

GravSphere2: A higher order Jeans method for mass modeling spherical stellar systems

Andrés Bañares-Hernández^{1,2,*}, Justin I. Read³, and Mariana P. Júlio^{4,5}

¹ Instituto de Astrofísica de Canarias, La Laguna, Tenerife 38200, Spain

² Departamento de Astrofísica, Universidad de La Laguna, Santa Cruz de Tenerife, Spain

³ Department of Physics, University of Surrey, Guildford, GU2 7XH, UK

⁴ Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany

⁵ Institut für Physik und Astronomie, Universität Potsdam, Karl-Liebknecht-Straße 24/25, 14476 Potsdam, Germany

Received 10 September 2025 / Accepted 10 November 2025

ABSTRACT

Aims. Mass-modeling methods are used to infer the gravitational field of stellar systems, from globular clusters to giant elliptical galaxies. While many methods already exist, most require assumptions on the form of the underlying distribution function or binning the data, leading to some loss of information. Furthermore, when only line-of-sight (LOS) data are available, many methods suffer from the well-known mass-anisotropy degeneracy. To overcome these limitations, we developed a new and publicly available mass modeling method, GRAVSPHERE2. It combines individual stellar velocities from LOS and proper motion (PM) measurements to solve the Jeans equations up to fourth order, without any data binning. Using flexible functional forms for the velocity anisotropy profiles at second and fourth order, we show how including additional constraints from a new observable, fourth-order PMs, allows us to obtain a full solution along the three dimensions and breaking the mass-anisotropy degeneracy at all orders. We tested our method on mock data for dwarf galaxies, showing how GRAVSPHERE2 improves on previous methods.

Methods. GRAVSPHERE2 introduces four key improvements over previous Jeans mass modeling methods in the literature: (i) we included fourth-order velocity moment equations in both the LOS and PM directions, for the first time, using them to break model degeneracies; (ii) we used a fully general treatment of both the second and fourth-order velocity anisotropies; (iii) we introduced a “bin-free” approach where we fit individual tracer velocities and positions using flexible and self-consistent probability density functions that include kurtosis; and (iv) we improved the likelihood sampling by using the nested sampler DYNESTY.

Results. GRAVSPHERE2 was able to recover the mass density, stellar velocity anisotropy, and the logarithmic slope of the mass density profile within its quoted 95% confidence intervals across almost all mocks over a wide radial range ($0.1 \lesssim r/R_{1/2} \lesssim 10$, where $R_{1/2}$ is the projected half-light radius). As the number of tracers is lowered (even down to just ten tracers) it gracefully degrades, with larger uncertainties but no induced bias. We find that GRAVSPHERE2 outperforms simple mass estimators, suggesting that it is worth using even when only a few LOS velocities are available. Using 1000 tracers without PMs, GRAVSPHERE2 recovers the logarithmic density slope at $R_{1/2}$ with 12% (25%) statistical errors for cuspy (cored) mock data, enabling us to make a distinction between the two. When including PMs, this result can be improved to 8% (12%). With only 100 tracers and no PMs, we were still able to recover slopes with ~30% (20%) errors. GRAVSPHERE2 will become a valuable new tool to hunt for massive black holes and invisible dark matter in spherical stellar systems, from globular clusters and dwarf galaxies to giant ellipticals and galaxy clusters.

Key words. globular clusters: general – galaxies: dwarf – galaxies: kinematics and dynamics – Local Group

1. Introduction

Many stellar systems are approximately spherical, including globular clusters (GCs) and nearby dwarf spheroidal galaxies (dSphs). Both GCs and dSphs contain seemingly invisible mass components that can be inferred from the anomalous motions of their stars and gas (see e.g., Bosma 1978; Rubin et al. 1980; Trimble 1987; Łokas 2002; Van der Marel & Anderson 2010; Battaglia et al. 2013, for more recent examples in GCs and dSphs; see e.g., Evans et al. 2022; Dickson et al. 2024; Smith et al. 2024a; Bañares-Hernández et al. 2025 and Read et al. 2019; Hayashi et al. 2020; Battaglia & Nipoti 2022; Bañares-Hernández et al. 2023; De Leo et al. 2024; Arroyo-Polonio et al. 2025; Pascale et al. 2025; Vitral et al. 2025). For GCs, the invisible mass most likely constitutes stellar remnants or an intermediate-mass black hole (IMBH), although some GCs may also contain dark matter. By contrast, dSphs appear to

be dominated at all radii by dark matter. With increasingly available spectroscopic and astrometric data from recent surveys, GCs promise an unprecedented natural laboratory to study stellar remnants and hunt for IMBHs (Mezcua 2017; Neumayer et al. 2020; Greene et al. 2020; Askar et al. 2023; Vitral et al. 2023; Häberle et al. 2024; Bañares-Hernández et al. 2025), while dSphs promise the same for dark matter (Boehm et al. 2017; Read et al. 2018, 2019; Battaglia & Nipoti 2022; De Leo et al. 2024; Vitral et al. 2024, 2025; Zimmermann et al. 2025).

