition technique, where the equations of motion are numerically integrated for a finite number of stellar trajectories embedded in an assumed gravitational potential with contributions from the stars and possibly dark components. The weighted superposition of the orbits is determined for which the surface brightness and projected velocity distributions of the model match the observed ones in a least squares sense (e.g. Richstone & Tremaine 1984). Besides the orbital weights, all unknown quantities like the central black hole mass and dark matter distribution are varied between different models. The model producing the best fit to the projected velocity distributions is then associated with the correct model parameters. Any galaxy in a steady state can be modelled by Schwarzschild’s orbit superposition technique. In order to determine both, the mass and internal motions of the stars, and solve an underlying mass-anisotropy-entanglement, one needs to describe the deviation of the observed absorption lines from a Gaussian profile by additional Gauss–Hermite functions of at least third and fourth order (Gerhard 1993; van der Marel & Franx 1993; Bender, Saglia & Gerhard 1994), or, preferably if the signal-to-noise ratio permits, measure the line shape non-parametrically (e.g. Mehrgan et al. 2019).

Early applications of Schwarzschild’s orbit superposition technique concentrated on spherical models (e.g. Richstone & Tremaine 1985; Rix et al. 1997). Since the simplified assumption of spherical symmetry is not true for most galaxies, later applications of Schwarzschild’s orbit superposition technique assumed axisymmetry (e.g. van der Marel et al. 1998; Cretton et al. 1999; Gebhardt et al. 2000; Thomas et al. 2004; Valluri, Merritt & Emsellem 2004).

However, it is nowadays known that the most massive galaxies are neither spherical nor axisymmetric but triaxial objects. Observational indications are provided by isophotal twists in the surface brightness distribution, velocity anisotropy, minor axis rotation, kinematically decoupled cores, and the statistical distribution of the ellipticity of the isophotes (Illingworth 1977; Bertola & Galletta 1978, 1979; Schechter & Gunn 1978; Williams & Schwarzschild 1979; Bender 1988b; Franx & Illingworth 1988; Vincent & Ryden 2005). Schwarzschild (1979) proved the existence of self-consistent triaxial stellar systems in dynamical equilibrium with numerical orbit superposition models. Also, N -body simulations supported the idea of triaxial ellipsoidal stellar bulges and dark matter haloes (e.g. Aarseth & Binney 1978; Hohl & Zang 1979; Miller & Smith 1979; Barnes 1992; Jing & Suto 2002; Naab & Burkert 2003; Bailin & Steinmetz 2005).

High-mass galaxies are of particular interest in several astrophysical aspects, e.g. the stellar initial mass function (IMF) is discussed to vary among galaxies, with the largest excess stellar mass compared to the locally measured Kroupa (Kroupa 2001) or Chabrier (Chabrier 2003) IMF occurring in the most massive galaxies (e.g. Treu et al. 2010; van Dokkum & Conroy 2010; Thomas et al. 2011; Cappellari et al. 2012; Posacki et al. 2015; Smith, Lucey & Conroy 2015; Vazdekis et al. 2015; Parikh et al. 2018). Moreover, different growth models for supermassive black holes (SMBHs) predict different amounts of intrinsic scatter at the high-mass end of SMBH scaling relations (e.g. Peng 2007; Jahnke & Macciò 2011; Somerville & Davé 2015; Naab & Ostriker 2017). In order to address these questions, precision dynamical mass measurements of the stars and SMBHs are required and these are directly linked to triaxial modelling to avoid artificial scatter introduced by wrong symmetry assumptions. Thomas et al. (2007) showed that the stellar mass-to-light ratio (and, thus, indirect inferences about the IMF) can be biased by up to

50 per cent in extreme cases when using axisymmetric models for a maximally triaxial galaxy. Moreover, a wrongly assumed mass-to-light ratio influences the determination of the mass of the central black hole in the model. The work by van den Bosch & de Zeeuw (2010) suggests that the assumption of axisymmetry may bias black hole measurements in massive ellipticals. They find that the best-fitting black hole mass estimate doubles when modelling NGC 3379 with their triaxial code (van den Bosch et al. 2008) in comparison to axisymmetric models. Triaxial dynamical modelling routines are therefore required to recover unbiased stellar mass-to-light ratios and black hole masses with the best possible accuracy.

To understand the uncertainties and ambiguities of triaxial modelling one has to understand the following three essentially different effects:

- 1) The intrinsic uncertainty of the applied dynamical modelling algorithm, which can only be tested under circumstances where the solution is designed to be unique. This is one of the aspects covered in this paper.
- 2) The uncertainty in the reconstruction of orbital and mass parameters from typical observational data given the right deprojection, which is also addressed in this paper.
- 3) The uncertainty of the deprojection routine. This topic is covered in de Nicola et al. (2020).

All previously described effects need to be combined to evaluate the uncertainties in the whole modelling process. This will be investigated in a future paper.

So far, there exist two dynamical modelling codes using Schwarzschild’s orbit superposition technique dealing with triaxiality by van den Bosch et al. (2008) and Vasiliev & Valluri (2019). Their estimated precision and efficiency will be later mentioned in the discussion. In this paper, we present our newly developed triaxial Schwarzschild code called SMART (‘Structure and MAAss Recovery of Triaxial galaxies’) and test the code on a realistic high-resolution numerical merger simulation including SMBHs by Rantala et al. (2018).

This paper is structured as follows. We introduce SMART and describe its specific benefits in Section 2. We then discuss the most important aspects for choosing this particular simulation in Section 3. In this section, we furthermore explain all relevant steps to extract the data needed for modelling the simulation. In Section 4, we will show the results of these models. Section 5 deals with the quasi-uniqueness of the anisotropy recovery when fitting full line-of-sight velocity distributions (LOSVDs). This is followed by a short discussion about remaining sources of systematics and comparison to other triaxial modelling codes in Section 7. We summarize our results and conclusions in Section 8.

2 TRIAXIAL SCHWARZSCHILD CODE SMART

SMART (‘Structure and MAAss Recovery of Triaxial galaxies’) is a fully 3D orbit superposition code based on the axisymmetric code of Thomas et al. (2004) and its original extension to three dimensions and non-axisymmetric densities by Finozzi (2018). It is written in FORTRAN 90/95 (Brainerd et al. 1996). SMART follows the classical Schwarzschild method consisting of the computation of the potential and forces for a given density, the setting up of an orbit library and the subsequent superposition of the orbits.

In the following sections, we will explain in more detail how SMART creates self-consistent density-, potential-, and orbit-configurations and how we weight the orbits in order to fit the input density and velocity structure. One main feature of our code is that

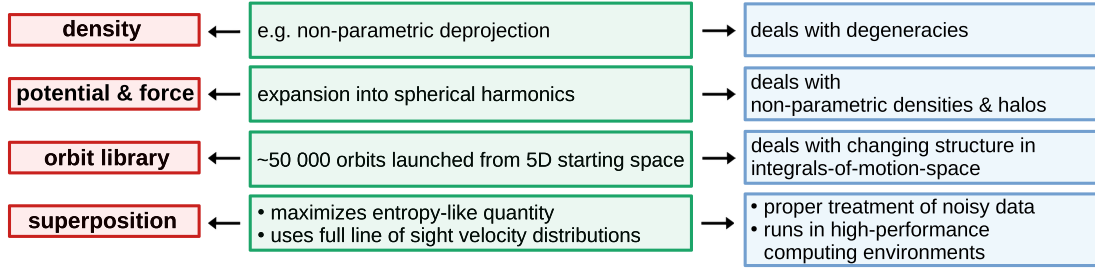


Figure 1. Our implemented triaxial Schwarzschild Modelling code SMART follows the classical Schwarzschild Modelling routine (red panels), but it is unique in calculating the potential by expansion into spherical harmonics, setting up an adaptive orbit library and computing the orbit superposition by maximizing an entropy-like quantity (green panels). This results in specific advantages (blue panels), e.g. that our code is able to deal with realistic changes in the gravitational potential, e.g. when the SMBH causes a more spherical potential in the centre. Our orbit library is adaptive and responds to changes in the integrals of motion. SMART can process density output from any deprojection routine, e.g. a non-parametric deprojection dealing with degeneracies.

it uses a 5D starting space for the orbit library to adapt to potentials with a radially varying structure of the integrals-of-motion space. Fig. 1 gives a schematic overview of the code’s main modules and the benefits of their specific implementation.

2.1 Coordinate systems and binning

We use two different coordinate systems to describe the intrinsic and projected properties of a galaxy. To transform between the intrinsic coordinates (x, y, z) , adapted to the symmetry of the object, and the coordinates (x', y', z') adapted to the sky projection, two matrices P and R are used:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = R \cdot P \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad (1)$$

with

$$R = \begin{pmatrix} \sin \psi & -\cos \psi & 0 \\ \cos \psi & \sin \psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2)$$

and

$$P = \begin{pmatrix} -\sin \vartheta & \cos \vartheta & 0 \\ -\cos \vartheta \cos \varphi & -\cos \vartheta \sin \varphi & \sin \vartheta \\ \sin \vartheta \cos \varphi & \sin \vartheta \sin \varphi & \cos \vartheta \end{pmatrix}. \quad (3)$$

P and its corresponding viewing angles ϑ and φ project to the plane of the sky with z' being the line of sight. R and its corresponding rotation angle ψ rotate the coordinates x' and y' in the plane of the sky along z' . If not stated otherwise, the intrinsic long axis is hereafter assumed to coincide with x , the intrinsic intermediate axis with y and the intrinsic short axis with z .

SMART works with a cell structure based on spherical coordinates. Intrinsic properties like the stellar or dark matter distribution or individual orbital properties are integrated over small cells in configuration and/or velocity space. We use a linear sampling for the longitude $\theta \in [-90^\circ, 90^\circ]$ and the azimuth $\phi \in [0^\circ, 360^\circ]$. Radial bins are spaced in even intervals of the radial binning index

$$i_r = \frac{1}{a} \log \left(c + \frac{a}{b} r \right). \quad (4)$$

The constant c allows to adapt the central binning scheme from logarithmic ($c = 0$) to linear ($c = 1$). The constants a and b are determined once the radial extent of the library $r_{\min} \leq r \leq r_{\max}$ and

the number of radial bins N_r are set and they are chosen so that the minimum radius r_{\min} lies within the first radial bin and the maximum radius r_{\max} in the last one (see also Siopis et al. 2009).

Similar to the spatial properties, the LOSVDs are integrated over small cells in phase space given by the spatial pattern of the observations (e.g. Voronoi bins; Cappellari & Copin 2003) and the velocity resolution of the LOSVD data. Like its axisymmetric predecessor (Thomas et al. 2004), the code uses the entire information contained in the full LOSVDs. See Sections 3.1.3 and 3.1.4 for more details.

2.2 Density and potential

The total gravitational potential

$$\Phi = \Phi_* + \Phi_{\text{DM}} + \Phi_{\text{SMBH}} \quad (5)$$

is composed as the sum of a Keplerian contribution from a supermassive black hole (Φ_{SMBH}) and the contributions from the stars (Φ_*) and dark matter (Φ_{DM}). SMART allows to use non-parametric densities for both the stars and the dark matter. The stellar density is generally assumed to be provided in 3D tabulated form (as for example returned from a non-parametric deprojection, e.g. de Nicola et al. 2020). The same holds for the dark matter halo. However, the code can also run with quasi-parametric deprojections (e.g. Multi Gaussian Expansion or MGE models, Monnet, Bacon & Emsellem 1992; Emsellem, Monnet & Bacon 1994; Cappellari 2002). It can also run with parametric dark matter haloes (e.g. NFW profiles; Navarro, Frenk & White 1996).

The solution to the Poisson equation is obtained with the help of an expansion in spherical harmonics (Binney & Tremaine 2008). For this, the stellar and dark matter density are individually interpolated by first performing a bi-linear interpolation among the elevation and azimuthal angle bins and afterwards a linear interpolation among the logarithm of the radial bins. For integrating these interpolated densities we use a 10-point Gaussian quadrature algorithm from Press et al. (1996). The advantage of calculating the potential by expansion into spherical harmonics in comparison to other techniques, e.g. by using the MGE method, is its ability to deal with non-parametric densities and haloes.

2.3 Orbit library

Every orbit in a gravitational potential is uniquely defined by its integrals of motion (e.g. Binney & Tremaine 2008). The number of (isolating) integrals of motion depends on the given potential.

Furthermore, every integral of motion reduces the dimensionality of the trajectories of the stars in the galaxy. Regular potentials admit in general three integrals of motion, one of which being the energy E . In the axisymmetric case, another classical integral is known explicitly: the z -component of the angular momentum, L_z . The third integral, I_3 , is usually not known explicitly. In the general triaxial case, only the energy is given explicitly. Near the central SMBH, however, the potential becomes more and more Keplerian and the number of isolating integrals of motion in a Keplerian potential is five. An example for a system in an almost Keplerian potential that is described by a five-integral distribution function is the asymmetric disc in the centre of M31 (Bender et al. 2005; Brown & Magorrian 2013). Since our code is not restricted to axisymmetric or triaxial symmetries, we aim for a 5D starting space for the stellar orbits, which we gain by systematically sampling E , L_z , v_r , r , and ϕ . The details of the initial conditions sampling technique are described in Section C.

In total, SMART sets up and integrates $\sim 50\,000$ orbits for 100 surfaces of section (SOS) crossings, i.e. for 100 crossings of the equatorial plane in upwards direction (for a more detailed description of SOS, see Appendix C). This number of orbits was intentionally chosen to be higher than in the axisymmetric predecessor code of Thomas et al. (2004) to address all necessary complexities given by, e.g. a radially changing structure in the integrals-of-motion space. Moreover, this amount of orbits proves to be sufficient to directly recover the (phase-space) density without any dithering of orbits in the starting space.

If the potential at hand is, e.g. axisymmetric, then all orbits conserve L_z and precess around the rotation axis. In this case, the orbits will fill the ϕ dimension automatically. Likewise, orbits will be represented by invariant curves in the (r, v_θ) -plane, due to the conservation of I_3 . Hence, when sequentially sampling the orbital launch conditions, the dimension of the submanifold containing all the orbital initial conditions that are not yet represented shrinks automatically, according to the number of integrals of motion provided by the gravitational potential under study.

Since, in general, triaxial potentials have three integrals of motion, a 2D starting space at a given energy (de Zeeuw 1985; Schwarzschild 1993) would provide a sufficient orbit sampling: One could sample initial conditions from the (x, z) -plane producing mainly tube orbits and compensate this with launching additional box orbits from the equipotential surfaces (van den Bosch et al. 2008). However, it is not clear whether the distribution function of realistic triaxial galaxies requires a 5D starting space near the SMBH in the centre. With our choice of a 5D starting space, we guarantee that our set of orbits adapts to the actual complexity of the integrals-of-motion space. In a realistic triaxial galaxy, like in the studied simulation, it changes from a more spherical centre (requiring at least four integrals of motion) into nearly prolate outskirts. Furthermore, it allows us to model systems like eccentric discs with distribution functions that obviously depend on more than three integrals of motion. In Fig. C1, we show that our implemented orbit sampling and integration routine (cf. Section 2.3.1) yields a homogeneous and dense coverage of phase space.

2.3.1 Orbit integration and classification

SMART integrates the orbital equations of motions $\frac{d\vec{x}_i}{dt} = -\vec{\nabla}\phi(\vec{x}_i)$, where i denotes the orbit index, in Cartesian coordinates by means of the Cash–Karp algorithm (Cash & Karp 1990). The fifth-order Runge–Kutta method is implemented by using an adaptive integra-

tion step-size (see Press et al. 1996). The default integration time of the individual orbits corresponds to 100 SOS crossings.

At each integrated time-step, the contribution of orbit i to the luminosity, internal velocity moments, and projected LOSVDs is calculated as the fraction of time the orbit spends in the corresponding bins. Projected quantities are convolved with the relevant PSF (point spread function) in every time-step and before binning. The PSF can either be provided as a parametrized 2D Gaussian or in terms of a PSF image. The convolution is performed via Monte Carlo method by randomly perturbing the coordinates $x'(t)$, $y'(t)$ (cf. Section 2.1) in dependence of the respective PSF.

Modelling a galaxy with SMART does not require an exact orbit classification analysis. However, we built in an approximate classification method. For this, SMART checks the sign conservation of the angular momentum in x -, y -, and z -direction for every SOS-crossing event. If the sign of L_x is conserved for the whole integration time and if this is not true for L_y and L_z , the orbit gets classified as x -tube. The same applies to the other directions (cf. Barnes 1992). If the 100 SOS crossings do not hold an angular momentum sign conservation along any direction, the orbit gets classified as box/chaotic orbit. If the sign conservation is true for every direction or if the orbit shows no radial and azimuthal change during the integration time, the orbit is classified as spherical/Kepler orbit.

2.4 Orbit superposition

The orbital weights w_i , which are decisive for the consistency between the observed and modelled luminosity as well as for the projected velocity profiles, are iteratively changed until the difference χ^2 between the observed LOSVDs $\mathcal{L}_{\text{data}}$ and modelled LOSVDs \mathcal{L}_{mod} is minimal:

$$\chi^2 = \sum_{j'}^{N_{\text{losvd}}} \sum_k^{N_{\text{vlos}}} \left(\frac{\mathcal{L}_{\text{data}}^{j'k} - \mathcal{L}_{\text{mod}}^{j'k}}{\Delta\mathcal{L}_{\text{data}}^{j'k}} \right)^2. \quad (6)$$

Here, j describes the spatial bin index of the N_{losvd} data cells and k describes the velocity bin index of the N_{vlos} velocity bins. $\Delta\mathcal{L}_{\text{data}}^{j'k}$ is the error of the data in the specific bin. An advantage of SMART is that it uses the full information contained in the LOSVDs and not only the Gauss–Hermite parameters alone (cf. e.g. Mehrgan et al. 2019 for a discussion of the benefits of using non-parametric LOSVDs in measuring galaxy masses). The luminosity density serves as a boundary condition for the choice of the orbital weights.

The problem of solving for the weights w_i is usually underdetermined because the number of orbits is much larger than the number of data points. We therefore regularize our models by maximizing an entropy-like quantity

$$\hat{S} \equiv S - \alpha \chi^2, \quad (7)$$

where

$$S = - \sum_i w_i \ln \left(\frac{w_i}{\omega_i} \right). \quad (8)$$

In the absence of any other constraints, the entropy maximization yields $w_i \propto \omega_i$ (cf. Section 5). Thus, the ω_i are bias factors for the orbital weights w_i and can be used to smooth the orbit model. Moreover, they can be used to construct orbit models with specific properties, e.g. orbit models dominated by certain families of orbits (cf. Section 5 for examples).

The particular form of the entropy in equation (8) guarantees the positivity of the orbital weights. The maximum-entropy technique is

flexible, however. Other choices for the entropy allow for negative weights as well (Richstone & Tremaine (1988)).

Technically, for each regularization value α , SMART maximizes \hat{S} by computing the relevant Lagrange multipliers. The iterative adjustment of the w_i 's is performed by using Newton's method. The implemented method is based on Richstone & Tremaine (1988). A detailed description of the algorithm as well as tests demonstrating the high accuracy performance of SMART can be found in Appendix D.

As we will describe in full detail in Section 5, the entropy term in equation (7) makes the solution of the orbital weights unique. While this might be advantageous from an algorithmic point of view it comes in principle with the danger of a potential bias. As we will show, by varying the orbital bias factors ω_i the maximum-entropy technique allows in principle to reconstruct any of the potentially degenerate solutions for the orbital weights (cf. Section 5). However, the results of the following sections imply that when fitting all the information contained in the full LOSVDs, the remaining degeneracies in the weight reconstruction have only little impact on the 'macroscopic' galaxy parameters of interest, like the anisotropy for example.

One natural choice for the orbital bias factors is $\omega_i = V_i$, where V_i is the phase-space volume represented by orbit i (cf. e.g. Richstone & Tremaine 1988; Thomas et al. 2004). With this choice,

$$S = - \sum_i w_i \ln \left(\frac{w_i}{V_i} \right) = - \int f \ln(f) d^3r d^3v \quad (9)$$

equals the Boltzmann entropy. Since the Boltzmann entropy increases during dissipationless evolutionary processes due to phase mixing and violent relaxation, galaxy models with large Boltzmann entropy are more likely than those with a small one (Richstone & Tremaine 1988; Thomas et al. 2007). However, in collisionless self-gravitating systems, every entropy-like functional is assumed to increase in phase space, such as the generalized H-function (Tremaine, Hénon & Lynden-Bell 1986), the entropy of the ideal gas (White & Narayan 1987) or the Tsallis entropy (Tsallis 1988). The choice of ω_i is thus arbitrary to some degree. This and the fact that in case of a triaxial potential it is computationally expensive to calculate the correct phase-space volume V_i for every orbit motivated us to set

$$\omega_i = \text{const.} = 1. \quad (10)$$

This functional form was also tested by de Lorenzi et al. (2007) in a slightly different context of a made-to-measure (M2M; Syer & Tremaine 1996; Bissantz, Debattista & Gerhard 2004) algorithm for N -body particle models. Compared to the Boltzmann entropy, $\omega_i = \text{const.}$ leads to relative preference of orbits with actually small V_i , while orbits with large V_i are relatively suppressed. With this choice of constant orbital bias factors ω_i , the entropy equation (cf. equation 8) resembles the Shannon entropy and yields the least 'informed' set of orbital weights.

2.5 Mass optimization with SMART

SMART is conceived to determine the viewing angles and the mass components, like the dark matter halo, the stellar mass-to-light ratio and black hole mass by looking for the model with the smallest χ^2 . To deal with this multidimensional parameter space, SMART uses NOMAD (Nonlinear Optimization by Mesh Adaptive Direct search), a software optimized for time-consuming constrained black-box optimizations (Audet & Dennis 2006; Le Digabel 2011). NOMAD is able to optimize a noisy function with unknown derivatives to

converge to the best-fitting model by using a direct-search scheme. SMART runs on multiple computer cores. The orbit processing (including the setup of the initial conditions, orbit integration, orbit classification, and computation of the internal and projected velocity distributions) as well as the relevant linear algebra operations applied for their superposition are parallelized.

3 THE N -BODY SIMULATION

In order to test SMART on a realistic mock galaxy, we use a high-resolution collisionless numerical merger simulation by Rantala et al. (2018). The simulation is a single generation binary galaxy merger of two equal-mass elliptical galaxies with an effective radius of 7 kpc hosting an SMBH of $8.5 \times 10^9 M_\odot$ each and corresponds to the so-called γ -1.5-BH-6 simulation in Rantala et al. (2018). The two initial galaxies are set up by using a spherically symmetric Dehnen density-potential (Dehnen 1993) with an initial inner stellar density slope of $\rho \propto r^{-3/2}$ for both progenitor galaxies. The merger results in a remnant triaxial galaxy with a SMBH of $1.7 \times 10^{10} M_\odot$ with a sphere of influence¹ of $r_{\text{SOI}} \sim 1$ kpc and an effective radius of $r_e \sim 14$ kpc. With that, the remnant resembles NGC 1600, a galaxy showing a very large core with a tangentially biased central stellar orbit distribution (Thomas et al. 2016). The simulation is based on the hybrid tree- N -body code KETJU (Rantala et al. 2017; Karl et al. 2015) which is able to accurately compute the dynamics close to the black hole due to the algorithmic chain regularization method AR-CHAIN (Mikkola & Merritt 2006, 2008). The computation of the global galactic dynamics is based on the tree code GADGET-3 (Springel 2005).

We analyse a snapshot of the simulation ~ 1.4 Gyr after the galaxy centres have merged, such that the remnant can be assumed to be in a steady state. At this stage, the actual distance of the two merging black holes in the simulation is 5 pc. The merger remnant shows a radially varying triaxiality parameter, being more oblate at smaller radii and increasingly more round in the centre with a maximum of triaxiality, i.e. $T = 0.5$, at about 3 kpc and more prolate outskirts. The simulation contains 8.3×10^6 stellar particles with masses of $10^5 M_\odot$ each, leading to a total stellar mass of $8.3 \times 10^{11} M_\odot$. The mass ratio of one SMBH and one stellar particle is $M_{\text{BH}}/M_* = 8.5 \times 10^4$ and therefore sufficiently large to investigate a realistic interaction of the SMBH binary with the stars (Mikkola & Valtonen 1992). The number of dark matter particles is 2×10^7 with masses of $7.5 \times 10^6 M_\odot$ each, leading to a total dark matter mass of $1.5 \times 10^{14} M_\odot$.

The simulation is particularly suitable to test SMART because of its (1) very high resolution (including properly resolved black hole dynamics); (2) realistic orbital structure and shape; (3) realistic mass composition (black hole, stars, and dark matter) with a realistic stellar density core (e.g. Thomas et al. 2009; Rantala et al. 2019). Finally, the number of stellar particles is large enough to measure fully resolved LOSVDs from the central sphere of influence of the black holes out to dark-matter-dominated regions.

3.1 Processing the simulation data

3.1.1 Orientation of the simulation

We aim at orienting our intrinsic coordinate system as closely as possible to the intrinsic symmetry axes of the merger remnant.

¹We here use the definition of the sphere of influence as the radius within which the total stellar mass equals the black hole mass, i.e. $M_*(r_{\text{SOI}}) = M_{\text{BH}}$.

However, the stellar and dark matter principal axes of the simulated remnant are not aligned. Hence, the orientation of the main axes depends on the radius and on the mass component for which the reduced inertia tensor (see e.g. Bailin & Steinmetz 2005) is calculated. Such a shift between the stellar and dark matter halo axes is not unexpected for collisionless merger simulations (see e.g. Novak et al. 2006). We decided to centre the remnant on the stars and black holes and afterwards orientate it by using the reduced inertia tensor for stars and dark matter within 30 kpc. With this, the stellar elliptical isophotes for the three different projections are well aligned with the projected principal axes within the field of view of $15 \text{ kpc} \times 15 \text{ kpc}$. There is a negligible residual misalignment which is strongest for the major axis projection but nowhere larger than $\sim 5^\circ$ (see Figs 3 and A1).

3.1.2 Density

Due to the good alignment and taking advantage of the nearly triaxial intrinsic symmetry of the merger remnant, we increase the resolution of the simulation for computing the density by a factor of 8 by folding all stellar and dark matter particles into one octant. The stars are then binned into concentric radial shells with 1000 stars in each shell and the dark matter particles are binned into radial shells with 5000 dark matter particles each. The single shells are subdivided into angular bins such that the elevation angle $\theta \in [0^\circ, 90^\circ]$ increases in constant $\sin(\theta)$ -steps of 0.1 and that the azimuthal angle $\phi \in [0^\circ, 90^\circ]$ increases in constant ϕ -steps of 10° .

Within $r < 0.28 \text{ kpc}$ for the stellar and $r < 10 \text{ kpc}$ for the dark matter particles, the resolution is too low to extract a smooth density. Here, we extrapolate the logarithmic densities from the outer parts by a first-order polynomial fit in the logarithm of the radius. For the stellar densities, we use the slope and vertical intercept averaged over all angular bins. The dark matter density is extrapolated by using the slopes and vertical intercept values of the individual angular bins. We ensure that the total enclosed mass is well covered by the extrapolated density. To smooth the radial density profiles, SCIPY's `Gaussian.filter()`-function (Virtanen et al. 2019) is used.

3.1.3 Spatial binning

For spatially binning the kinematic input data, we use the Voronoi tessellation method of Cappellari & Copin (2003). For each tested projection (see Section 3.2), we construct a separate set of Voronoi bins. To end up with a roughly constant number of stellar simulation particles N_* in each bin, we define the signal-to-noise ratio as

$$\frac{\text{signal}}{\text{noise}} = \sqrt{\text{signal}} = \sqrt{N_*} = \begin{cases} 70 & \text{for } r < r_{\text{SOI}}, \\ 150 & \text{for } r_{\text{SOI}} < r < 15 \text{ kpc}. \end{cases} \quad (11)$$

When calculating the average over the five different projections, this results in a total number of $N_{\text{losvd}} = N_{\text{voronoi}} = 227$ Voronoi bins within the whole field of view and 54 Voronoi bins within r_{SOI} .

This resolution is chosen to conform with realistic observational data and does not exceed high-resolution wide-field spectral observations by, e.g. MUSE (cf. Mehrgan et al. 2019) but still proves to be sufficiently high for this analysis.

3.1.4 Kinematic data and velocity binning

For each spatial bin, we calculate the LOSVDs for $N_{\text{vlos}} = 45$ equally-sized velocity bins with $v_{\text{min}}^{\text{max}} = \pm 1600 \text{ km s}^{-1}$, which is chosen so

that it covers about 10 times the velocity dispersion. This results in a velocity resolution of $\Delta v_{\text{vlos}} = 71.11 \text{ km s}^{-1}$.

It is not trivial which ‘error’ for the kinematic input data of the simulation should be used since we do not have an error in the simulation in the sense that repeated measurements give the same results. However, the χ^2 -minimization formally requires information about an ‘error’. We tested several assumed ‘error-bars’, such as

- (i) the difference between two kinematic data sets, each determined by using half of the simulation particles,
- (ii) the Poisson noise, and
- (iii) a constant absolute error for each LOSVD as 10 per cent of the maximum per LOSVD.

In order to prevent an underestimation of the relative error for the major axis projection holding more particles along the line of sight than the other projections, the constant absolute error proves to be the most suitable method and is used in the present analysis. The choice of setting the error value to 10 per cent of the maximum value of each LOSVD corresponds to a velocity uncertainty of $\Delta v = 13 \text{ km s}^{-1}$, dispersion error of $\Delta \sigma = 13 \text{ km s}^{-1}$ (or 5 per cent relative error) and $\Delta h_n = 0.03$ for the higher order Gauss–Hermite moments. This is a reasonable choice both in terms of real observational errors and of the scatter in the kinematic maps of the merger simulation (see Figs 3 and B1).

3.2 Applying SMART to the simulation

For the purpose of testing our code, SMART is provided with the correct viewing angles and with the 3D normalized stellar (i.e. luminosity) density ρ_* from the simulation as well as with the normalized 3D dark matter density ρ_{DM} . The DM scaling parameter s_{DM} is to be determined by SMART. We skipped any surface brightness deprojection, since degeneracies in the deprojection (see e.g. de Nicola et al. 2020) would only hamper a correct evaluation of our code. We parametrize the density as

$$\rho = M_{\text{BH}} \times \delta(r) + \Upsilon \cdot \rho_* + s_{\text{DM}} \cdot \rho_{\text{DM}}, \quad (12)$$

where our fit parameters are the black hole mass M_{BH} , stellar mass-to-light ratio Υ , and the multiplication factor s_{DM} defining the magnitude of the dark matter density profile favoured by SMART. We determine these parameters by finding the minimum in χ^2 .

We model and analyse five different projections with (1) $\vartheta = 90^\circ$, $\varphi = 0^\circ$, i.e. the major axis projection, (2) $\vartheta = 90^\circ$, $\varphi = 90^\circ$, i.e. the intermediate axis projection, (3) $\vartheta = 0^\circ$, $\varphi = 90^\circ$, i.e. the minor axis projection, (4) $\vartheta = 90^\circ$, $\varphi = 10^\circ$, i.e. a projection 10° off the major axis in azimuthal direction, and (5) $\vartheta = 90^\circ$, $\varphi = 45^\circ$, i.e. a projection in between the major and intermediate axis projection. Without loss of generality, ψ was set to 90° for all three viewing directions, i.e. R (see equation 2) equals the unit matrix.

The field of view is chosen to be $15 \text{ kpc} \times 15 \text{ kpc}$. The minimum sampled starting radius is set to $r_{\text{min}} = 0.05 \text{ kpc}$ and the maximum sampled starting radius is set to $r_{\text{max}} = 80 \text{ kpc}$. We find optimal results for a central binning with $c = 0.5$ in equation (4) because this guarantees that in case of the simulated merger remnant the difference of the circular velocity within one radial bin equals the model's velocity resolution $\Delta v_{\text{vlos}} = 71.11 \text{ km s}^{-1}$ at a radius of $r = 0.16 \text{ kpc} = 0.16 \cdot r_{\text{SOI}}$. For $c = 1$, this would be only reached at $r = 0.38 \text{ kpc} = 0.38 \cdot r_{\text{SOI}}$, resulting in a deteriorated black hole mass recovery by ~ 10 per cent.

For each tested projection, we model two halves of the LOS kinematic data. After correct projection on to the plane of the sky (see Section 2.1), we separately model the half of the kinematic data with

positive x' -coordinates (hereafter called ‘right half’ of the galaxy) and the half of the kinematic data with negative x' -coordinates (hereafter called ‘left half’ of the galaxy).

4 RESULTS

4.1 Choice of regularization

When applying the code to realistic noisy measurements, regularization becomes important to prevent the orbital weights to fit the noise in the data. The optimal regularization parameter α for a specific observational data set can be determined by running Monte Carlo simulations on kinematic mock data (Thomas et al. 2005) and is given as the one providing the minimum deviation of the intrinsic properties (like the distribution function, or velocity moments, or mass parameters) in comparison to the default model. When fitting noiseless ideal data, one would expect best results for $\chi^2 \rightarrow 0$, or $\alpha \rightarrow \infty$ (neglecting recovery degeneracies and assuming an ‘error’ can be defined). Even in that case, however, due to residual systematics (like finite resolution of the orbit library etc.) and due to the intrinsic noise in the N -body simulation, we still expect that the best result may not necessarily be achieved asymptotically for very large α , but already for some finite value of the regularization parameter. To take this into account, we split the code test into two phases: (1) We fit the orbit model with the correct mass parameters and determine that value of α for which the internal structure of the simulation is best recovered. Specifically, we use the second-order velocity moments for this comparison. (2) We then also vary the mass parameters and test how well they can be recovered. Our benchmark is the optimized α from the comparison of the moments, but we will discuss the results for all α to demonstrate their robustness.

For determining the deviation between the model’s velocity dispersions σ_r , σ_θ , and σ_ϕ and the real ones from the simulation, we define

$$\text{rms}_\sigma = \frac{1}{3} \sum_i \text{rms}_{\sigma_i} = \frac{1}{3} \sum_i \sqrt{\frac{1}{N_{\text{data}}} \sum_{j=1}^{N_{\text{data}}} \left(\frac{\sigma_{i,\text{data}} - \sigma_{i,\text{mod}}}{\sigma_{i,\text{data}}} \right)^2}, \quad (13)$$

where the index i denotes the three coordinates r , θ , ϕ .

These rms_σ profiles of the five tested projections and their respective modelled halves in dependence of the regularization parameter α are plotted as thick lines in the top row of Fig. 2. The x -axis ticks, thereby mark all α -values that were tested in this analysis.

The second row in Fig. 2 shows the quality of the fit as χ^2/N_{data} profile (for definition of χ^2 see formula 6) again plotted against the regularization parameter. χ^2 as deviation from the kinematic input data with the modelled fit is here normalized over the number of input data N_{data} , which is composed of the number of Voronoi bins N_{voronoi} times the number of kinematic bins N_{vos} . As expected, the fit to the data is poor when α is low (high χ^2/N_{data}). In this regime, it is the entropy term which is essentially maximized and the data (via $\alpha \cdot \chi^2$, cf. equation 7) have little influence on the fit. With this, the anisotropy strongly depends on the ω_i and, in our case, happens to be a poor representation of the internal moments of the merger (high rms values, see Fig. 2, first row). With increasing α , both the fit quality and the agreement with the merger structure improve. However, at very high α -values further improvements of the fit do not make the internal moments better since we are dominated by the noise of the N -body simulation. The most suitable choices of regularization can be read off from the minima $\min(\text{rms}_\sigma)$ of the rms_σ profiles and are marked as thick vertical lines in Fig. 2. Their average value is

$\alpha(\min(\text{rms}_\sigma)) = 0.41$. This value, however, depends on the specific implemented set-up of SMART as well as of the input data. Evaluated at the individual most suitable regularization values and afterwards averaged over all 10 models (5 projections and two halves each), we get a minimum value of only $\min(\text{rms}_\sigma) = 0.008$. This extremely good agreement demonstrates that our orbit sampling represents the phase space very well.

We also test the influence of rotation on the comparison. Due to the overall small angular momentum and thus small absolute velocity amplitude however, the relative errors in v are sometimes large. The absolute error of the first-order velocity moments, averaged over all angular and radial bins, is only $\Delta v = 3.9 \text{ km s}^{-1}$, but the maximum velocity over these bins is likewise only 23.4 km s^{-1} . We therefore here define $\text{rms}_{v,\sigma}$ as normalized deviation between the first internal moments and the velocity dispersions together as

$$\text{rms}_{v,\sigma} = \frac{1}{3} \sum_i \text{rms}_{v_i,\sigma_i} = \frac{1}{3} \sum_i \sqrt{\frac{1}{2N_{\text{data}}} \sum_{j=1}^{N_{\text{data}}} (\Delta v_i^2 + \Delta \sigma_i^2)},$$

$$\text{with } \Delta v_i = \frac{v_{i,\text{data}} - v_{i,\text{mod}}}{\sqrt{v_{i,\text{data}}^2 + \sigma_{i,\text{data}}^2}} \text{ and } \Delta \sigma_i = \frac{\sigma_{i,\text{data}} - \sigma_{i,\text{mod}}}{\sqrt{v_{i,\text{data}}^2 + \sigma_{i,\text{data}}^2}}, \quad (14)$$

where the index i again denotes the three coordinates r , θ , ϕ .

The $\text{rms}_{v,\sigma}$ profiles are plotted against the regularization parameter as thin lines in the top row of Fig. 2. As expected, their minima appear at regularization values similar to the ones of the rms_σ profiles. They are marked as thin vertical lines and their average value is $\alpha(\min(\text{rms}_{v,\sigma})) = 0.012$. Again, this value depends on the specific set-up. Evaluated at the individual $\alpha(\min(\text{rms}_{v,\sigma}))$ -values and afterwards averaged over the five different projections and their respective modelled halves we gain a value of only $\min(\text{rms}_{v,\sigma}) = 0.012$.

All values in the vicinity of $\alpha(\min(\text{rms}_\sigma))$ and $\alpha(\min(\text{rms}_{v,\sigma}))$ are good regularization choices. Within this regularization region, SMART is able to well fit the kinematic input data for each tested projection (see Figs 3 and B1). Averaged over the five different projections and their respective modelled halves we receive mean values and deviations of $\bar{v} = (0.11 \pm 2.09) \text{ km s}^{-1}$, $\bar{\sigma} = (309.75 \pm 2.54) \text{ km s}^{-1}$, $\bar{h}_3 = 0.00 \pm 0.01$ and $\bar{h}_4 = 0.01 \pm 0.01$, again evaluated at $\alpha(\min(\text{rms}_\sigma))$. The χ^2 -maps in Figs 3 and B1 show that the models for each projection are able to fit the kinematic input data homogeneously well over the field of view with slightly larger deviations in the centre. If not specifically annotated, the results shown in the further analysis are evaluated at $\alpha(\min(\text{rms}_\sigma))$. However, the quality of the models does not strongly depend on the exact regularization value because a broader range of regularization values around these determined α -values is sufficiently appropriate and results in equally good mass parameter reproductions within the overall scatter (see also third and fourth lines in Fig. 2 that will be explained in Section 4.3).

4.2 Reproduction of internal moments and orbit structure

The previously shown small rms_σ and $\text{rms}_{v,\sigma}$ values already demonstrate the very good recovery of the internal moments by the model when providing the correct mass parameters and viewing angles. The high level of agreement between model and simulation is further illustrated in Fig. 4. The anisotropy parameter $\beta = 1 - \frac{\sigma_\theta^2 + \sigma_\phi^2}{2\sigma_r^2}$ of the model matches the profile of the simulated one (see Fig. 4, middle panel). The model is also able to reproduce the negative β within the core radius r_b , which equals the black hole sphere of influence

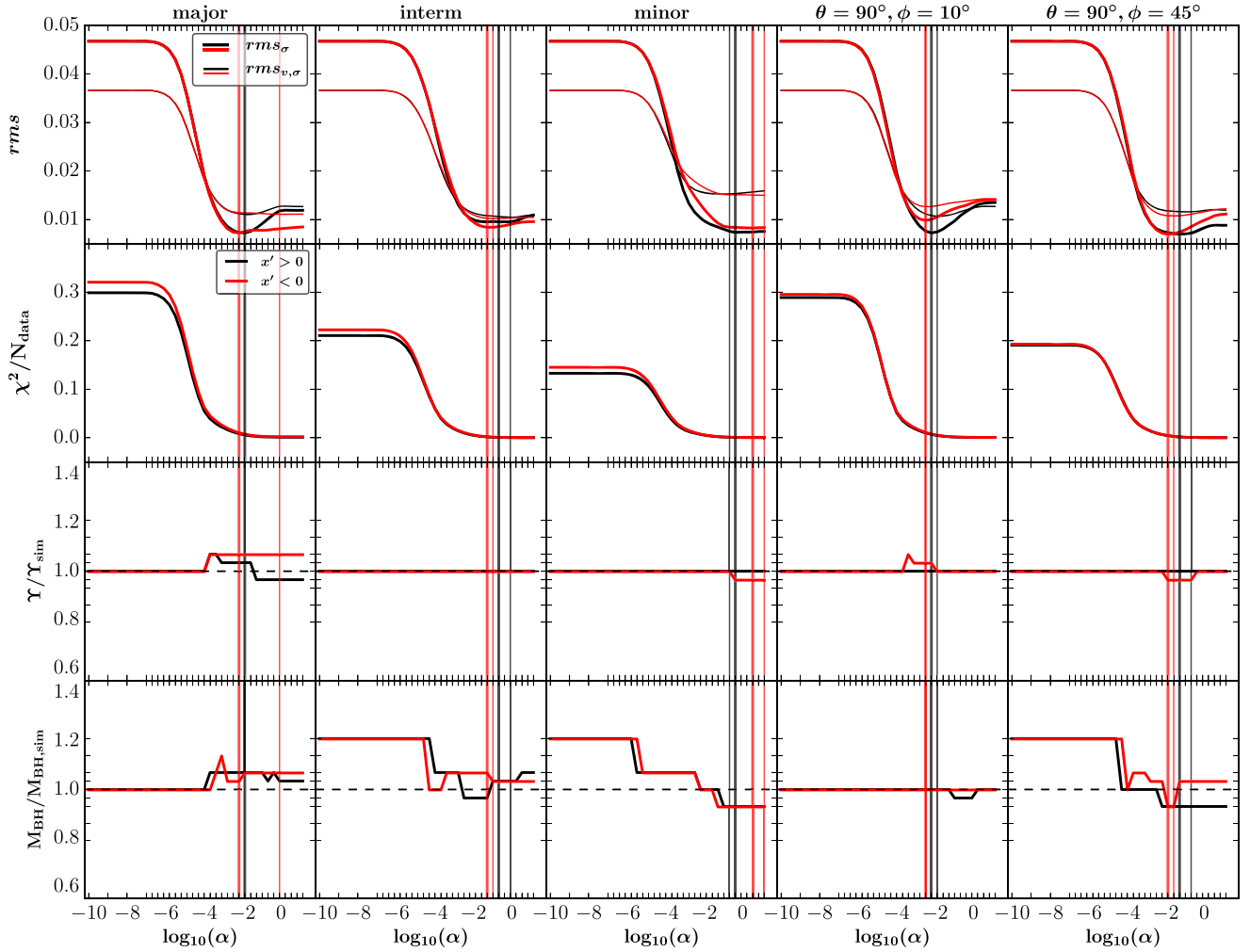


Figure 2. Choice of regularization and 1D mass recovery results for the five different projections (different columns). The first row shows the rms_σ - (thick line) and $\text{rms}_{v,\sigma}$ -profile (thin line) for the modelled right half of the galaxy (black) and left half of the galaxy (red) with the correct black hole mass, stellar mass-to-light ratio and dark matter scale factor as input. All values are plotted against the increasing regularization value α in logarithmic units. The x -axis-ticks thereby symbolize the tested α -values. The minima $\min(\text{rms}_\sigma)$ (thick line) and $\min(\text{rms}_{v,\sigma})$ (thin line) are marked as vertical lines and suggest suitable regularization values. The thin black vertical line in the major axis panel thereby overlaps with the thick black vertical line and the thin red vertical line in the $\theta = 90^\circ, \phi = 10^\circ$ panel overlaps with the thick red vertical line. The second row shows the corresponding χ^2/N_{data} values. The third and fourth rows show the 1D mass recovery results for the stellar mass-to-light ratio $\Upsilon/\Upsilon_{\text{sim}}$ and black hole mass $M_{\text{BH}}/M_{\text{BH,sim}}$ normalized over the correct values of the simulation. The y -axis-ticks for the third and fourth row symbolize the concrete masses that were tested and used as input values for the models. The black dotted line marks unity which is achieved when the model correctly recovers the mass. The stellar mass-to-light ratio and black hole mass were recovered with an accuracy better than 5 per cent. Such an intrinsic precision under similar conditions has not yet been demonstrated with other Schwarzschild modelling codes.

(Thomas et al. 2016), reflecting the tangential orbit distribution due to black hole ‘core scouring’: Within the sphere of influence ETGs at the high-mass end exhibit central regions that are fainter than an extrapolation of a Sérsic function (Sérsic 1963) as fit to the outer surface brightness profile would suggest. The commonly accepted theory for the formation of these ‘cores’ is a gravitational slingshot process of stars on radial orbits caused by SMBH binaries that were arised by galaxy mergers (e.g. Begelman, Blandford & Rees 1980; Hills & Fullerton 1980; Ebisuzaki, Makino & Okumura 1991; Milosavljević & Merritt 2001; Merritt 2006; Rantala et al. 2018).

The classification of the integrated orbits is also plotted in Fig. 4 (bottom panel). Orbits in the immediate vicinity of the black hole are spherical and Keplerian orbits as expected due to the SMBH. Z-tubes are the pre-dominant orbits within $0.3 \text{ kpc} < r < 3 \text{ kpc}$ causing the more oblate shape of the galaxy in this range. For $r > 3 \text{ kpc}$,

the majority of orbits are classified as x -tubes corresponding to the prolate shape in the outskirts of the simulation as described in Section 3. Our orbit classification analysis well matches the one by Frigo et al. (in preparation), which is done via an orbit frequency analysis of the simulation.

4.3 Mass recovery

So far, we have provided SMART with the correct mass parameters of the stellar mass-to-light ratio Υ_{sim} , black hole mass $M_{\text{BH,sim}}$, and dark matter multiplication factor $s_{\text{DM,sim}}$ of the simulation. Even though SMART can be provided with any type of dark matter profile, we here transfer the dark matter density profile shape of the simulation and concentrate on finding the correct mass multiplication scale factor s_{DM} of this pre-determined halo shape (cf. equation 12).

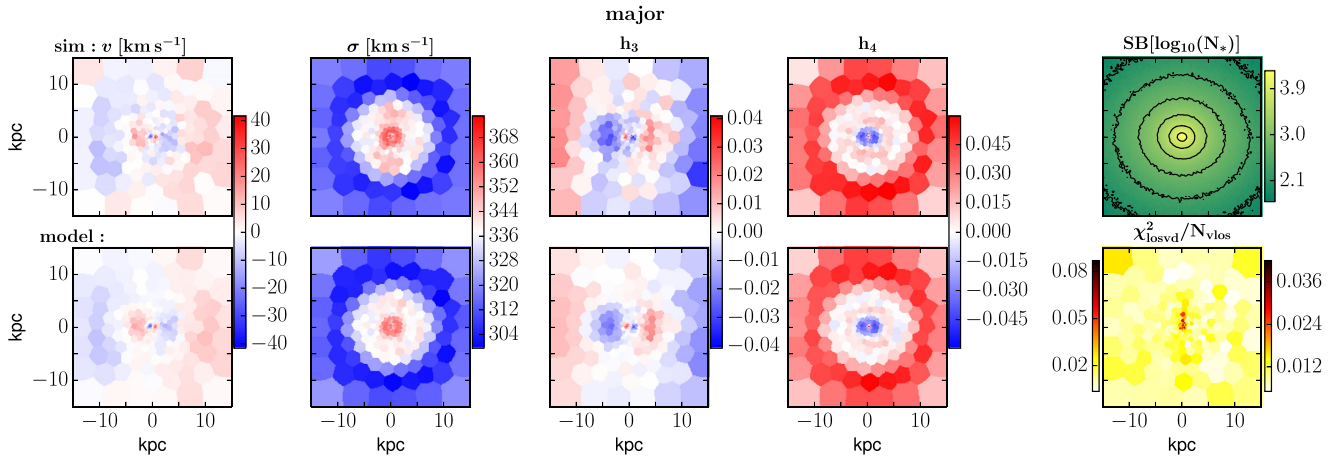


Figure 3. Velocity maps of the simulation (top row) and of the model (bottom row) for the major axis projection, i.e. the projection with the line of sight being parallel to the major axis of the simulation based on the simulation’s orientation described in Section 3.1.1 (for the other projections see Fig. B1). The different panels show the velocity in km s^{-1} (first column), velocity dispersion in km s^{-1} (second column), the h_3 -parameter (third column), and h_4 -parameter (fourth column) plotted over the whole field of view. The fifth panel in the top row shows the corresponding surface brightness map from the simulation in units of logarithmic numbers of stellar particles N_* . The contour lines correspond to isodensity surfaces. The fifth panel in the bottom row shows $\chi^2_{\text{losvd}}/N_{\text{vlos}}$ as deviation from the kinematic input data with the modelled fit. We show the result for the model with the correct stellar mass-to-light ratio, black hole mass and dark matter scale factor as input parameters evaluated at the most suitable regularization parameter of $\alpha(\min(\text{rms}_\sigma))$. The maps in the second row consist of the results of the modelled left and right half of the galaxy. Therefore, the χ^2 -maps show two colourbars for the two different halves.

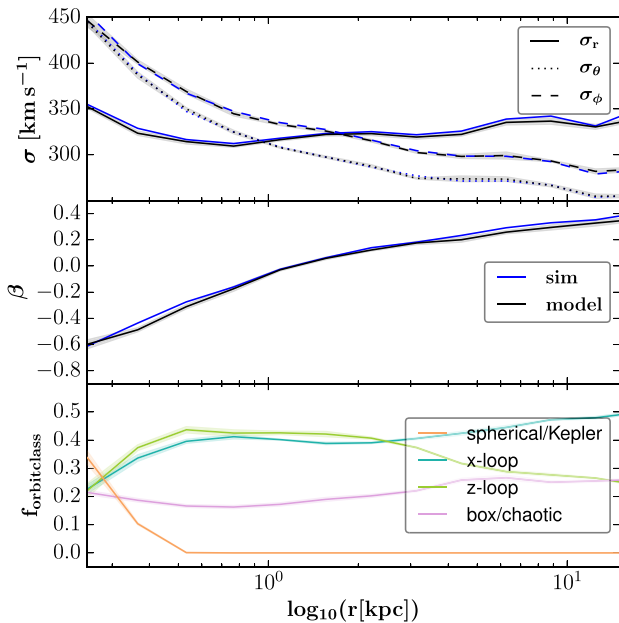


Figure 4. Reproduction of the internal properties within $r \in [0.25 \text{ kpc}, 15 \text{ kpc}]$. Top panel: The internal velocity dispersions in radial direction σ_r (solid line), elevation direction σ_θ (dotted line), and azimuthal direction σ_ϕ (dashed line) from the model (black lines) averaged over the five different projections accurately follow the real ones from the simulation (blue lines) out to 15 kpc as field of view. The grey shaded lines mark the deviations between the different projections. Middle panel: Also, the anisotropy parameter β is well reproduced and represents the tangentially anisotropic orbit distribution ($\beta < 0$) within the core radius $r_b = r_{\text{SOI}} \sim 1 \text{ kpc}$ as well as the radially anisotropic orbit distribution ($\beta > 0$) outside r_b . Bottom panel: Radial distribution of the orbit fractions $f_{\text{orbitclass}}$ classified by SMART.

The following sections will show the 1D mass reproduction results of individually determining the favoured stellar mass-to-light ratio Υ or black hole mass M_{BH} (Section 4.3.1) as well as the 2D mass recovery results of simultaneously determining the favoured Υ and M_{BH} (Section 4.3.2) or Υ and s_{DM} (Section 4.3.3). We skip any 3D mass parameter recovery, since this would not provide more information than the combined 2D recoveries in the context of testing the orbit library.

4.3.1 1D mass recovery of Υ and M_{BH}

Fig. 2 shows the 1D mass recovery results of Υ (third row) and M_{BH} (fourth row) for the five different projections and their respective modelled halves. For testing the recovery of the black hole mass, we provide the model with the correct Υ_{sim} and $s_{\text{DM,sim}}$ values and run nine models with different black hole masses within $M_{\text{BH}} \in [0.79M_{\text{BH,sim}}, 1.21M_{\text{BH,sim}}]$ including the correct one. The tested mass grid has a smaller grid size close to $M_{\text{BH}}/M_{\text{BH,sim}} = 1$ and the exact tested values can be read off from the ordinate ticks in Fig. 2.

For testing the 1D mass recovery of the stellar mass-to-light ratio, we provide $M_{\text{BH,sim}}$ and $s_{\text{DM,sim}}$ and test the same $\Upsilon/\Upsilon_{\text{sim}}$ values as for the black hole mass analysis.

Fig. 2 shows the favoured mass parameters, i.e. the mass parameters where χ^2/N_{data} is smallest, as a function of α . As one can see, for our fiducial choice of α (i.e. $\alpha = \alpha(\min(\text{rms}_\sigma))$ as the best recovery of the velocity dispersions), the average mass recovery performs excellently, with $\Delta M_{\text{BH}} = 5$ per cent and $\Delta \Upsilon = 2$ per cent. In fact, above $\log \alpha \gtrsim -3$, the results are very robust with little dependency on α . Within the overall minor scatter, all models, independent from the chosen half of the galaxy or projection, show equally good fits and reproductions of the internal moments and mass parameters.

4.3.2 2D mass recovery of Υ and M_{BH}

For simultaneously recovering Υ and M_{BH} by SMART, we sample a 2D grid of input masses for 49 models per projection with Υ

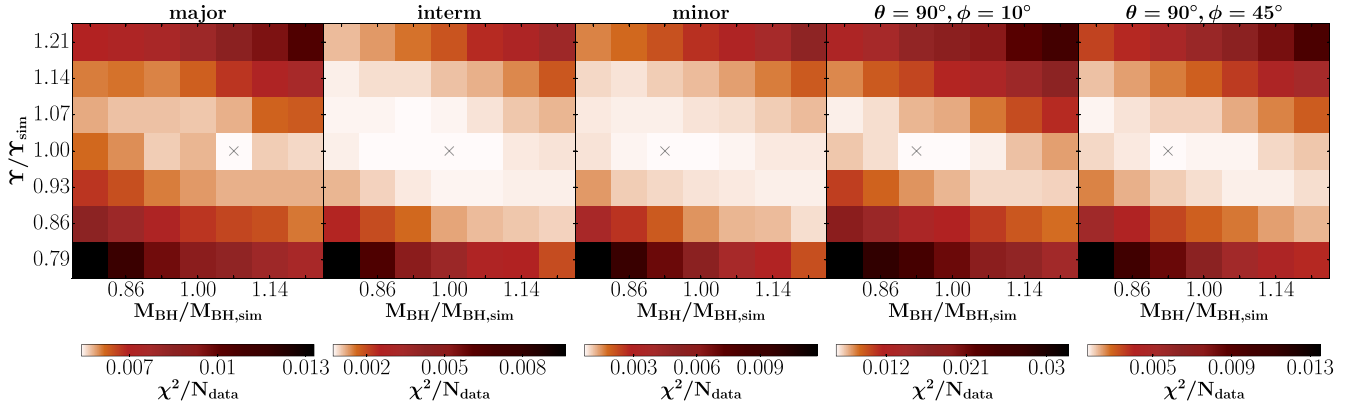


Figure 5. 2D mass recovery results of Υ and M_{BH} for the positive halves of the five different projections. For each projection, we evaluate 49 models with different Υ and M_{BH} input masses, covering a 2D grid with a step size of 7 per cent around the correct mass parameter. Each plot contains the χ^2/N_{data} colourbar for the individual projection. The favoured models are marked with a grey cross. The model always finds the correct stellar mass-to-light ratio and the black hole mass with a minor averaged deviation of 6 per cent.

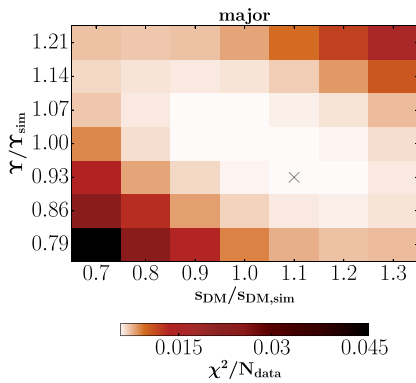


Figure 6. 2D mass recovery results of Υ and s_{DM} for the positive half of the major axis projection. s_{DM} thereby is the mass multiplication scale factor of the pre-determined halo shape of the simulation. We evaluate 49 models with different Υ and s_{DM} input masses, covering a 2D grid with a step size of 7 per cent for the stellar mass-to-light ratio and 10 per cent for the dark matter scaling multiplication factor. The favoured model is marked with a grey cross. The model slightly overestimates the dark matter halo scale factor by 10 per cent and slightly underestimates the stellar mass-to-light ratio by 7 per cent.

$\in [0.79\Upsilon_{\text{sim}}, 1.21\Upsilon_{\text{sim}}]$ and $M_{\text{BH}} \in [0.79M_{\text{BH,sim}}, 1.21M_{\text{BH,sim}}]$. We again model all five different projections but fit only the right half of the galaxy since the 1D mass recovery showed no significant difference between the respective halves of each projection. The results are plotted in Fig. 5. Evaluated at the regularization value $\alpha(\min(\text{rms}_\sigma))$ the stellar mass-to-light ratio is in every case correctly recovered and the black hole mass is reproduced with an accuracy of 6 per cent averaged over the different projections.

4.3.3 2D mass recovery of Υ and s_{DM}

For recovering Υ and s_{DM} , we sample a 2D grid of input masses for 49 models along the major axis projection with $\Upsilon \in [0.79\Upsilon_{\text{sim}}, 1.21\Upsilon_{\text{sim}}]$ and $s_{\text{DM}} \in [0.7s_{\text{DM,sim}}, 1.3s_{\text{DM,sim}}]$. We here model the right half of the major axis projection. The result is shown in Fig. 6. Evaluated at $\alpha(\min(\text{rms}_\sigma))$, the dark matter scale factor is slightly overestimated by 10 per cent and the stellar mass-to-light ratio is

slightly underestimated by 7 per cent, confirming the accurate mass reconstruction from the previous tests.

In conclusion, these mass recovery results demonstrate that our orbit sampling and superposition algorithms allow for a very accurate reconstruction of the mass composition and orbital structure of triaxial systems with known density shapes.

4.4 Beyond second-order velocity moments

So far, we have focused on the reproduction of the first- and second-order internal velocity moments. The previous sections have shown that the full shape of the LOSVDs contains enough information to accurately reconstruct the mass and the anisotropy structure of the orbit distribution. In Fig. 7, we show the full mass-weighted stellar velocity distributions per velocity bin against the velocity in km s^{-1} in radial, longitudinal, and azimuthal direction when integrating over the other velocity components and respective spatial bins of the model. We find that the central azimuthal and longitudinal velocity distributions within $r < r_{\text{SOI}}$ have two maxima that become more pronounced closer to the centre (see Fig. 7). This likely reflects the strong tangential anisotropy produced during the formation of the core and is probably linked to the negative h_4 parameter at the centre of the merger remnant (see Figs 3 and B1), which will be investigated in more detail in a separate paper. The whole internal velocity distributions contain more information about the formation process than the velocity moments alone can do. So, we extended SMART to calculate the internal velocity distributions for 31 velocity bins within the positive and negative escape velocity, i.e. $v_{\text{min}}^{\text{max}} = \pm v_{\text{esc}}(r, \theta, \phi)$, evaluated at each radial and angular bin of the SMART-specific grid (see Section 2.1). The internal velocity distributions averaged within spherical shells reproduce the ones from the simulation sufficiently well, with a deviation of $\text{rms} = 0.07$ averaged over all velocity bins with $v < 1000 \text{ km s}^{-1}$ and radial bins within $r \in [0.25 \text{ kpc}, 15 \text{ kpc}]$ (see Fig. 7). Including the outer wings of the internal velocity distributions with $v > 1000 \text{ km s}^{-1}$, the rms increases, however, the number of simulation particles in these bins is very small. We will apply this ability of SMART to model the whole internal velocity distribution in future studies of real observational data.

We checked the behaviour of the rms profile in dependence of the regularization as deviation of the whole internal velocity distributions and compared it with the rms_σ and $\text{rms}_{v,\sigma}$ profiles. It thereby showed

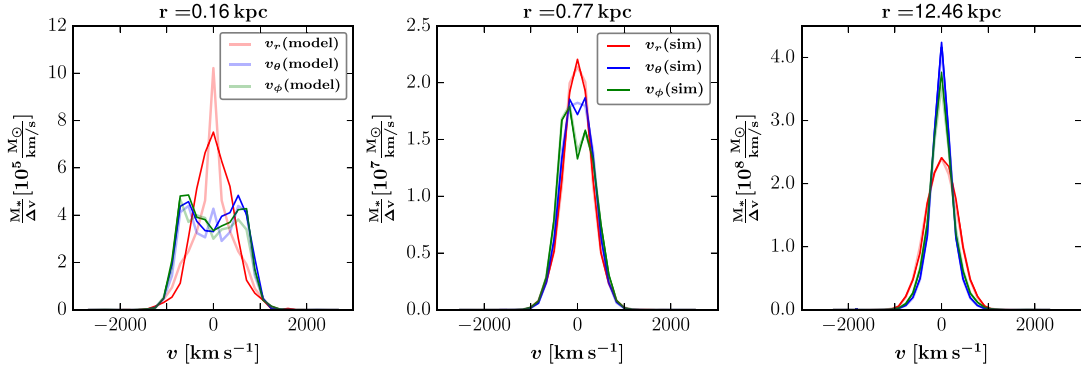


Figure 7. Internal velocity distributions in radial (red), elevation (blue), and azimuthal (green) direction when integrating over the respective spatial bin of the model and the other velocity components for three different radii of $r = 0.16$ kpc (left-hand panel), $r = 0.77$ kpc (middle panel), and $r = 12.46$ kpc (right-hand panel). The internal velocity distributions are calculated for 31 velocity bins within the positive and negative escape velocity evaluated at each radial and angular bin of the SMART-specific grid. We here show the mass-weighted stellar internal velocity distributions per velocity bin averaged within spherical shells. The non-shaded lines correspond to the simulation data and the shaded lines show the modelled results. In closer vicinity of the black hole, the elevation and azimuthal distribution show two maxima corresponding to the tangentially anisotropic orbit distribution. SMART is able to reproduce the internal velocity distributions in general and is also able to follow this specific behaviour in the central bins.

the same form and its minimum appeared in the same regularization region and therefore does not provide additional information when determining the most suitable regularization value.

4.5 Robustness and uniqueness checks

In order to test the robustness of the results against modifications of our fiducial set-up and in order to test the uniqueness of the results, we make the following checks and considerations:

4.5.1 PSF convolution and noise

While Valluri et al. (2004) describe that three-integral, axisymmetric, orbit-based modelling algorithms in general show a flat-bottomed χ^2 distribution, being unable to determine the black hole mass to better than a factor of ~ 3.3 , Magorrian (2006) demonstrates that this is only true for noiseless data. According to this, our model should not be able to precisely determine the correct black hole mass of the simulation due to the lack of an ‘error’. However, the previous results have already shown that SMART achieves a well-defined black hole mass due to a well-defined minimum in the χ^2 -profile, which was probably supported by the intrinsic noise of the N -body simulation. Nevertheless, we check whether the minimum in the χ^2 -curve changes when simulating an ‘error’ in the kinematic input data. For this, we model the positive half of the intermediate axis projection by adding Gaussian noise to the simulation chosen so that the velocity dispersion of the noisy kinematic simulation data results in an observationally realistic error of ~ 3 per cent ($\bar{v} = (0.14 \pm 7.64) \text{ km s}^{-1}$, $\bar{\sigma} = (288.86 \pm 7.80) \text{ km s}^{-1}$, $\bar{h}_3 = 0.00 \pm 0.02$, $\bar{h}_4 = 0.02 \pm 0.02$). To achieve even more realistic conditions we furthermore smooth the data by simulating a psf convolution with a FWHM of 2.43 arcsec which corresponds to 0.24 kpc, i.e. about a fourth of the sphere of influence, when assuming the galaxy to be at a distance of 20 Mpc. The so constructed velocity maps can be seen in Fig. B2. We provide SMART with the information about the used FWHM-value for the 2D psf convolution and test the same Υ and M_{BH} input masses as in Section 4.3. The corresponding results for the 1D mass recoveries when modelling this modified input data are plotted as turquoise lines in Fig. 8, and do not differ decisively from the ones

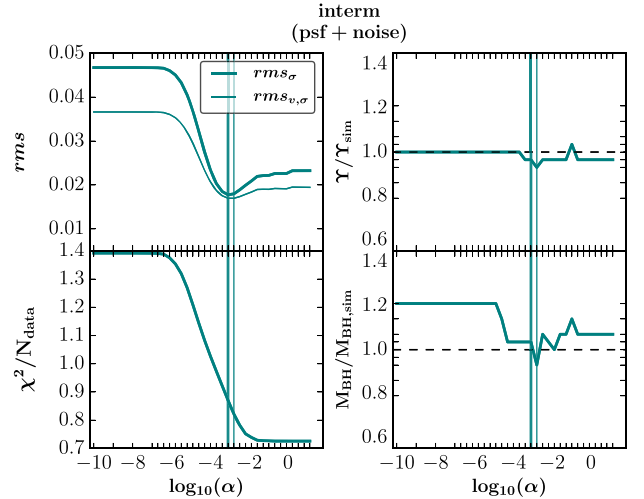


Figure 8. Choice of regularization and 1D mass recovery results for the positive half of a psf convolved noisy version of the intermediate axis projection. The rms and χ^2/N_{data} profiles show the same shape and suitable regularization region than for the noiseless and non-psf-convolved case. Also, the Υ - and M_{BH} reproduction shows no remarkable change within the overall scatter.

without psf convolution and noise. The stellar mass-to-light ratio is slightly underestimated by $\Delta\Upsilon(\alpha(\min(\text{rms}_{\sigma}))) = 3.5$ per cent or $\Delta\Upsilon(\alpha(\min(\text{rms}_{v,\sigma}))) = 7$ per cent and the black hole mass is overestimated by $\Delta M_{\text{BH}}(\alpha(\min(\text{rms}_{\sigma}))) = 3.5$ per cent or underestimated by $\Delta M_{\text{BH}}(\alpha(\min(\text{rms}_{v,\sigma}))) = 7$ per cent. The rms_{σ} , $\text{rms}_{v,\sigma}$, and χ^2/N_{data} profiles are of course shifted upwards but follow the same form than the ones without noise and psf convolution (cf. Fig. 2). Fig. 9 shows the 2D mass recovery result of Υ and M_{BH} for this model. The stellar mass-to-light ratio is again correctly reproduced and the black hole mass is underestimated by only 7 per cent. With this, SMART is able to model the noisy and psf convolved kinematics as well as the default kinematics without any ‘error’ equally well within the overall scatter.

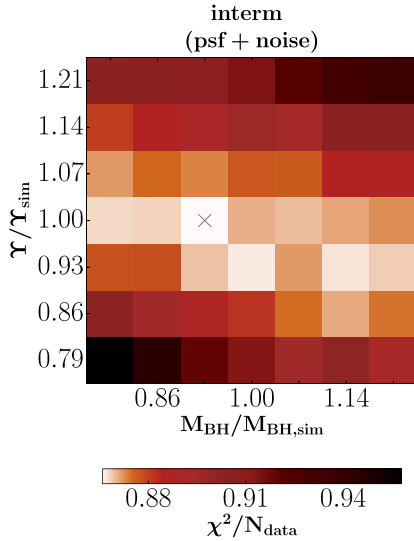


Figure 9. 2D mass recovery results of Υ and M_{BH} for the positive half of a psf convolved noisy version of the intermediate axis projection. We again evaluate 49 models with different Υ and M_{BH} input masses, covering a 2D grid with a step size of 7 per cent around the correct parameters. The favoured model is marked with a grey cross. The model finds the correct stellar mass-to-light ratio and slightly underestimates the black hole mass by 7 per cent. Within the overall scatter this resembles the results of the models without noise and psf convolution.

4.5.2 Changing the orbital bias factors ω_i

As discussed in Section 2.4, the orbital bias factors ω_i of equation (8) can be used to control the orbital weights w_i . In the absence of other constraints $w_i \sim \omega_i$. Our choice of $\omega_i = 1$ is somewhat arbitrary. In fact, it biases the w_i s strongly away from the true solution. Thus, we want to test whether this affects our fits. Specifically, we test the opposite extreme. We remodel the positive half of the major axis, abbreviated below as X^+ , by setting $\omega_i = w_i(X^+)$, where $w_i(X^+)$ are the orbital weights of the best-fitting model for X^+ . We also remodel X^+ by setting the bias factors $\omega_i = w_i(X^-)$ to the orbital weights of the best-fitting model for the negative half of the major axis X^- . Fig. 10 shows the results for these completely independent model fits. As expected, the rms and χ^2/N_{data} profiles start with smaller values, since at $\alpha = 0$, the ω_i s bias the orbital weights the strongest. As motivated above, in the specific case here, the weights are biased towards a previous fit, which explains the better initial χ^2 and rms . However, at our fiducial α range, the reproduction of the internal moments and the quality of the fit is the same as for the case with identical ω_i s. As a consequence, this implies that the simplifying assumption of constant ω_i does not change the modelling results significantly.

4.5.3 Degeneracy

When redoing the just described analysis (Section 4.5.2) but using the orbital bias factors of the right half of the *intermediate* axis projection $\omega_i(Y^+)$ (brown line in Fig. 11) as initial values for remodelling X^+ we gain a minor improvement in the reconstruction of the internal moments since the rms_σ and $rms_{v,\sigma}$ values are a bit smaller compared to the default model (black line), though the mass recovery shows equal results. The fact that, for the same quality of fit (i.e. same χ^2), the rms of the internal moments is smaller for the model with ω_i assumed to be equal to orbital weights from another modelled

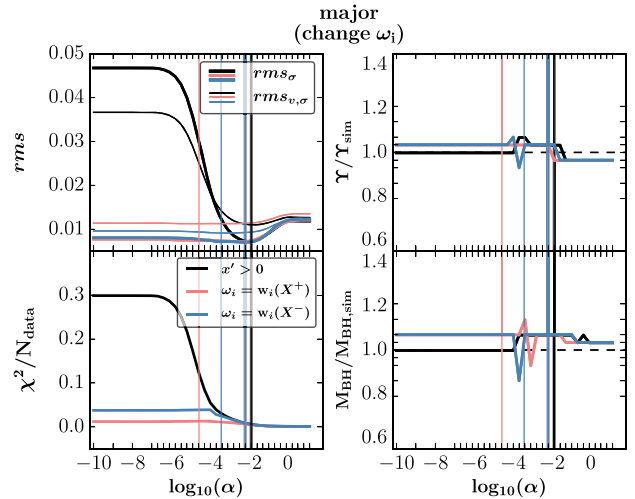


Figure 10. Choice of regularization and 1D mass recovery results for the positive half of the major axis projection when using the default constant orbital bias factors $\omega_i = \text{const.}$ (black line), when using the best-fitting orbital weights $w_i(X^+)$ of the positive half of the major axis as orbital bias factors, i.e. $\omega_i = w_i(X^+)$ (pink line), and when using $w_i(X^-)$ of the negative half of the major axis as orbital bias factors, i.e. $\omega_i = w_i(X^-)$ (blue line). The modified orbital bias factors do not improve the results, indicating that the choice of $\omega_i = \text{const.}$ is sufficient.

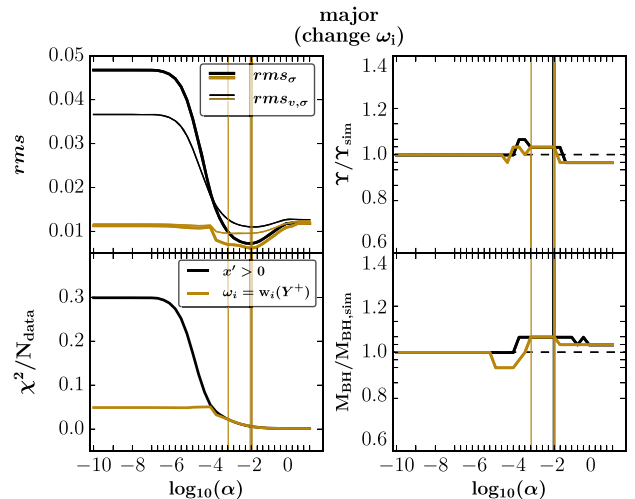


Figure 11. Choice of regularization and 1D mass recovery results for the positive half of the major axis projection when using the default constant initial orbital bias factors $\omega_i = \text{const.}$ (black line) and when using the best-fitting orbital weights $w_i(Y^+)$ of the positive half of the intermediate axis as orbital bias factors, i.e. $\omega_i = w_i(Y^+)$ (brown line). The additional information from the second projection axis appears to reduce the degeneracy leading to a minor improvement in the internal moments reproduction as seen in the lower rms_σ and $rms_{v,\sigma}$ values.

projection direction (in this case from the intermediate axis projection Y) suggests that these ω_i contain some information in addition to the kinematics of the given projection (in this case the major axis projection X). The fact that information from another line of sight can improve the model is not surprising. In fact, it shows that some of the already very small residual rms in the internal velocity moments is due to remaining degeneracies in the recovery of the

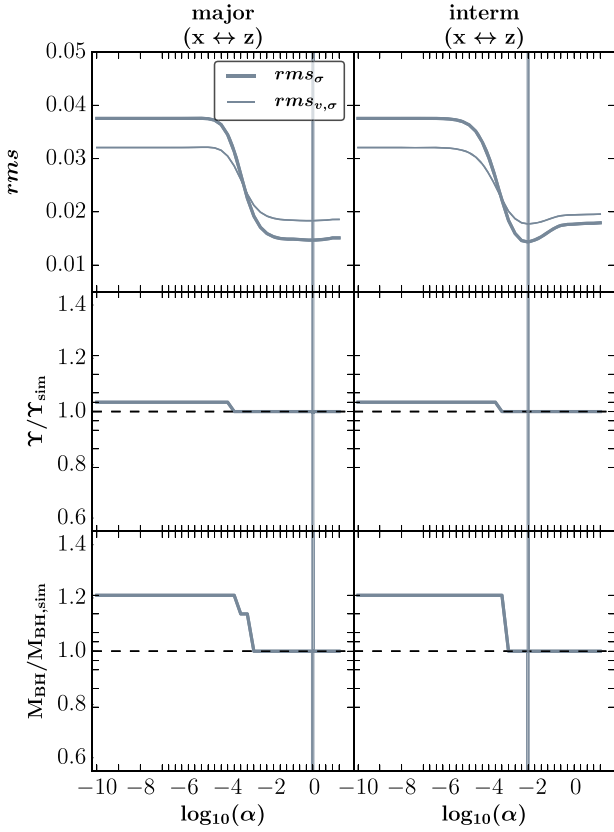


Figure 12. Choice of regularization and 1D mass recovery results for the positive half of the major axis projection (left column) and positive half of the intermediate axis projection (right column) when changing the x - and z -coordinates in the simulation so that the orbit sampling in SMART changes from a sampling of v_ϕ to v_θ . Whereas the mass recovery gets slightly improved, the rms_σ and $\text{rms}_{v,\sigma}$ profiles show slightly deteriorated values. Overall, both techniques basically lead to the same results.

orbital weights. These degeneracies seem to be surprisingly small. However, our results imply that as long as the deprojected light profile and normalized DM halo are known and the orbit sampling is dense, masses and anisotropies can be recovered with very high accuracy, independent of the viewing angle.

4.5.4 Change in orbit sampling technique

Even though our results already prove the efficient ability of our orbit library to completely reproduce all necessary orbits in a triaxial potential, we want to check the robustness of our orbit sampling method (cf. Appendix C). We therefore change the angular momentum sampling direction from the minor axis to the major axis. This corresponds to a change of v_ϕ to v_θ samplings. The basic idea is that a more homogeneous angular momentum sampling along the major axis instead of the minor axis might improve the mass recovery. Thus, we change the major and minor axis coordinates within the simulation and rerun the models for the new major- (left column in Fig. 12) and intermediate-axis-kinematics (right column in Fig. 12). With this, the angular momentum orbit sampling is proceeded along the major axis of the simulated remnant. The results are shown in Fig. 12. One can see that both orbit libraries produce the same results. In fact, the modified orbit sampling procedure reveals a slightly better mass recovery but slightly worse internal moments reproduction

(higher rms_σ and $\text{rms}_{v,\sigma}$ values) for equally good kinematic fits. Nevertheless, regardless of the chosen angular momentum sampling axis our orbit sampling technique (cf. Appendix C) of creating initial conditions, which belong to certain energy shells and angular momentum sequences, has proven to be a highly efficient technique. It produces a general and complete set of orbits for triaxial potentials being able to deal with changing structure in the integrals-of-motion space since it manages to reproduce all relevant properties and to radially adapt itself to the more spherical centre as well as the more prolate outskirts of the simulated galaxy.

4.5.5 Summary of robustness and uniqueness checks

In conclusion, these checks prove that SMART is robust against minor internal modifications as well as input data changes. SMART has proved its ability to handle with noisy and psf convolved data. Furthermore, we have shown that constant orbital bias factors are a good approximation. Even though it is impossible to get observations from two viewing points, we have demonstrated that this would allow to reduce the minor degeneracies allowed by the kinematic data even further. We have verified this by using the orbital bias factors from a second projection direction. Moreover, the results are not affected by changing the orbit sampling technique from setting up L_z sequences to setting up L_x sequences. This shows that, as expected, the choice of sampling axis is not decisive and that the orbit sampling routine is universal.

5 THE QUASI-UNIQUENESS OF THE ANISOTROPY RECONSTRUCTION WHEN FITTING FULL LOSVDs

The results of the previous section have shown that the anisotropy of the N -body merger remnant can be reconstructed with very high accuracy from the Schwarzschild models that we fitted to the full LOSVDs. This not only demonstrates the high accuracy of our orbit superposition model, but it also implies that when models can exploit the full information contained in the entire LOSVDs, the remaining degeneracy in the recovery of the distribution function cannot affect the anisotropy or mass recovery significantly. In this section, we want to use the maximum entropy technique to explore this in more depth.

5.1 The Maximum Entropy Technique and the Mathematical Structure of the Solution Space

As already described in Section 2.4, we solve for the orbital weights w_i by using a maximum entropy technique (cf. equations 7 and 8). The χ^2 term in equation (7) contains the kinematical constraints and is the deviation between observed LOSVDs $\mathcal{L}_{\text{data}}$ and the model prediction \mathcal{L}_{mod} , i.e. the weighted sum over the contributions of all orbits to LOSVD j' and line-of-sight velocity bin k ,

$$\mathcal{L}_{\text{mod}}^{j',k} \equiv \sum_{i=1}^{N_{\text{orbit}}} w_i \mathcal{L}_{\text{orb}}^{j',k,i}. \quad (15)$$

It is convenient to think of the observed $\vec{\mathcal{L}}_{\text{data}}$ and the model predictions $\vec{\mathcal{L}}_{\text{mod}}$ as vectors with $N_{\text{data}} = N_{\text{losvd}} \times N_{\text{vlos}}$ elements. Then,

$$\vec{\mathcal{L}}_{\text{mod}} = \mathcal{L}_{\text{orb}} \cdot \vec{w}, \quad (16)$$

where \mathcal{L}_{orb} is a matrix with N_{data} rows and N_{orbit} columns.

In addition to the kinematical observations $\vec{\mathcal{L}}_{\text{data}}$, the orbital weights \vec{w} are subject to photometric constraints. In analogy to

equation (16):

$$\vec{p}_{\text{mod}} = P_{\text{orb}} \cdot \vec{w}, \quad (17)$$

where \vec{p}_{mod} is a vector with the model predictions for the 3D luminosity density at spatial position j_3 in the galaxy ($j_3 = 1, \dots, N_{\text{phot}}$). To guarantee the self-consistency of our model, the respective observed \vec{p}_{data} are not included via a χ^2 term. Instead, we treat them as boundary conditions for the fit:

$$\vec{p}_{\text{data}} \stackrel{!}{=} \vec{p}_{\text{mod}}. \quad (18)$$

Hence, we seek for the maximum of equation (7) subject to the linear equality constraints

$$\vec{p}_{\text{data}} - P_{\text{orb}} \cdot \vec{w} = 0. \quad (19)$$

For convenience, we normalize the \vec{p}_{data} such that $\sum \vec{p}_{\text{data}} = 1$. Since we are only interested in positive orbital weights (see below), the orbital weights obey $0 \leq w_i \leq 1$ and we can restrict the maximization to the respective N_{orbit} -dimensional convex quader.

To show that equation (7) has a unique global maximum that can be controlled through the bias factors ω_i it is convenient to consider the equivalent minimization problem for $f \equiv -\hat{S}$, given by multiplying equation (7) with -1 .

Let us first consider equation (7) without the entropy term S . The χ^2 term can be written as

$$\chi^2 = -\vec{\mathcal{L}}_{\text{data}}^T C_v \vec{\mathcal{L}}_{\text{data}} + 2\vec{\mathcal{L}}_{\text{data}}^T C_v \mathcal{L}_{\text{orb}} \vec{w} - \vec{w}^T \mathcal{L}_{\text{orb}}^T C_v \mathcal{L}_{\text{orb}} \vec{w}, \quad (20)$$

where C_v is the covariance matrix of the observed LOSVDs and is positive definite. The Hesse matrix of χ^2 reads

$$\nabla^2 \chi^2 = 2\mathcal{L}_{\text{orb}}^T C_v \mathcal{L}_{\text{orb}}. \quad (21)$$

Because \mathcal{L}_{orb} is positive by construction, the symmetric matrix $\nabla^2 \chi^2$ is at least positive semidefinite and χ^2 is convex.

The minimization of χ^2 alone, subject to the linear equality constraints equation (19), is therefore a convex optimization problem with affine equality constraints in standard form. As such, it only has a global minimum (e.g. Boyd & Vandenberghe 2004). In general, we cannot assume that χ^2 is *strictly* convex (i.e. that $\nabla^2 \chi^2$ is positive definite). The set of orbital weights that solve $\chi^2(\mathbf{w}) = \chi_{\text{min}}^2$ may therefore be non-unique. This is not surprising given that the linear equation

$$\vec{\nabla} \chi^2 = 2\vec{w}^T \mathcal{L}_{\text{orb}}^T C_v \mathcal{L}_{\text{orb}} + 2\vec{\mathcal{L}}_{\text{data}}^T C_v \mathcal{L}_{\text{orb}} \equiv 0 \quad (22)$$

will in general be underconstrained if $N_{\text{orbit}} > N_{\text{data}}$. As already mentioned above, the reconstruction of the distribution function is not unique even if the deprojected density and the potential are known. We give an example in Appendix D.

Because the Hesse matrix of $-S$ is diagonal with the i th element equal to $1/w_i$, the entropy $-S$ is *strictly* convex. In contrast to the case of χ^2 , the set of orbital weights that minimize $-S$ (or, equivalently, that maximise S) is *unique*. For the entropy alone this is easy to see since

$$\nabla S = -\log \frac{w_i}{\omega_i} - 1 \equiv 0 \quad (23)$$

can be solved analytically: $w_i = \exp(-1) \cdot \omega_i$. The constraints from equation (19) will shift the solution, but the strict convexity still guarantees it to remain unique (e.g. Boyd & Vandenberghe 2004).

For general $\alpha > 0$, the Hesse matrix of f is the sum $\nabla^2(-S) + \alpha \nabla^2 \chi^2$ and f is always strictly convex. Equation (7) subject to the constraints (19) hence always has a unique solution.

As already mentioned above, from an algorithmic point of view, it is advantageous to have a unique solution. From the physical point of

view this is not desirable because any algorithm that picks up only one of the potentially many solutions that minimize χ^2 in equation (6) may lead to a *bias*. However, suppose \vec{s}_1 and \vec{s}_2 are two solutions which lead to the same χ_{min}^2 . By setting $\omega_i = s_{1,i} \cdot \exp(1)$ we can make \vec{s}_1 the global solution of equation (7). Likewise, by setting $\omega_i = s_{2,i} \cdot \exp(1)$ we can make \vec{s}_2 the global solution of equation (7). This shows that the maximum-entropy formulation of the problem allows in principle the reconstruction of the entire solution space (via variation of the ω_i).

5.2 Testing the uniqueness of the anisotropy recovery

The previous section has shown that the recovery of the distribution function (or, equivalently, of the orbital weights) is in general not unique, even when the deprojection and the potential are known. The maximum entropy technique recovers one of the many possible solutions, which is unique for every given set of ω_i . Variation of the ω_i allows to sample the full solution space. On the other hand, the fits of the N -body merger remnant have shown that the anisotropy recovery is very accurate and stable even to variations of the ω_i . Moreover, in Appendix D we explicitly construct two different phase-space distribution functions that fit a given set of kinematics equally well. Even though they have different orbital weights, they reveal very similar anisotropies in the second-order velocity moments. This suggests that while the recovery of the full distribution function is non-unique, the anisotropy of the second-order moments is actually very well constrained by the information contained in the full LOSVDs.

To investigate this further, we create kinematic data of several toy models with different intrinsic properties: with randomized orbit weights (called RANDOM in the following), with an overpopulation of box/chaotic orbits (called BOX), with an overpopulation of z -tubes (called ZTUBE), with an overpopulation of only the prograde z -tubes (called ZROT) and with an overpopulation of x -tubes (called XTUBE). We then fit the mock kinematic data of these toy models under different choices for the ω_i , trying to push the fitted orbit model towards extreme shapes. The goal is to test how accurate and stable the recovery of the internal velocity anisotropy is, when we fit the entire information contained in the LOSVDs.

The toy models are constructed as maximum-entropy models [i.e. through maximization of equation (7) with $\alpha = 0$; cf. Thomas et al. 2007]. For the BOX model, we increase the ω_i of all box orbits by a factor of 1000, for the XTUBE and ZTUBE models we increase the ω_i of the respective tube orbits by the same amount. For the ZROT toy model we increase the ω_i only for prograde z -tubes and for the RANDOM model we use randomized ω_i (cf. Appendix D). The weights are then still forced to satisfy the density constraints of the N -body simulation. For each toy model, we create kinematic mock data for four different projections (major-, intermediate-, minor-axis projections and the $\theta = 45^\circ$, $\phi = 45^\circ$ projection). We then model every projection of all toy models (in total 20 different input data) and test four different methods for the fits: We use (i) our default constant orbital bias factors, i.e. $\omega_i = 1$ corresponding to the Shannon-entropy (abbreviated as ‘shannon’ in Fig. 13), (ii) increased orbital bias factors by a factor of 10 for the box/chaotic orbits (box10), (iii) increased orbital bias factors by a factor of 10 for the z -tubes (ztube10), and (iv) increased orbital bias factors by a factor of 10 for the x -tubes (xtube10).

Fig. 13 shows the resulting anisotropy profiles (when averaging over shells) for these models. The different rows correspond to the different input toy models and the different columns show the individual projection directions. The dotted data points symbolize the

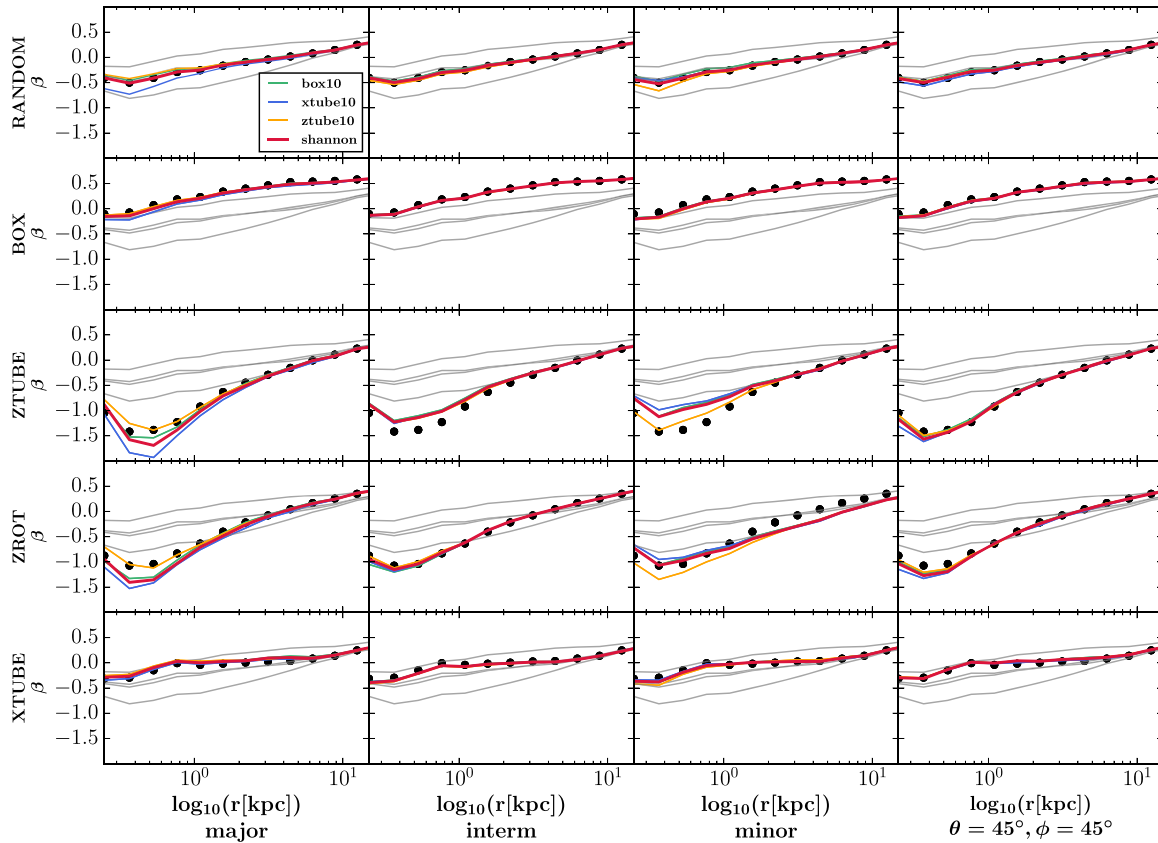


Figure 13. Recovery of the anisotropy profiles of different input toy models by SMART fits with different entropy functions. The input models (black dots) are constructed to have specific orbit classes overrated, resulting in different anisotropy profiles (the overrated orbit type is labelled on the y-axis of each individual row, details in the text). All the five toy models can be well recovered. This is independent of the choice of the projection axis (different columns; from left to right major, intermediate, minor and a diagonal axis) and from the assumed entropy function in the fit (coloured lines). For example, the anisotropy profile of the toy model with an overpopulation of box/chaotic orbits (called BOX; second row) is well reproduced by models maximizing the Shannon-entropy (red lines, labelled shannon) but also with other entropy functions that use a 10 times higher bias factor for box-orbits (green, labelled box10), for x -tubes (blue, xtube10) or for z -tubes (orange, ztube10). The grey lines correspond to the anisotropies implied by maximizing these four different entropy functions without fitting the kinematic data. They symbolize the variety of anisotropy profiles that are in principle possible for different choices of ω_i . After fitting the kinematic data, the average deviation (averaged over all radii and toy models fitted with different entropy techniques) between recovered and input anisotropy is very small, $|\Delta\beta| = 0.05$.

anisotropy profiles of the input toy models (which are of course the same for the different projections). The coloured lines (green, blue, orange, red) show the recovered anisotropies of SMART fits using entropy functions with enhanced bias factors for specific orbit types as described under (i)–(iv) above. The grey lines, for comparison, show the anisotropies that result when we maximize the above entropy functions (i)–(iv) without fitting the mock LOSVDs of the toy models. The grey lines therefore illustrate the variety of different anisotropy profiles that can be constructed by varying the orbital bias factors ω_i . They also indicate the range of different anisotropy profiles that are consistent with the given density distribution.

As one can see, even though the entropy functions tend to push the fits into extreme directions, the range of anisotropy profiles recovered after the fit to the LOSVDs is very narrow. When averaging over all radii and toy models fitted with different entropies, the mean deviation to the input models is $|\Delta\beta| = 0.05$. The average spread in beta inside the sphere of influence is slightly larger $|\Delta\beta(r < r_{\text{SOI}})| = 0.09$ than $|\Delta\beta(r_{\text{SOI}} < r < r_{\text{FOV}})| = 0.03$ outside r_{SOI} . One possible explanation for this might be the increase of degrees of freedom of spherical orbits near the centre. In addition, because β involves the ratio of the intrinsic dispersions, the same fractional error in the intrinsic dispersions results in a four times larger $|\Delta\beta|$

when the anisotropy is as tangential as $\beta = -1.5$ compared to the isotropic case. Overall, the Shannon entropy (red line) is able to recover the beta anisotropy best. This is our default entropy used in SMART.

These results together with Section 4 strongly suggest that the information contained in the full LOSVDs constrains the anisotropy in the second-order velocity moments very well. In turn, this is the reason why our models can reproduce the mass of the black hole and of the stars in the N -body simulation very well.

At larger radii, solely the reconstruction of the intrinsic anisotropy of the toy model with enhanced prograde z -tubes (i.e. ZROT) turns out to be difficult when viewed along the minor axis. This, however, is expected, since any rotation around the minor-axis cannot be observed and, thus, not be reconstructed from this viewing direction. Since we use equal ω_i for prograde and retrograde orbits in our *fitted* models, these models do not have intrinsic rotation in the z -tubes for this projection. Consequently, the tangential velocity dispersion is larger than in the toy model and the fitted β becomes too negative. We checked that if we use the true second-order velocity moment rather than the velocity dispersion in the tangential direction, then the differences between the outer profiles of the ZROT model and the fits along the minor axis disappear.

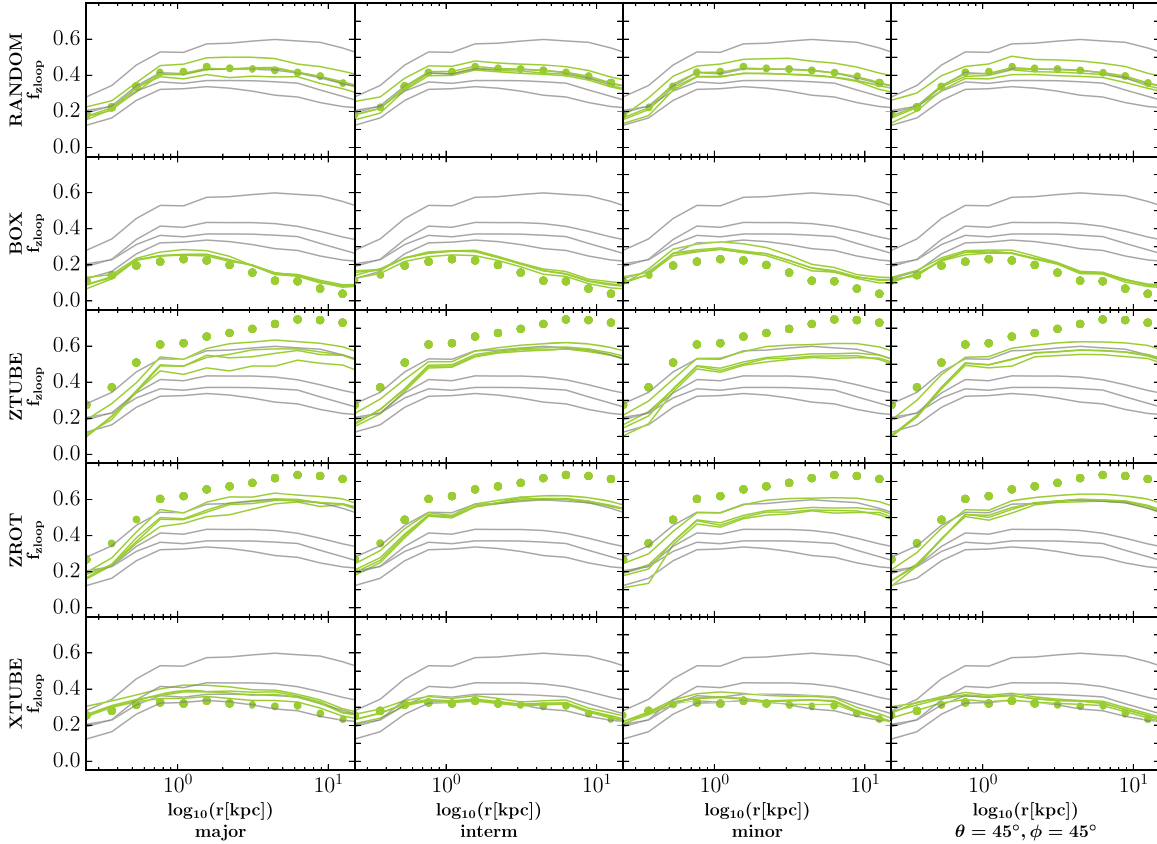


Figure 14. Recovery of the z -loop orbit fractions of different input toy models by models with different entropy methods. We here show the same analysis as in Fig. 13 but now for the reconstruction of the fraction of orbits classified as z -tubes. The colour coding is adapted to Figs 4 and C1. Independent of the tested projection, the z -loop fractions of the individual input toy models are well recovered by the models using different entropy methods. The same is true for the other orbit class fractions (see Figs E1–E3).

For dynamical models which aim for a full phase-space reconstruction (like Schwarzschild models) and which use the full information encoded in the LOSVDs (see also Vasiliev & Valluri 2019) the anisotropy should be recoverable with a typical error of $|\Delta\beta| \approx 0.05$. We solely found larger anisotropy discrepancies (up to $|\Delta\beta| = 0.5$) in extremely tangentially biased regions inside the sphere of influence.

As an example of the corresponding accordance of the reconstructed orbit fractions, Fig. 14 shows the case of z -tubes. The intended overpopulation of the z -tubes in the ZTUBE toy model (third row) and the ZROT toy model (fourth row) can be clearly seen in comparison to the other toy models. Independent of the chosen line of sight and entropy method (i.e. the bias factors ω_i), the fraction of z -tubes is qualitatively recovered and follows the enhancement tendencies. The orbit fractions however are less well determined than the anisotropy by the data and show a stronger dependence on the entropy. The same is true for the x -tubes, box/chaotic, and spherical/Kepler orbits shown in Figs E1–E3.

6 INTERMEDIATE AXIS ROTATION

One of the first investigations to check whether tube orbits with net rotation around the intermediate axis are stable in a triaxial ellipsoid was done by Heiligman & Schwarzschild (1979), who studied a triaxial model with fixed axis ratios of 1:1.25:2 by numerical methods and stated that ‘ Y -tube orbits nearly certainly do not exist in the

adopted model’. Also Binney & Tremaine (2008, p. 263) assert that ‘tube orbits around the intermediate axis are unstable’ in a triaxial potential. Adams et al. (2007) analysed the orbit instability of orbits in a triaxial cusp potential, which are initially confined to one of the three principal axes, under a perturbation along the perpendicular direction. They found that orbits around any of the principal axes are unstable to perpendicular motions. However, according to previous results, they again state that orbits around the intermediate axis are more likely to be unstable. This instability is strongest for original box orbits lying in the x - z plane when the axial ratio of these two axes in the original plane is largest.

Our orbit classification routine in SMART finds no y -tubes in the sense that there is no sign conservation of the angular momentum along the intermediate axis over our default integration period of 100 SOS-crossings, agreeing with the aforementioned works done by other groups. However, some of the orbits integrated for the N -body models do show y -rotation for a limited time-span. When providing SMART with artificial projected input LOSVDs that mimic a net rotation along the intermediate axis, the model is able to produce a y -rotation signal of the order of 10 km s^{-1} in the fit. A rotation of this magnitude is small but in principle detectable with today’s telescopes’ resolution. Fig. 15 shows the SMART fit to the major-axis projection of the simulation (cf. Fig. 3) when assuming the viewing angles of the minor-axis projection. With this, we simulate a hypothetical rotation along the intermediate axis. The top row shows the velocity- and h_3 -map when stopping the orbit integration after 2

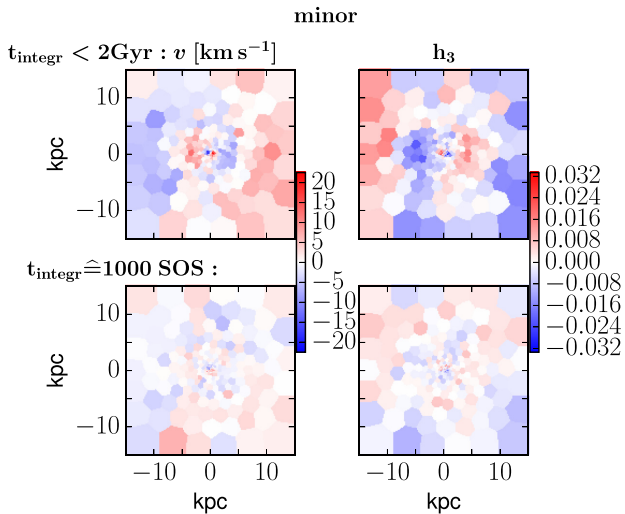


Figure 15. Fitting mock kinematic input data containing intermediate axis rotation. When modelling artificial LOSVDs showing a net rotation along the intermediate axis, SMART is able to reproduce a y -rotation signal when integrating the orbits for a limited integration time of 2 Gyr (first row). The y -rotation becomes visible in the v (left-hand panel) and h_3 (right-hand panel) maps. When integrating for 1000 SOS-crossings (second row), the y -rotation signal cannot be fitted by the model any more.

Gyr if this is shorter than the time needed for 100 SOS-crossings. Indeed, the model reproduces a y -rotation signal. The amplitude of this residual y -rotation becomes smaller and smaller when the orbital integration time is increased, and vanishes when all orbits are integrated for 1000 SOS-crossings (bottom row).

This analysis indicates that the model’s triaxial potential, which is constructed based on the 3D density from the realistic N -body merger simulation – contrary to expectations – contains orbits with y -rotation for a physically relevant time span. What remains unclear at this moment is whether such a y -rotation indeed appears in real elliptical galaxies. If so, then the connection between kinematic misalignments and photometric twists is less constrained than often assumed when only rotation around the intrinsic long and short axes is considered.

7 DISCUSSION

7.1 Remaining sources of systematics

All relevant properties of the simulated merger remnant galaxy were proven to be recovered with a convincing precision and the deviations we found are almost negligible. The remaining deviations (in the $\sim 5 - 10$ per cent level) can be either originated by SMART or the simulation. One remaining contribution to the scatter in the final mass recovery certainly comes from the finite binning resolution of the simulation data and of the SMART models (see Sections 2 and 3). Especially, the need of extrapolating the density towards the centre due to limited resolution of the simulation holds uncertainties.

One more inaccuracy is potentially induced by the softening length in the simulation used to avoid unrealistic two-body encounters between massive particles. Force calculations for radii smaller than the softening length are consequently modified by the softening. In the close vicinity of the black hole, the stellar particles in the simulation are modelled using a non-softened algorithmic chain regularization technique (ARCHAIN; Mikkola & Merritt 2006, 2008) including post-Newtonian corrections (e.g. Will 2006). The

particles outside this chain radius are treated by using softened gravitational force calculations based on the GADGET-3 (Springel 2005) leapfrog integrator. The chain radius r_{chain} is chosen to be at least 2.8 times larger than the GADGET-3 softening length ϵ ($r_{\text{chain}} > 2.8\epsilon$) to ensure that the particles within the chain remain non-softened. The softening length in the simulation for the stellar particles is $\epsilon_* = 3.5$ pc and the softening length for the dark matter particles is $\epsilon_{\text{DM}} = 100$ pc. The mass recovery of the global stellar mass-to-light ratio and dark matter scale factor will be unaffected by these relatively small values. However, there might be a remaining influence on the black hole mass recovery within r_{SOI} . Nevertheless, this effect is expected to be small.

7.2 Comparison to other triaxial Schwarzschild models

We found two other dynamical modelling codes in the literature using Schwarzschild’s orbit superposition technique dealing with triaxiality by van den Bosch et al. (2008) and Vasiliev & Valluri (2019). The code FORSTAND by Vasiliev & Valluri (2019) is applicable to galaxies of all morphological types. When assuming that the deprojection and dark matter halo is known and provided to the code, the models of noise-free axisymmetric disc mock data sets taken from N -body simulations showed very weak constraints on M_{BH} : any value between zero and 5–10 times the true black hole mass was equally consistent with the data. Read from fig. 2 in their paper, the stellar mass-to-light ratio showed a variation of around 20 per cent.

Jin et al. (2019) tested the triaxial Schwarzschild code by van den Bosch et al. (2008) by applying it to nine triaxial galaxies from the large-scale, high-resolution Illustris-1 simulation (Vogelsberger et al. 2014), which provides a stellar and dark matter resolution of $\sim 10^6 M_{\odot}$. When fixing the black hole mass and allowing the model to deproject the mock data set, the stellar mass within an average effective radius is underestimated by ~ 24 per cent and the dark matter is overestimated by ~ 38 per cent. Their averaged model results obtained from mock data with different viewing angles tend to be too radial in the outer regions with better anisotropy matches in the inner region.

Of course, these results cannot be used for direct comparison due to a widely varying resolution and in case of the analysis by Jin et al. (2019) the deprojection probably causes the major deviation. However, SMART for sure is able to add further progress in modelling triaxial galaxies and convinces with its proved precision.

8 SUMMARY AND CONCLUSION

We have developed a new triaxial dynamical Schwarzschild code called SMART and tested its efficiency and reliability by applying it to an N -body merger simulation including SMBHs. The simulation was deliberately selected due to its high accuracy, reasonable formation process, realistic internal structure and ability to precisely calculate the dynamics close to the central black hole. This ensured the possibility to check whether SMART is able to recover all relevant properties including the mass of an SMBH of a realistic triaxial galaxy when providing the deprojected light profile and normalized DM halo.

SMART is assembled with the feature to compute the potential and force by expansion into spherical harmonics, allowing to deal with non-parametric densities and haloes. Its orbit library contains 50000 integrated orbits that are set up by creating random initial radial and velocity values within given energy shells and angular momentum sequences and by filling the surfaces of section. This ensures the ability to adapt itself to a radially changing number of integrals

of motion. The orbit superposition is executed by maximizing an entropy-like quantity and by using the full LOSVDs instead of only Gauss–Hermite parameters alone.

These benefits enable SMART to reconstruct all relevant properties and features of the merger remnant with an excellent precision. We will now recap the requirements that were set on SMART and proved to be fulfilled in this analysis:

(i) SMART is able to reproduce the anisotropy profile and internal velocity dispersions with an rms_σ of only 1.2 per cent.

(ii) SMART reproduces the stellar mass-to-light ratio; black hole mass; and mass scale factor of the dark matter density profile with a precision on the 5–10 per cent level. To our knowledge, this is the first time that the intrinsic precision at given deprojected light profile has been quantified to be so high. This sets the basis for further investigations of the whole modelling procedure.

(iii) SMART well fits the LOSVDs with mean values and deviations of only $\bar{v} = (0.11 \pm 2.09) \text{ km s}^{-1}$, $\bar{\sigma} = (309.75 \pm 2.54) \text{ km s}^{-1}$, $\bar{h}_3 = 0.00 \pm 0.01$ and $\bar{h}_4 = 0.01 \pm 0.01$.

For the determination of these accuracy values, the simulation is modelled from up to five different projections. The mass recovery precision can be achieved for noiseless as well as noisy and psf convolved kinematic input data.

We extensively discuss that the maximum-entropy technique provides an elegant technique to study the range of possible orbit distributions consistent with a given set of data. Our tests with the N -body data and with additional toy models strongly suggest that when the full information contained in the entire LOSVDs is used to constrain the model, then the remaining degeneracies in the recovery of the exact phase-space distribution function do not affect ‘macroscopic’ properties of the galaxy models, like the anisotropy in the second-order velocity moments. This is the basis for the very good reconstruction of the orbital structure and mass of the black hole and stars with SMART.

It was shown that the orbit library is robust against axis changes and generates a complete set of well superpositioned orbits necessary to model a triaxial galaxy with all corresponding internal structures.

Also the accurate mass parameter recovery accomplished by SMART suggests only minor degeneracies contained in the projected kinematic data, provided that the deprojection is known. We showed that these remaining minor degeneracies could in principle be narrowed even more if information about the orbital bias factors from a second projection direction were provided.

When analysing the elevation and azimuthal internal velocity distributions of the simulation we find that the central radial bins show two maxima. This corresponds to the negative h_4 -parameter in the centre and the strong tangential anisotropy produced during the core formation. SMART is able to reconstruct this phenomenon with an accuracy of ~ 7 per cent.

One more discovery of scientific interest is intermediate axis rotation which is produced by orbits contained in the model’s orbit library representing the simulation’s triaxial potential. Independent of the question whether such intermediate axis rotation really appears in the real universe, it was shown that our model contains orbits with y -rotation whose stability was empirically found to be maintained up to at least 2 Gyr.

ACKNOWLEDGEMENTS

We acknowledge the support by the DFG Cluster of Excellence ‘Origin and Structure of the Universe’. The dynamical models have been done on the computing facilities of the Computational Centre

for Particle and Astrophysics (C2PAP) and we are grateful for the support by F. Beaujean through the C2PAP.

DATA AVAILABILITY STATEMENT

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

- Aarseth S. J., Binney J., 1978, *MNRAS*, 185, 227
 Adams F. C., Bloch A. M., Butler S. C., Druce J. M., Ketchum J. A., 2007, *ApJ*, 670, 1027
 Anderson E. et al., 1999, *LAPACK Users’ Guide*, 3rd edn., Society for Industrial and Applied Mathematics, Philadelphia, PA
 Audet C., Dennis J., Jr, 2006, *SIAM J. Optim.*, 17, 188
 Bailin J., Steinmetz M., 2005, *ApJ*, 627, 647
 Barnes J. E., 1992, *ApJ*, 393, 484
 Barnes J. E., Hernquist L., 1996, *ApJ*, 471, 115
 Begelman M. C., Blandford R. D., Rees M. J., 1980, *Nature*, 287, 307
 Bender R. et al., 2005, *ApJ*, 631, 280
 Bender R., 1988a, *A&A*, 193, L7
 Bender R., 1988b, *A&A*, 202, L5
 Bender R., Moellenhoff C., 1987, *A&A*, 177, 71
 Bender R., Surma P., Doebereiner S., Moellenhoff C., Madejsky R., 1989, *A&A*, 217, 35
 Bender R., Burstein D., Faber S. M., 1992, *ApJ*, 399, 462
 Bender R., Saglia R. P., Gerhard O. E., 1994, *MNRAS*, 269, 785
 Bertola F., Galletta G., 1978, *ApJ*, 226, L115
 Bertola F., Galletta G., 1979, *A&A*, 77, 363
 Binney J., Tremaine S., 2008, *Galactic Dynamics*. 2nd edn., Princeton Univ. Press, Princeton, NJ
 Bissantz N., Debattista V. P., Gerhard O., 2004, *ApJ*, 601, L155
 Boyd S., Vandenberghe L., 2004, *Convex Optimization*. Cambridge Univ. Press, Cambridge
 Brainerd W. S., Goldberg C. H., Adams J. C., Garcia A., McKay S., Christian W., 1996, *Comput. Phys.*, 10, 135
 Brown C. K., Magorrian J., 2013, *MNRAS*, 431, 80
 Cappellari M. et al., 2007, *MNRAS*, 379, 418
 Cappellari M. et al., 2012, *Nature*, 484, 485
 Cappellari M., 2002, *MNRAS*, 333, 400
 Cappellari M., 2016, *ARA&A*, 54, 597
 Cappellari M., Copin Y., 2003, *MNRAS*, 342, 345
 Cash J., Karp A., 1990, *ACM Trans. Math. Softw.*, 16, 201
 Chabrier G., 2003, *PASP*, 115, 763
 Cretton N., de Zeeuw P. T., van der Marel R. P., Rix H.-W., 1999, *ApJS*, 124, 383
 de Lorenzi F., Debattista V. P., Gerhard O., Sambhus N., 2007, *MNRAS*, 376, 71
 de Nicola S., Saglia R. P., Thomas J., Dehnen W., Bender R., 2020, *MNRAS*, 496, 3076
 de Zeeuw T., 1985, *MNRAS*, 216, 273
 Dehnen W., 1993, *MNRAS*, 265, 250
 Ebisuzaki T., Makino J., Okumura S. K., 1991, *Nature*, 354, 212
 Emsellem E. et al., 2007, *MNRAS*, 379, 401
 Emsellem E., Monnet G., Bacon R., 1994, *A&A*, 285, 723
 Faber S. M., Dressler A., Davies R. L., Burstein D., Lynden Bell D., Terlevich R., Wegner G., 1987, in Faber S. M., ed., *Nearly Normal Galaxies. From the Planck Time to the Present*, p. 175
 Finozzi F., 2018, *Triaxial Models of Massive Elliptical Galaxies*, PhD thesis. Ludwig-Maximilians-University, Munich
 Franx M., Illingworth G. D., 1988, *ApJ*, 327, L55
 Gebhardt K. et al., 2000, *AJ*, 119, 1157
 Genzel R., Tacconi L. J., Rigopoulou D., Lutz D., Tecza M., 2001, *ApJ*, 563, 527
 Gerhard O. E., 1993, *MNRAS*, 265, 213
 Heiligman G., Schwarzschild M., 1979, *ApJ*, 233, 872
 Henon M., Heiles C., 1964, *AJ*, 69, 73

Hills J. G., Fullerton L. W., 1980, *AJ*, 85, 1281
Hohl F., Zang T. A., 1979, *AJ*, 84, 585
Hopkins P. F., Cox T. J., Hernquist L., 2008, *ApJ*, 689, 17
Illingworth G., 1977, *ApJ*, 218, L43
Jahnke K., Macciò A. V., 2011, *ApJ*, 734, 92
Jin Y., Zhu L., Long R. J., Mao S., Xu D., Li H., van de Ven G., 2019, *MNRAS*, 486, 4753
Jing Y. P., Suto Y., 2002, *ApJ*, 574, 538
Johansson P. H., Naab T., Burkert A., 2009, *ApJ*, 690, 802
Karl S. J., Aarseth S. J., Naab T., Haehnelt M. G., Spurzem R., 2015, *MNRAS*, 452, 2337
Kormendy J., Bender R., 1996, *ApJ*, 464, L119
Kroupa P., 2001, *MNRAS*, 322, 231
Le Digabel S., 2011, *ACM Trans. Math. Softw.*, 37, 1
Magorrian J., 2006, *MNRAS*, 373, 425
Mehrgan K., Thomas J., Saglia R., Mazzalay X., Erwin P., Bender R., Kluge M., Fabricius M., 2019, *ApJ*, 887, 195
Merritt D., 2006, *ApJ*, 648, 976
Mikkola S., Merritt D., 2006, *MNRAS*, 372, 219
Mikkola S., Merritt D., 2008, *AJ*, 135, 2398
Mikkola S., Valtonen M. J., 1992, *MNRAS*, 259, 115
Miller R. H., Smith B. F., 1979, *ApJ*, 227, 407
Milosavljević M., Merritt D., 2001, *ApJ*, 563, 34
Monnet G., Bacon R., Emsellem E., 1992, *A&A*, 253, 366
Moster B. P., Naab T., White S. D. M., 2019, preprint (arXiv:1910.09552)
Naab T., Burkert A., 2003, *ApJ*, 597, 893
Naab T., Ostriker J. P., 2017, *ARA&A*, 55, 59
Naab T., Jesseit R., Burkert A., 2006, *MNRAS*, 372, 839
Navarro J. F., Frenk C. S., White S. D. M., 1996, *Astrophys. J.*, 462, 563
Novak G. S., Cox T. J., Primack J. R., Jonsson P., Dekel A., 2006, *ApJ*, 646, L9
Oser L., Ostriker J. P., Naab T., Johansson P. H., Burkert A., 2010, *ApJ*, 725, 2312
Parikh T. et al., 2018, *MNRAS*, 477, 3954
Peng C. Y., 2007, *ApJ*, 671, 1098
Posacki S., Cappellari M., Treu T., Pellegrini S., Ciotti L., 2015, *MNRAS*, 446, 493
Press W., Vetterling W., Metcalf M., Flannery B., Teukolsky S., 1996, Cambridge University Press, Cambridge
Rantala A., Pihajoki P., Johansson P. H., Naab T., Lahén N., Sawala T., 2017, *ApJ*, 840, 53
Rantala A., Johansson P. H., Naab T., Thomas J., Frigo M., 2018, *ApJ*, 864, 113
Rantala A., Johansson P. H., Naab T., Thomas J., Frigo M., 2019, *ApJ*, 872, L17
Richstone D. O., 1982, *ApJ*, 252, 496
Richstone D. O., Tremaine S., 1984, *ApJ*, 286, 27
Richstone D. O., Tremaine S., 1985, *ApJ*, 296, 370
Richstone D. O., Tremaine S., 1988, *ApJ*, 327, 82
Rix H.-W., de Zeeuw P. T., Cretton N., van der Marel R. P., Carollo C. M., 1997, *ApJ*, 488, 702
Röttgers B., Naab T., Oser L., 2014, *MNRAS*, 445, 1065
Schechter P. L., Gunn J. E., 1978, *AJ*, 83, 1360
Schwarzschild M., 1979, *ApJ*, 232, 236
Schwarzschild M., 1993, *ApJ*, 409, 563
Sérsic J. L., 1963, *Boletín de la Asociación Argentina de Astronomía La Plata Argentina*, 6, 99
Siopis C. et al., 2009, *ApJ*, 693, 946
Smith R. J., Lucey J. R., Conroy C., 2015, *MNRAS*, 449, 3441
Somerville R. S., Davé R., 2015, *ARA&A*, 53, 51
Springel V., 2005, *MNRAS*, 364, 1105
Syer D., Tremaine S., 1996, *MNRAS*, 282, 223
Tacconi L. J. et al., 2005, *Messenger*, 122, 28
Thomas J. et al., 2009, *MNRAS*, 393, 641
Thomas J. et al., 2011, *MNRAS*, 415, 545
Thomas J., Saglia R. P., Bender R., Thomas D., Gebhardt K., Magorrian J., Richstone D., 2004, *MNRAS*, 353, 391
Thomas J., Saglia R. P., Bender R., Thomas D., Gebhardt K., Magorrian J., Corsini E. M., Wegner G., 2005, *MNRAS*, 360, 1355

Thomas J., Jesseit R., Naab T., Saglia R. P., Burkert A., Bender R., 2007, *MNRAS*, 381, 1672
Thomas J., Ma C.-P., McConnell N. J., Greene J. E., Blakeslee J. P., Janish R., 2016, *Nature*, 532, 340
Toomre A., Toomre J., 1972, *ApJ*, 178, 623
Tremaine S., Hénon M., Lynden-Bell D., 1986, *MNRAS*, 219, 285
Treu T., Auger M. W., Koopmans L. V. E., Gavazzi R., Marshall P. J., Bolton A. S., 2010, *ApJ*, 709, 1195
Tsallis C., 1988, *J. Stat. Phys.*, 52, 479
Valluri M., Merritt D., Emsellem E., 2004, *ApJ*, 602, 66
van den Bosch R. C. E., de Zeeuw P. T., 2010, *MNRAS*, 401, 1770
van den Bosch R. C. E., van de Ven G., Verolme E. K., Cappellari M., de Zeeuw P. T., 2008, *MNRAS*, 385, 647
van der Marel R. P., Franx M., 1993, *ApJ*, 407, 525
van der Marel R. P., Cretton N., de Zeeuw P. T., Rix H.-W., 1998, *ApJ*, 493, 613
van Dokkum P. G., Conroy C., 2010, *Nature*, 468, 940
Vasiliev E., Valluri M., 2019, A New Implementation of the Schwarzschild Method for Constructing Observationally Driven Dynamical Models of Galaxies of All Morphological Types, Vol. 889, preprint (arXiv:1912.04288)
Vazdekis A. et al., 2015, *MNRAS*, 449, 1177
Vincent R. A., Ryden B. S., 2005, *ApJ*, 623, 137
Virtanen P. et al., 2019, SciPy 1.0: fundamental algorithms for scientific computing in Python, 17, Springer Science and Business Media LLC, preprint (arXiv:1907.10121)
Vogelsberger M. et al., 2014, *Nature*, 509, 177
White S. D. M., Narayan R., 1987, *MNRAS*, 229, 103
Will C. M., 2006, *Living Rev. Rel.*, 9, 3
Williams T. B., Schwarzschild M., 1979, *ApJS*, 41, 209

APPENDIX A: SURFACE BRIGHTNESS, ELLIPTICITY, AND POSITION ANGLE PROFILE

When fitting elliptical isophotes (Bender & Moellenhoff 1987) to the surface brightness maps for the principal axis directions of the

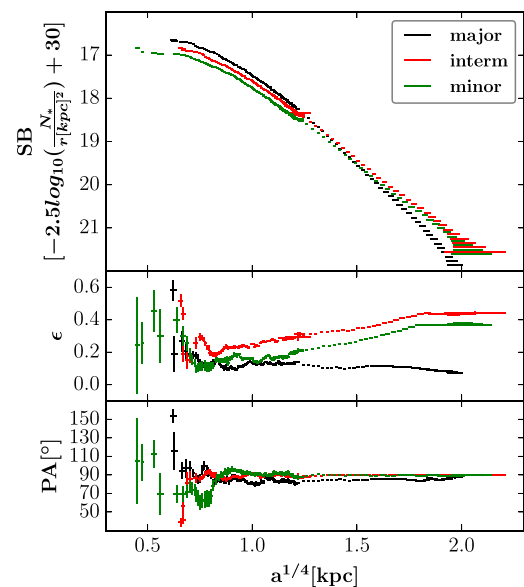


Figure A1. Surface brightness (first row), ellipticity (second row), and position angle profile (third row) for the major- (black), intermediate- (red), and minor-axis (green) projections of the simulation plotted against the semimajor axis a of the elliptical isophotes. The simulation demonstrates to be generic triaxial.

simulation in units of stellar simulation particles one can see that all three projections differ in their surface brightness (SB) and ellipticity profile ($\epsilon = 1 - b/a$ with a being the semimajor and b being the semiminor axis of the elliptical isophotes). The position angle (PA) profile shows some isophotal twists. Overall, the simulation appears to show typical triaxial behaviour (see Fig. A1).

APPENDIX B: VELOCITY, SURFACE BRIGHTNESS, AND χ^2 MAPS

Figs B1 and B2 show the velocity maps of the kinematic input data and their fits by SMART for four different projection axes and an additional psf convolved version of the intermediate axis. The plots

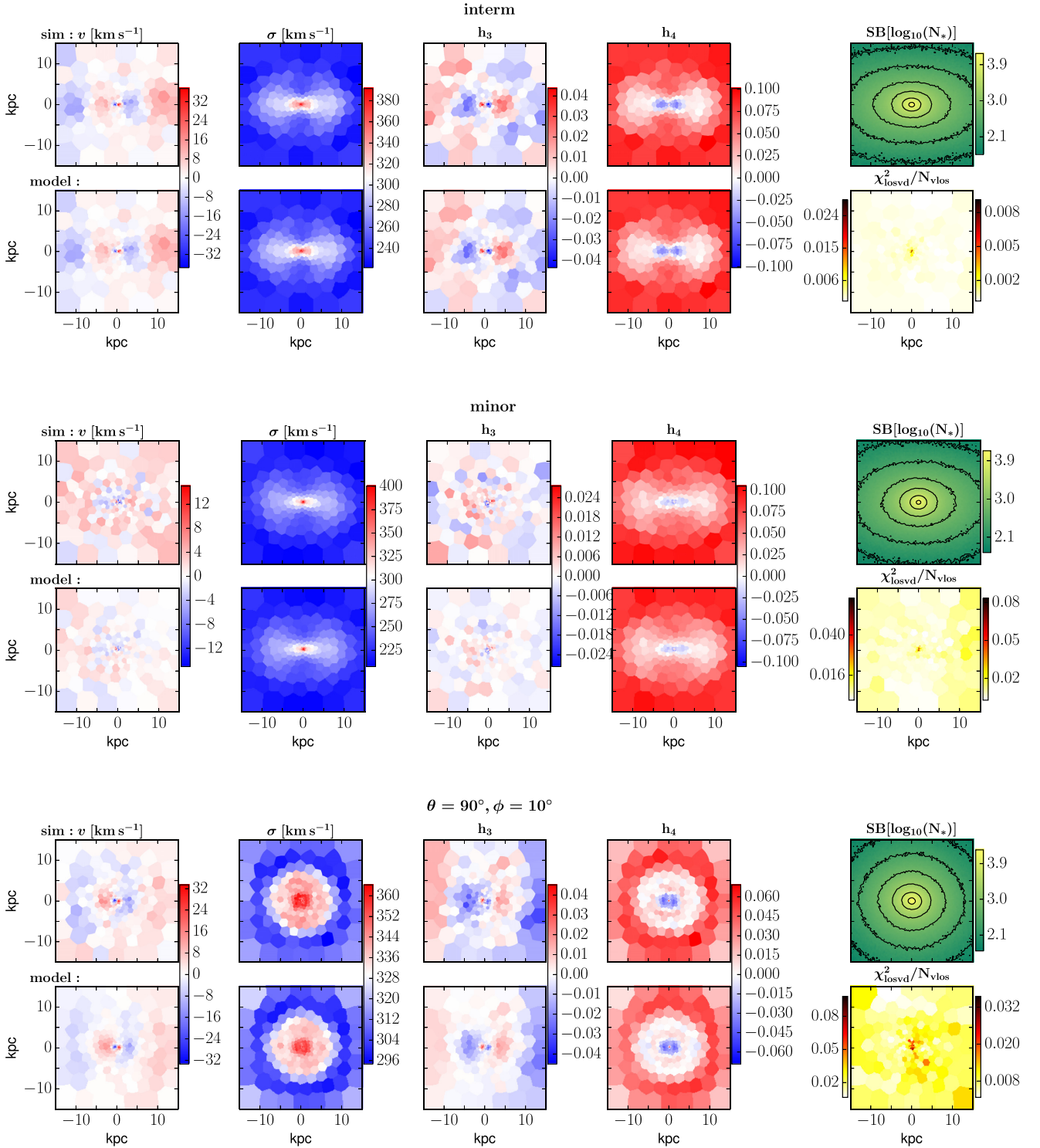


Figure B1. Velocity and surface brightness maps of the simulation (top row) as well as velocity maps of the model and χ^2 map as deviation from the kinematic input data with the modelled fit (bottom row) for different projections. The individual projection axis can be read from the title (for the major axis projection and a more detailed caption description see Fig. 3).

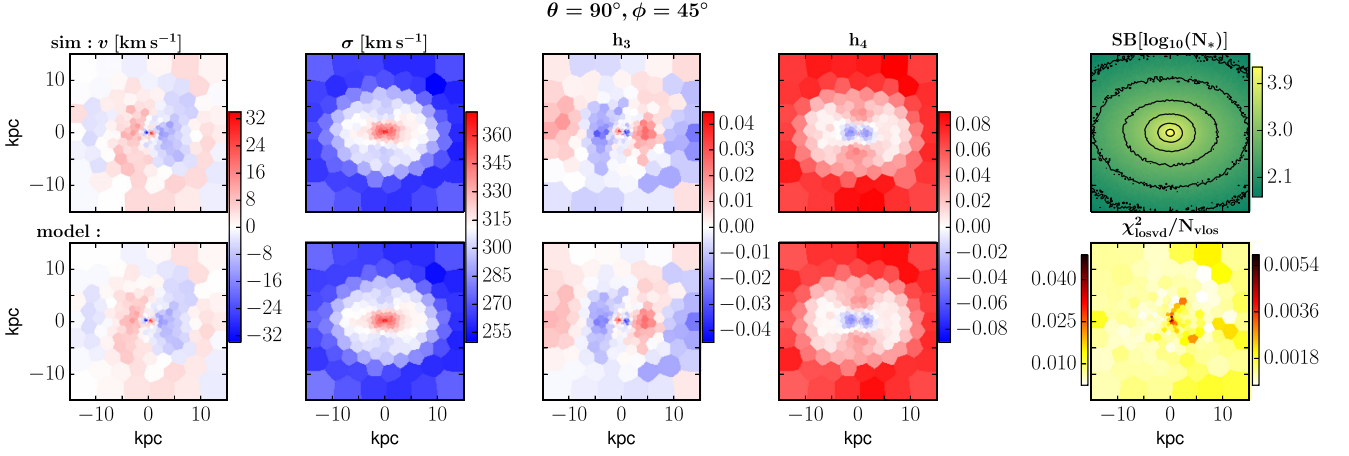


Figure B1 – continued

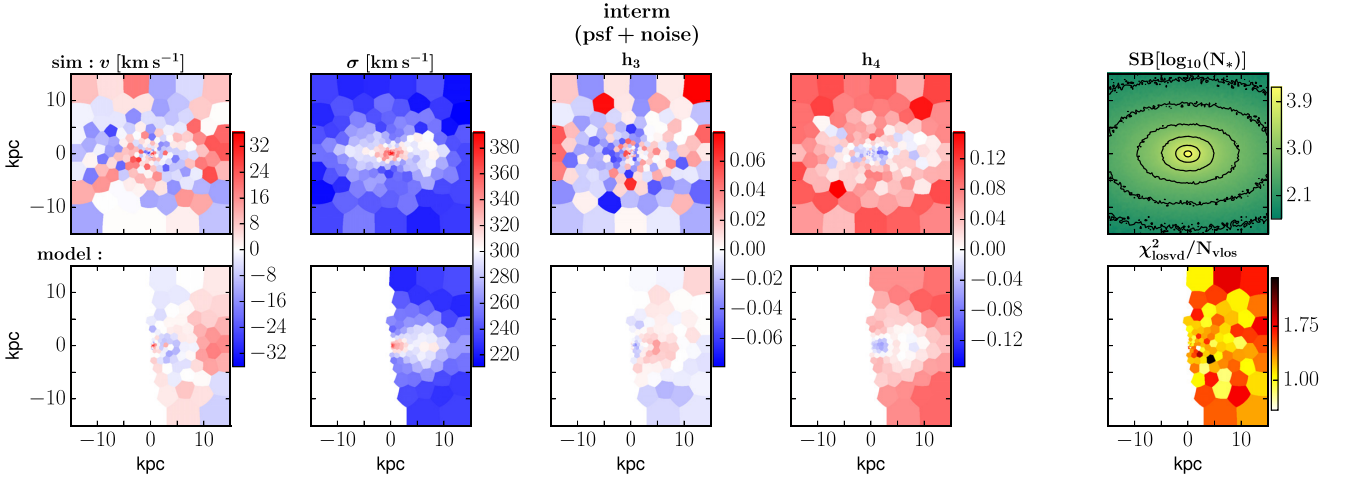


Figure B2. Velocity and surface brightness maps of the simulation (top row) as well as velocity maps of the model and χ^2 -map as deviation from the kinematic input data with the modelled fit (bottom row) for the noisy and psf convolved intermediate axis projection (for a more detailed description, see caption in Fig. 3 and Section 4.5.1).

also contain surface brightness maps in form of stellar simulation particles and χ^2 -maps as deviation between the simulation data and modelled fit. As already described in Section 4.1 one can see that SMART manages to fit the kinematic input data independent of the chosen projection.

APPENDIX C: ORBITAL REPRESENTATION OF THE PHASE SPACE

For generating our orbital initial conditions we set up energy shells by calculating the maximum allowed potential energy for N_E radii between the minimum allowed pericenter chosen to be $r_{\text{peri,min}} = r_{\text{min}}$ and the maximum allowed apocenter of $r_{\text{apo,max}} = r_{\text{max}}$. The radial bins are sampled by using formula 4. Each energy shell $i_E \in [1, N_E]$ is then subdivided in $i_E(L_z)$ sequences. We empirically find $i_E(L_z) = i_E^2$ as appropriate choice for the number of sequences per energy shell. The orbits for a specific sequence share the same initial z -component of the angular momentum L_{z,i_E} . These L_{z,i_E} sequences are logarithmically sampled between the minimum and maximum

allowed L_z -value in dependence of the radius $r(i_E)$ of the specific bin.

The final orbital initial spatial conditions (r_i, θ_i, ϕ_i) are then given

- (i) by randomly sampling azimuthal angles $\phi_i \in [0, 2\pi]$,
- (ii) by setting the elevation angles $\theta_i = 0$ because every orbit crosses the equatorial plane, and
- (iii) by randomly sampling possible radial values r_i between the minimum and maximum allowed apsis, i.e. $r_i \in [r_{\text{peri}}(L_{z,i_E}), r_{\text{apo}}(L_{z,i_E})]$.

While sampling the energy as an integral of motion is straightforward, the sampling of L_z is an arbitrary way to qualify the azimuthal velocity v_ϕ as the parameter we actually want to sample. In analogy, one could also sample the elevation velocity v_θ in terms of the x -component of the angular momentum L_x , which is discussed in more detail in Section 4.5.4. The computation of $r_{\text{peri}}(L_{z,i_E})$ and $r_{\text{apo}}(L_{z,i_E})$ is thereby done under the wrong assumption that L_{z,i_E} was conserved. Strictly speaking, this is only true for an axisymmetric potential as limiting case. However, our tests above show that this does not bias the results obtained with our orbit library. The remaining

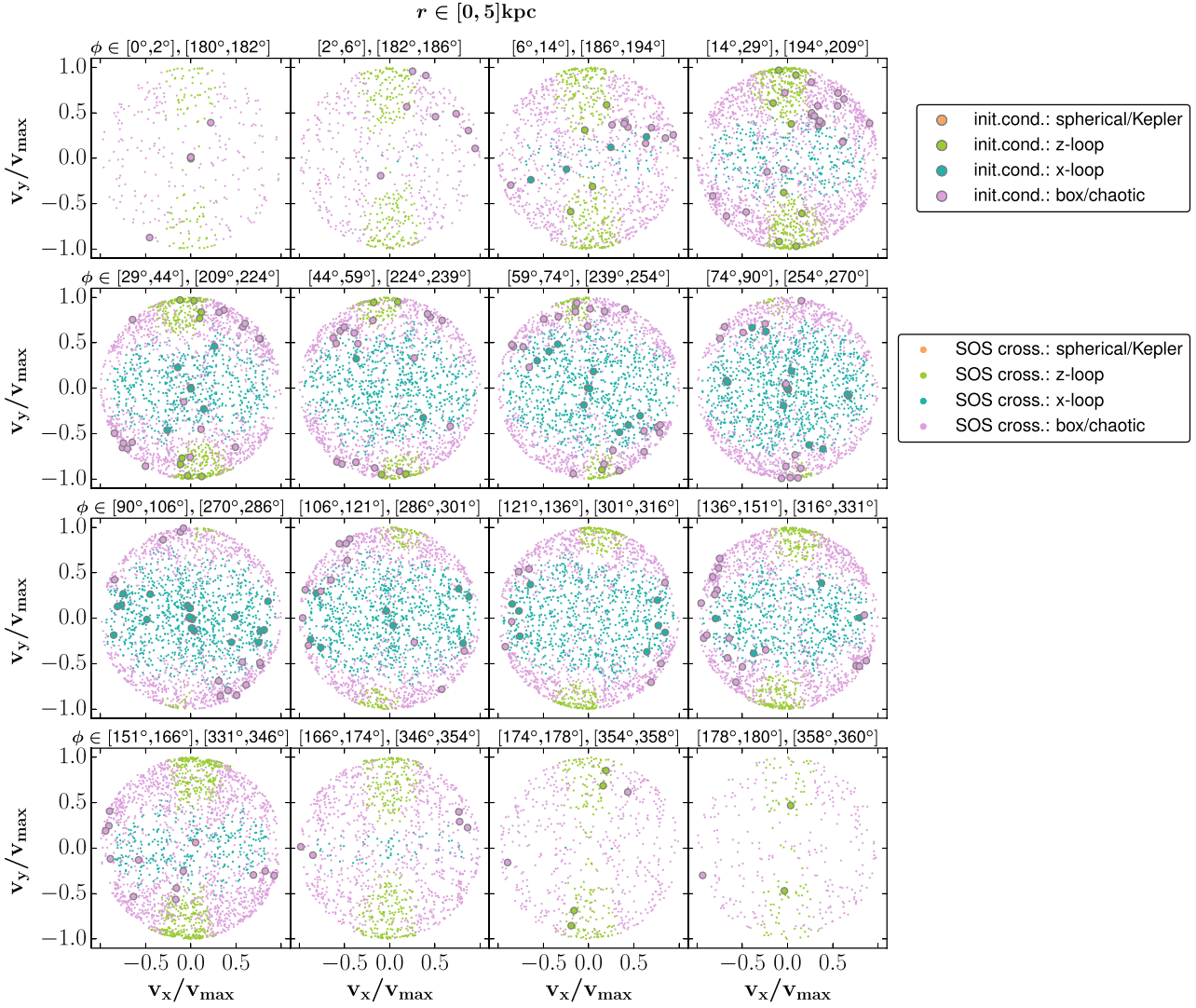


Figure C1. Orbital Phase-space coverage in the $x - y$ plane. We here plot the normalized velocity components v_x, v_y of all orbits in a single energy shell. This includes the orbital starting points (thick dots) as well as SOS-crossings during the time evolution of the orbits (thin dots). To consider all five dimensions of our starting space we distinguish between several radial and azimuthal annulus sectors of the plane (the chosen radial and azimuthal intervals are labelled above the individual panels). The orbit library produces a representative coverage of the phase space because the orbital imprints homogeneously fill the area enclosed by the unit circle.

initial velocity conditions $(v_{r,i}, v_{\theta,i}, v_{\phi,i})$ can then be subsequently computed as

$$\begin{aligned}
 & \text{(i) } v_{\phi,i} = \frac{L_{z,i} E}{r_i}, \\
 & \text{(ii) } v_{r,i} \in [0, v_{r,\max}] \quad \text{with} \quad v_{r,\max} = \\
 & \sqrt{2(E - \Phi(r_i, \theta_i = 0, \phi_i)) - v_{\phi,i}^2}, \text{ where } E \text{ is the total energy and } \Phi \text{ the potential and} \\
 & \text{(iii) } v_{\theta,i} = \sqrt{2(E - \Phi(r_i, \theta_i = 0, \phi_i)) - v_{\phi,i}^2 - v_{r,i}^2}.
 \end{aligned}$$

More initial conditions are generated by filling the surfaces of section (SOS; Henon & Heiles 1964; Richstone 1982). The surface of section, also called Poincaré section, consists of the r_i - and $v_{r,i}$ -values of the orbits crossing the equatorial plane, i.e. $\theta_i = 0$, in upwards direction. The topology of the SOS depends on the integrals of motion. The orbit library representatively fills the phase-space if the SOS is properly filled. For this, we randomly generate azimuthal

angles ϕ_i which get divided in N_{sector} azimuthal sectors. For each azimuthal sector, we then sample $(r_i, v_{r,i})$ -tuples that are chosen so that their distance to the nearest imprint-tuple is maximized. The remaining initial conditions $v_{\phi,i}$ and $v_{\theta,i}$ are generated as before. To create a homogeneous distribution of pro- and retrograde orbits, the orbits contained in the library which are later on classified as regular orbits (see Section 2.3.1) get duplicated by changing the sign of their direction of rotation.

For an illustration of our 5D starting space of the stellar orbits in Fig. C1, we show a series of 2D plots with velocity spheres on annulus sectors (chosen radial and azimuthal sections) in the equatorial plane (i.e. $\theta = 0$). Since the magnitude of the velocity vector $\sqrt{v_x^2 + v_y^2 + v_z^2}$ is determined by the energy E , it is sufficient to plot v_x, v_y -imprints for a certain energy shell (the remaining v_z -component is then given as velocity shell in dependence of the value of E). We show both, the velocity components v_x, v_y of the

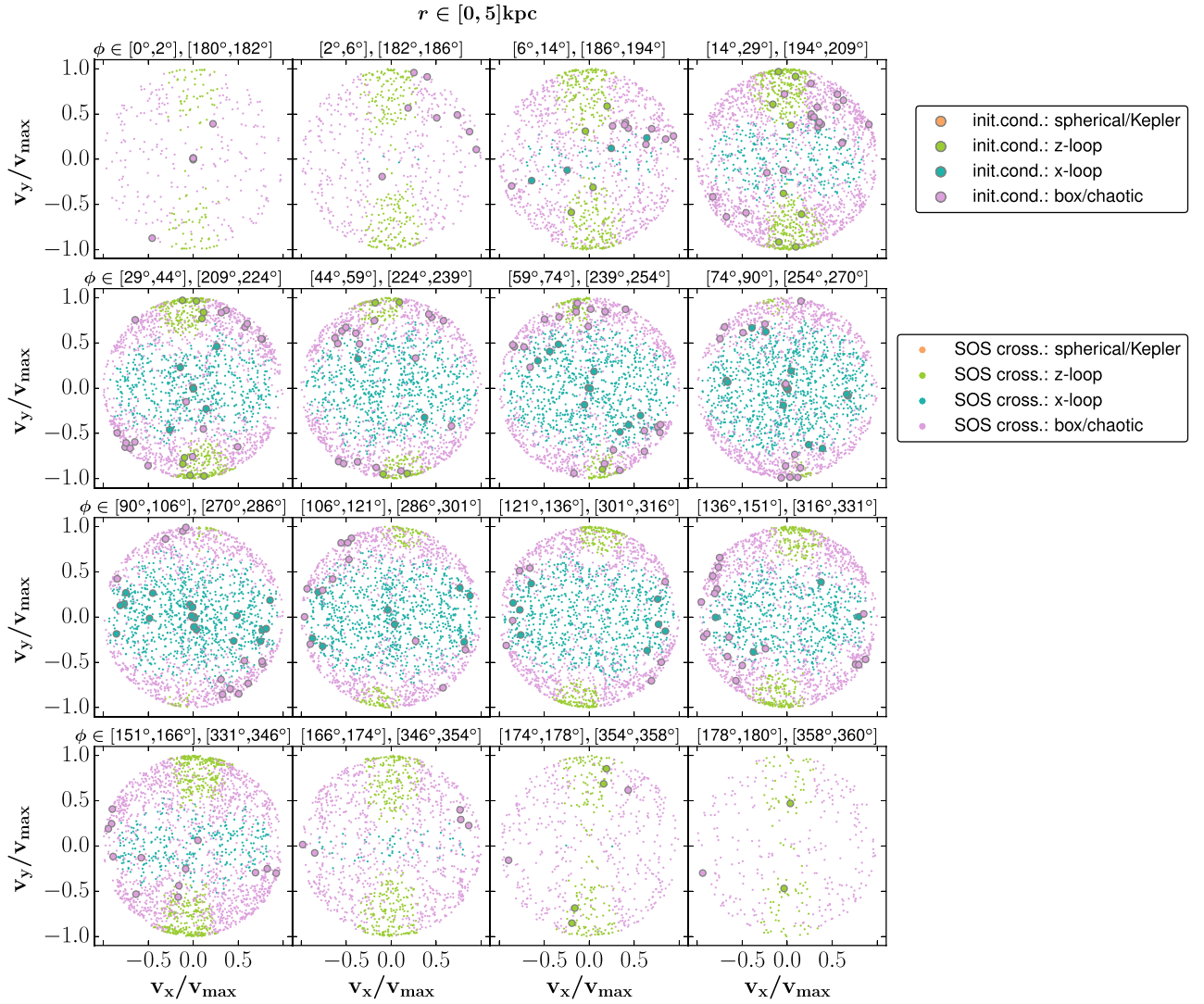
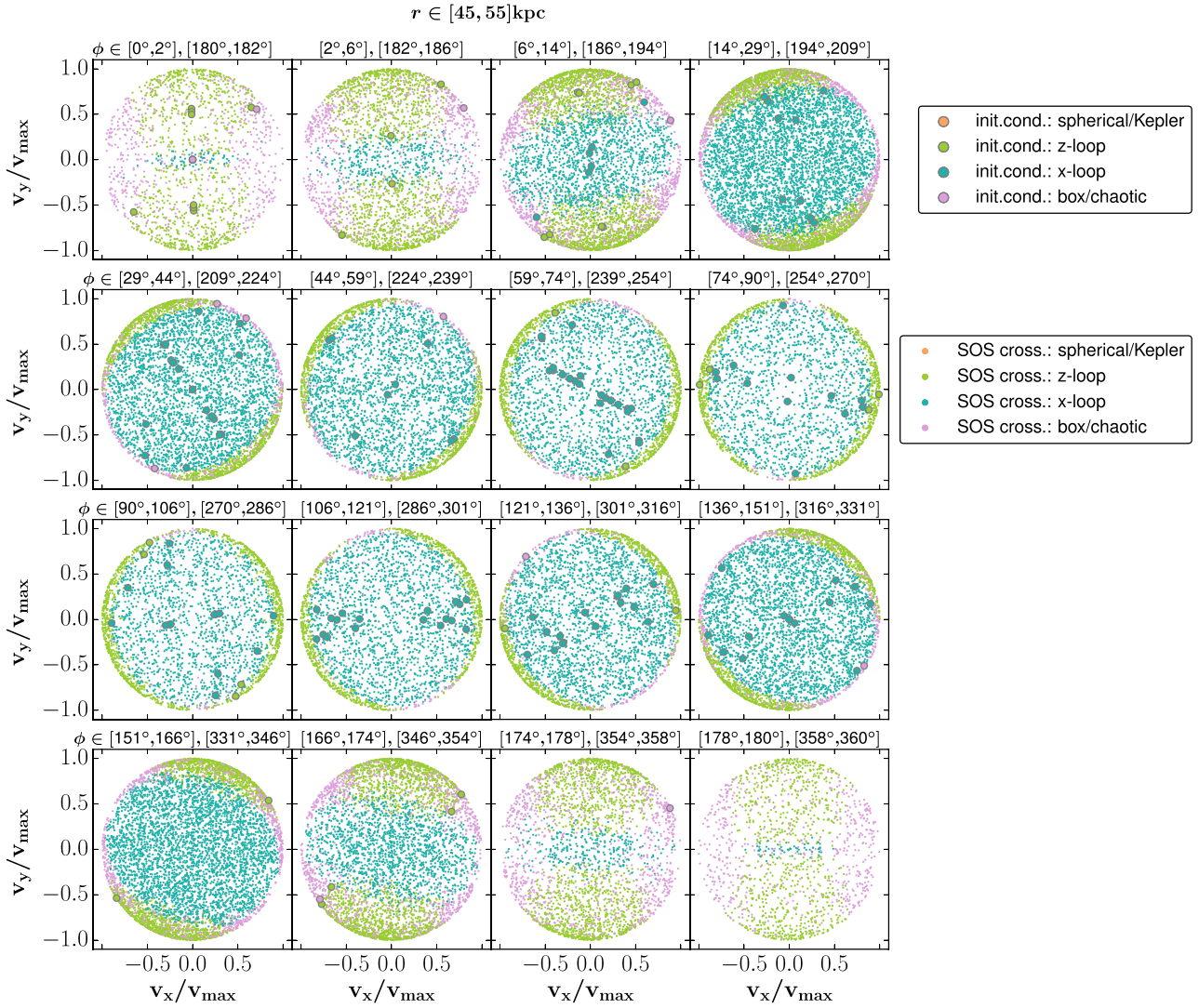


Figure C1 – continued

orbital initial starting point (marked as thick dots) as well as the orbital imprints for every crossing event in the subsequent time evolution (thin dots). The velocity components are normalized by $v_{\max}(r, \theta, \phi) = \sqrt{2(E - \Phi(r, \theta, \phi))}$, with Φ being the gravitational potential. The different colours correspond to the different orbit types. Fig. C1 shows that the orbits of our orbit library homoge-

neously represent the accessible phase space. A sufficient coverage of the phase space is crucial for a reliable model. In addition, one can see that the actual starting points of the orbits can be chosen in very different ways. Our random choice is only one possibility that leads to an appropriate phase-space representation. Most of the space gets automatically filled by the orbital imprints over time.



APPENDIX D: OPTIMIZATION ALGORITHM AND TEST

To find the unique set of orbital weights that maximizes \hat{S} our code conceptionally seeks for a solution

$$\nabla_{\mathbf{w}} \hat{S} = 0, \quad (\text{D1})$$

subject to the constraints (19).

Technically, we follow Richstone & Tremaine (1988) and treat the orbital weights \mathbf{w} and the kinematical model predictions

$$\vec{\mathbf{y}} \equiv \vec{\mathcal{L}}_{\text{data}} - \vec{\mathcal{L}}_{\text{mod}} \quad (\text{D2})$$

as independent variables. The actual dependence $\vec{\mathbf{y}} = \vec{\mathbf{y}}(\vec{\mathbf{w}})$ (equation 16) is added to the constraints.

Thus, let

$$\vec{\mathbf{x}} \equiv \begin{pmatrix} \vec{\mathbf{w}} \\ \vec{\mathbf{y}} \end{pmatrix} \quad (\text{D3})$$

be the $N_{\text{orbit}} + N_{\text{data}}$ new variables, then the model solves

$$\nabla_{\vec{\mathbf{x}}} \hat{S} - \mathbf{C}^T \cdot \boldsymbol{\lambda} = 0 \quad (\text{D4})$$

for the x_i and for $N_{\text{data}} + N_{\text{phot}}$ Lagrange multipliers λ_j (combined in the vector $\vec{\boldsymbol{\lambda}}$). The matrix

$$\mathbf{C} \equiv \begin{pmatrix} \mathbf{P}_{\text{orb}} & 0 \\ \mathcal{L}_{\text{orb}} & 1 \end{pmatrix} \quad (\text{D5})$$

contains the constraint equations

$$\begin{pmatrix} \vec{\mathbf{p}}_{\text{data}} \\ \vec{\mathcal{L}}_{\text{data}} \end{pmatrix} = \mathbf{C} \cdot \vec{\mathbf{x}}. \quad (\text{D6})$$

Equation (D4) is non-linear and solved with a Newton method:

$$\vec{\mathbf{x}}^{(n+1)} = \vec{\mathbf{x}}^{(n)} - \left(\frac{d\mathbf{g}}{d\mathbf{x}} \right)^{-1} \vec{\mathbf{g}}, \quad (\text{D7})$$

or,

$$\vec{\Delta} \mathbf{x}^{(n)} = - \left(\frac{d\mathbf{g}}{d\mathbf{x}} \right)^{-1} \vec{\mathbf{g}}, \quad (\text{D8})$$

where

$$\mathbf{g} \equiv \nabla_{\vec{\mathbf{x}}} \hat{S} - \mathbf{C}^T \cdot \boldsymbol{\lambda}. \quad (\text{D9})$$

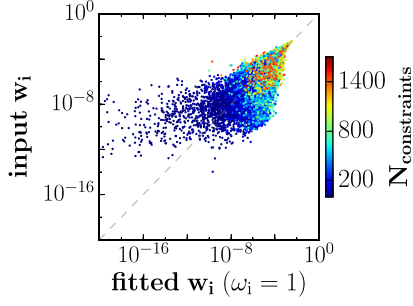


Figure D1. Recovery of the orbital weights of the RANDOM test model described in Section D. Each dot represents one orbit with the colour indicating the number of LOSVD data points $N_{\text{constraints}}$ that the orbit contributes to. For the recovery of the weights we fitted the kinematics of the RANDOM model using our fiducial constant bias factors $\omega_i = \text{const.}$ (Shannon entropy). Due to the degeneracy in the χ^2 minimization, the recovered orbital weights differ from the original weights. Larger differences are observed for orbits which are less constrained by data (blue points).

At each iteration, the Lagrange multipliers are updated via (D6),

$$\begin{pmatrix} \Delta \vec{p}_{\text{data}} \\ \Delta \vec{\mathcal{L}}_{\text{data}} \end{pmatrix} = \mathbf{C} \cdot \Delta \vec{x}, \quad (\text{D10})$$

which leads to

$$\mathbf{A} \cdot \vec{\lambda} = \begin{pmatrix} \Delta \vec{p}_{\text{data}} \\ \Delta \vec{\mathcal{L}}_{\text{data}} \end{pmatrix} + \mathbf{C} \cdot \left(\frac{d\mathbf{g}}{d\mathbf{x}} \right)^{-1} \cdot \nabla_x \hat{S}, \quad (\text{D11})$$

$$\mathbf{A} \equiv \mathbf{C} \left(\frac{d\mathbf{g}}{d\mathbf{x}} \right)^{-1} \mathbf{C}^T. \quad (\text{D12})$$

The Jacobian $\frac{d\mathbf{g}}{d\mathbf{x}}$ is diagonal and its inverse is easy to compute. The only matrix that needs to be inverted numerically is \mathbf{A} and its dimension is $N_{\text{data}} + N_{\text{phot}} < N_{\text{orbit}}$.

Modulo the scaling by $\left(\frac{d\mathbf{g}}{d\mathbf{x}}\right)^{-1}$, \mathbf{A} reads

$$\mathbf{A} = \begin{pmatrix} \mathbf{P}_{\text{orb}} \mathbf{P}_{\text{orb}}^T & \mathbf{P}_{\text{orb}} \mathcal{L}_{\text{orb}}^T \\ \mathcal{L}_{\text{orb}} \mathbf{P}_{\text{orb}}^T & \mathcal{L}_{\text{orb}} \mathcal{L}_{\text{orb}}^T + \mathbf{1} \end{pmatrix}. \quad (\text{D13})$$

The required linear algebra operations are implemented through the LAPACK library (Anderson et al. 1999).

In practice, to find the best-fitting orbit model, we start with a very small value α_0 and then iteratively increase α until χ^2 does not change anymore. Everytime we change $\alpha \rightarrow \alpha + d\alpha$ the unique solution to equation (7) will shift slightly from \vec{w} to $\vec{w} + d\vec{w}$ in the space of the orbital weights. Because $f \equiv -\hat{S}$ is differentiable with respect to α we can keep the new solution $\vec{w} + d\vec{w}$ in an arbitrarily small neighbourhood around \vec{w} by limiting $d\alpha$ accordingly. For strictly convex f , the Newton method converges locally (and the convergence is quadratic). This means that once we have found a solution for some value of α , the convergence of the Newton steps for neighbouring $\alpha + \Delta\alpha$ is guaranteed (and fast) if we use the solution at α as starting point for the search at $\alpha + \Delta\alpha$: it is only a question of limiting $\Delta\alpha$, if necessary. In this way, we can iteratively solve for any value of α starting from α_0 . In practice, we increase α by a factor of 2 in each iteration. Intermediate steps with a smaller α increment are inserted in case an iteration has not converged (which almost never happens).

At the initial α_0 we start from a homogeneous distribution of orbital weights. Methods with guaranteed convergence even from starting points outside of the constraint conditions have been developed. We do not use such methods, because in practice we always find the solution of equation (D4) even from an initially homogeneous

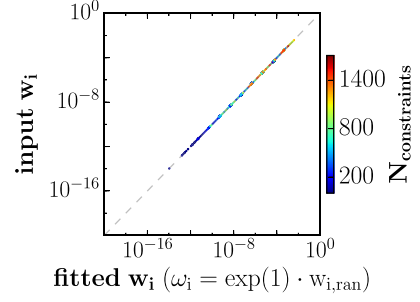


Figure D2. Same as Fig. D1, but here we fitted the RANDOM model using $\omega_i = \exp(1) \cdot w_{i,\text{ran}}$. With these bias factors the unique maximum of equation (7) should occur at the original weights $w_{i,\text{ran}}$ of the RANDOM model. And indeed, even though we start the fit with an initial guess where all the weights are equal, the true solution is reproduced with very high precision. This demonstrates the robustness and high accuracy of our optimization algorithm. The two different fits presented here and in Fig. D1 illustrate how the maximum entropy technique with variable ω_i can be used to sample different solutions that minimize the χ^2 .

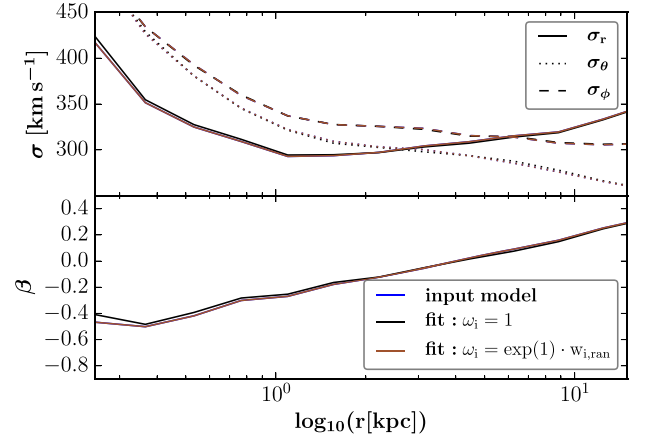


Figure D3. Internal velocity structure of the two fits presented in Fig. D1 (black lines) and D2 (red lines). Despite the different orbital weight distributions, the anisotropy in the second-order velocity moments is almost equally well recovered in both cases. This provides evidence that when using the full information contained in the entire LOSVDs, the remaining degeneracies in the recovery of the distribution function of triaxial galaxies (i.e. the differences in the detailed population of the various orbits) have only little impact on the ‘macroscopic’ internal galaxy properties like the anisotropy β . This, in turn, is crucial for a precise mass reconstruction.

distribution of orbital weights. Since the solution is global and unique for each α , convergence at any step of the α sequence can be tested based on equation (D4) itself.

To test our implementation, we reconstruct the kinematical data of a toy model with randomized orbital weights. The toy model is constructed as a maximum-entropy model [i.e. through maximization of \hat{S} in equation (7) with $\alpha = 0$]. The bias factors ω_i for the orbital weights are set to $\omega_i = 10^{r_i}$, where $r_i \in [0, 3]$ are random numbers. After maximization of \hat{S} subject to the density constraints, the resulting orbital weights w_i satisfy the density constraints of the N -body simulation. Due to these density constraints, the w_i are not directly proportional to the random ω_i , yet they are still very strongly affected by them and thus we call this toy model with randomised orbital weights RANDOM in the following. We adapt

$r_{\text{peri},\text{min}} = r_{\text{min}} = 0.05$ kpc and $r_{\text{apo},\text{max}} = r_{\text{max}} = 21$ kpc. If an orbit conserves the sign of at least one angular-momentum component, then this orbit has a duplicate companion orbit with opposite direction of rotation in the library (this concerns tube orbits and spherical orbits; see also Section C). Let i^+ and i^- denote the orbital indices of such a pair of orbits, where i^+ refers to the orbit with initial velocity components $v_r, v_\theta, v_\phi > 0$. To increase the degree of difficulty for reconstructing the RANDOM model, we modify the weights of orbits in such a pair:

$$w_{i^+,\text{ran}} = (1 - c_w) \cdot (w_{i^+} + w_{i^-}),$$

$$w_{i^-,\text{ran}} = c_w \cdot (w_{i^+} + w_{i^-}). \quad (\text{D14})$$

By using $c_w = 10^{-3}$, equation (D14) leads to increased weights $w_{i^+,\text{ran}}$ and almost eliminated weights $w_{i^-,\text{ran}}$, corresponding to an internal rotation of the model (see also Thomas et al. 2007). We then modelled the LOSVDs of the major axis projection of this RANDOM toy model with SMART in order to test our solver algorithm for the orbital weights.

Because the χ^2 minimization is non-unique (cf. Section 5), the solution of equation (7) will in general not occur at the true weights $w_{i,\text{ran}}$ of the RANDOM model. We therefore tried two different configurations: (i) we used the default orbital bias factors $\omega_i = 1$ (Shannon entropy) and (ii) we set $\omega_i = \exp(1) \cdot w_{i,\text{ran}}$. In the second case, we push the maximum entropy solution towards the original, randomly generated weights. Only in this case, we can predict the location of the global solution analytically and test our implementation rigorously. For ensuring fair conditions we set the initial weights to $w_{i,\text{ini}} = 1$ in both cases when we start the minimization. An efficient optimization algorithm should be able to find two equivalent degenerate solutions, one for case (i) and a different one for case (ii). The true orbital weights of the RANDOM model should be recovered (only) in the latter case.

SMART is indeed able to fit the input data for both cases very well. We stopped the minimization at $\bar{\chi}^2 = 2 \times 10^{-9}$ for $\omega_i = 1$ and $\bar{\chi}^2 = 5 \times 10^{-9}$ for $\omega_i = \exp(1) \cdot w_{i,\text{ran}}$. Fig. D1 shows the comparison between the true orbital weights of the RANDOM model and the recovered weights when using $\omega_i = 1$ (case (i) above). As expected, these weights *differ* from the RANDOM weights, since the

χ^2 minimization and, hence, the recovery of the entire phase-space distribution function, is non-unique. Orbits that contribute to a larger number of LOVSD data points and, in turn, are better constrained by the data have fitted weights that tend to be closer to the true ones. Fig. D2 shows the orbital weights of the input model and the modelled fit using $\omega_i = \exp(1) \cdot w_{i,\text{ran}}$. As explained above, in this case we expect the recovered solution to exactly match with the RANDOM model. Fig. D2 demonstrates the high quality of our solver algorithm since the fitted weights indeed match exactly with the RANDOM weights ($\Delta \bar{w} = 4 \times 10^{-9}$).

The two different fits presented in Figs D1 and D2 illustrate how the maximum entropy technique with variable ω_i can be used to sample different solutions that are equivalent with respect to the χ^2 minimization.

In Fig. D3, we compare the ‘macroscopic’ internal velocity structure of these two fits. Even though the detailed population of the orbits (i.e. the exact phase-space distribution function) of both models differ, they both reproduce the internal moments of the RANDOM model very well ($\text{rms}_\sigma = 5 \times 10^{-3}$ for $\omega_i = 1$ and $\text{rms}_\sigma = 3 \times 10^{-5}$ for $\omega_i = \exp(1) \cdot w_{i,\text{ran}}$). As further discussed in Section 5, this suggests that *when fitting the full LOSVDs* the remaining degeneracies in the recovery of the distribution function of a triaxial galaxy (with known normalized densities) show only little effect on the ‘macroscopic’ galaxy parameters of interest.

APPENDIX E: ORBIT FRACTION RECOVERY WHEN TESTING THE UNIQUENESS OF THE ANISOTROPY RECOVERY

We here show the remaining orbit fraction recovery plots for the analysis described in Section 5.2. The recovery of the z -loops was already shown in Fig. 14 of that section.

Here, we plot the recoveries of the fraction of orbits classified as x -loops (see Fig. E1), box/chaotic orbits (see Fig. E2) as well as spherical/Kepler orbits (see Fig. E3). Independent of the line of sight (different rows) or chosen entropy method (black lines), the orbit class fractions of the individual input toy-models (different rows) are well recovered.

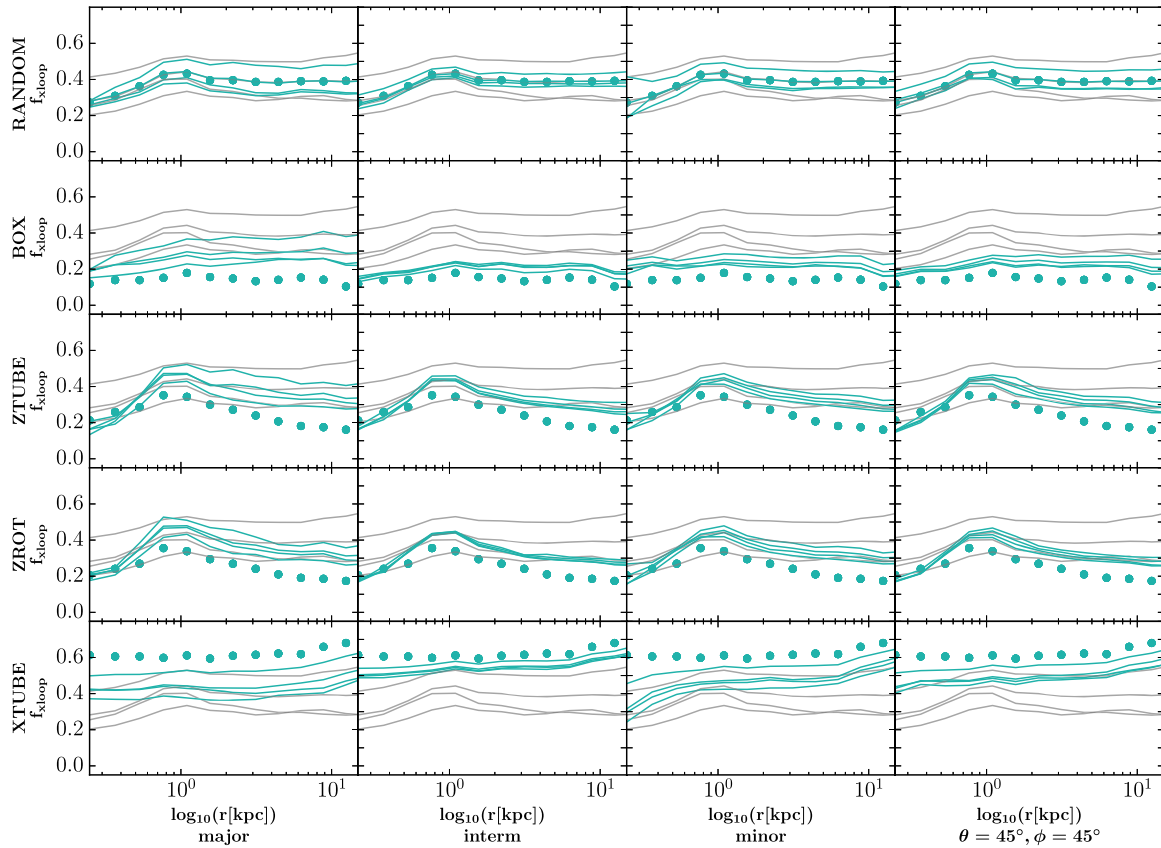


Figure E1. Recovery of the x -loop orbit fractions of different input toy models by models with different entropy methods. We here show the same analysis as in Figs 13 and 14 but now for the reconstruction of the fraction of orbits classified as x -tubes. Independent of the tested projection, the x -loop fractions of the individual input toy models are well recovered by the models using different entropy methods.

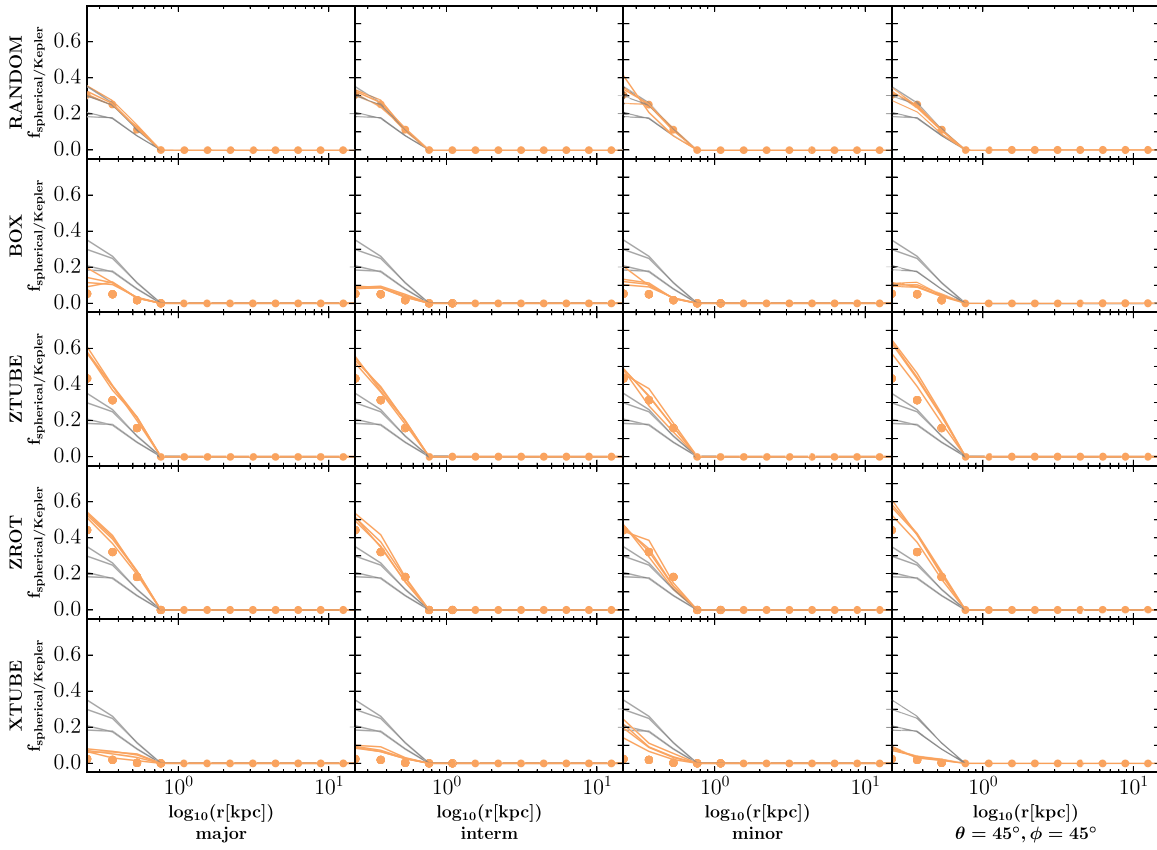


Figure E2. Recovery of the box/chaotic orbit fractions of different input toy models by models with different entropy methods. For a more detailed caption description see Figs 13 and E1.

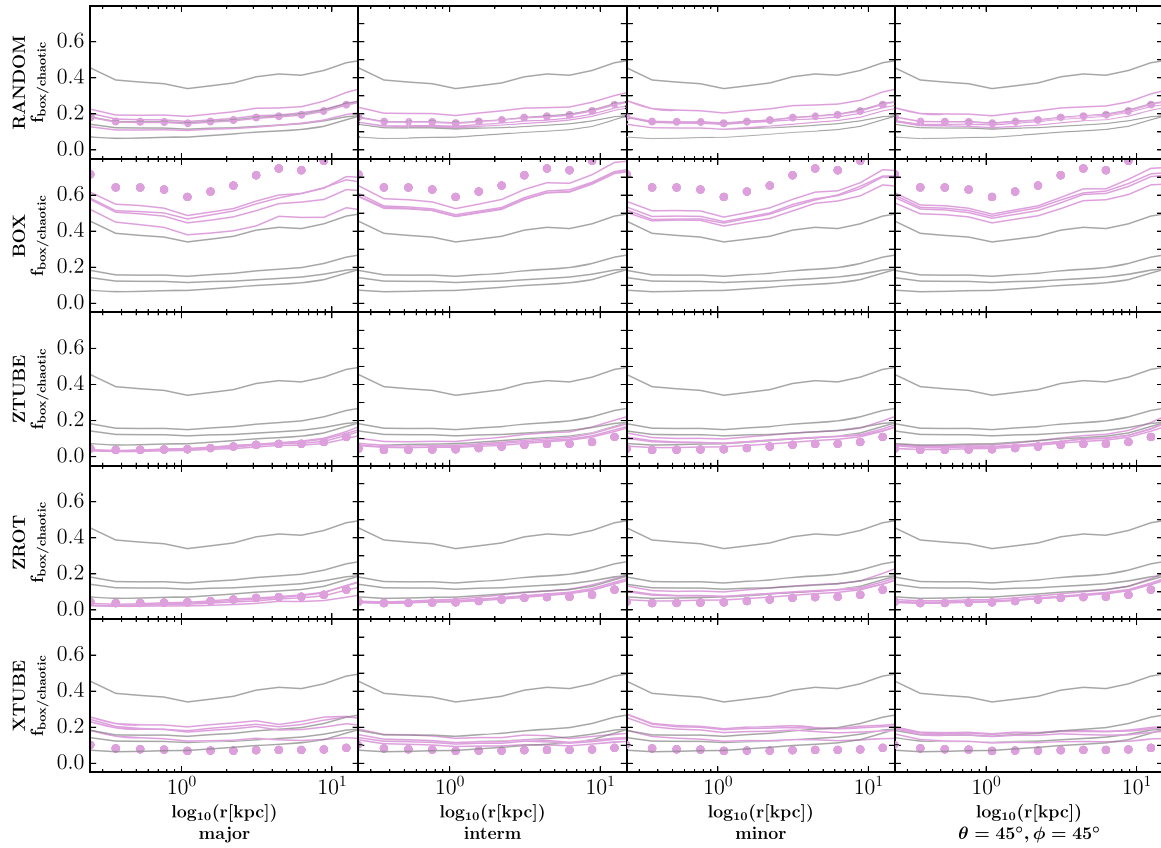


Figure E3. Recovery of the spherical/Kepler orbit fractions of different input toy models by models with different entropy methods. For a more detailed caption description see Figs 13 and E1.

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