The above proliferation in data has led to an associated proliferation in mass modeling methodologies. All methods seek to solve the steady-state Collisionless Boltzmann Equation (CBE), also known as the Vlasov Equation (e.g., Binney & Tremaine 2008), expressed as

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \nabla_x f \cdot \mathbf{v} - \nabla_v f \cdot \nabla_x \Phi = 0, \quad (1)$$

where Φ is the gravitational potential resulting from all mass components in the system. Then, satisfying the Poisson equation

* Corresponding authors: a.banareshernandez@gmail.com

with mass density, ρ , we have

$$\nabla_{\mathbf{x}}^2 \Phi = 4\pi G \rho, \quad (2)$$

and f is the phase-space distribution function.

By solving for the gravitational potential and subtracting the expected potential from all visible components we can infer the presence of invisible masses for future study, such as black holes and dark matter (e.g., [Van der Marel & Anderson 2010](#); [Walker et al. 2009](#)), which is usually done in the steady state approximation, assuming $\frac{\partial f}{\partial t} = 0$.

There are number of methods for solving the CBE. Jeans methods solve moments of the CBE ([Jeans 1922](#)), having the advantages that they are fast¹ and minimize potential biases that could creep in due to assumptions about the form of the distribution function, f (e.g., [Cappellari 2008](#); [Watkins et al. 2013](#); [Mamon et al. 2013](#); [Cappellari 2015](#); [Read & Steger 2017](#)). The main disadvantages, however, are that the data typically need to be binned, while f is not determined. This leads to some loss of information and makes it hard to incorporate stars of uncertain membership, binaries, and/or rigorously model selection effects (e.g. [Read et al. 2021](#)). Furthermore, since f is not determined, a physical f is not guaranteed. This can result in models that formally require a negative f in some regions of phase space ([Dejonghe & Merritt 1992](#); [Richardson & Fairbairn 2013](#)) which should then be discarded (e.g., [An & Evans 2006](#)).

Distribution function methods assume some form for f and then fit this to the data (e.g., [Gieles & Zocchi 2015](#); [Vasiliev 2019](#); [Read et al. 2021](#); [Dickson et al. 2023](#)). In such cases, f can also be implicitly expressed as a superposition of orbits with Schwarzschild methods (e.g., [Schwarzschild 1979](#); [Chanamé et al. 2008](#); [van den Bosch et al. 2008](#); [Breddels et al. 2013](#)) or with made-to-measure methods, in which N -body simulations are evolved to fit some data constraints (e.g., [Syer & Tremaine 1996](#); [de Lorenzi et al. 2007](#); [Dehnen 2009](#); [Hunt & Kawata 2013](#)). In addition, [Li et al. \(2025\)](#) proposed a novel nonparametric distribution function method. Distribution function methods have the advantages that no data binning is required and f is determined by the data. However, if the functional form for f (or its implicit specification via other methods) does not encompass the true solution, then bias can creep in.

Finally, numerous machine-learning methods have been proposed that learn from training data to either accelerate distribution function fitting or incorporate cosmological priors into the mass modeling pipeline (e.g., [Expósito-Márquez et al. 2023](#); [Nguyen et al. 2025](#); [Lim et al. 2025](#); [Sarrato-Alós et al. 2025](#)).

Given the proliferation of mass modeling methods, we may wonder why yet another method is needed. However, there are two key problems that remain to be addressed, even for spherical stellar systems. Firstly, while the amount and quality of data continue to improve, there will always be interesting systems with very few stellar tracers (e.g., [Collins et al. 2021](#); [Smith et al. 2024b](#); [Pickett et al. 2025](#)). Typically, simple mass estimators are used for low tracer numbers, but these have been shown to be potentially significantly biased ([Laporte et al. 2019](#)). More sophisticated, unbiased, mass modeling methods for low tracer numbers are only just being developed (e.g., [Nguyen et al. 2025](#)).

Secondly, there is a well-known degeneracy between the cumulative mass (which we are typically most interested in) and the velocity anisotropy profile, which is a measure of how circular or radial the orbit distribution is (e.g., [Dejonghe 1987](#); [Merritt](#)

[1987](#); [Merrifield & Kent 1990](#); [Wilkinson et al. 2002](#); [Łokas & Mamon 2003](#); [de Lorenzi et al. 2009](#); [Read & Steger 2017](#); [Read et al. 2021](#)). For Jeans modeling, this occurs because the Jeans equations do not constitute a closed system, meaning that for a given mass model and tracer profile, a unique solution is not guaranteed (e.g., [Binney & Tremaine 2008](#)). This typically arises when only projected line-of-sight (LOS) velocities are available.

Two key methods have been proposed to break the mass-anisotropy degeneracy. Firstly, we can incorporate higher moments of the velocity distribution as data constraints. At infinite order, these fully specify the phase-space distribution function and ensure its positivity ([Merrifield & Kent 1990](#); [Dejonghe & Merritt 1992](#); [Richardson & Fairbairn 2013](#)). After the usually-used second moment, the next nontrivial moment for a spherical, nonrotating, pressure-supported system is the fourth moment. This encodes information about deviations from Gaussianity, including the “tailedness” of the distribution and the kurtosis. It can be incorporated directly into the Jeans equations, thereby introducing a new fourth-order anisotropy parameter ([Merrifield & Kent 1990](#); [Łokas & Mamon 2003](#); [Łokas et al. 2005](#); [Richardson & Fairbairn 2013](#); [Wardana et al. 2025](#)). Alternatively, we can perform luminosity-weighted integrals over the fourth moment to derive the virial shape parameters (VSPs) that marginalize out the higher-order anisotropy ([Merrifield & Kent 1990](#); [Richardson & Fairbairn 2014](#); [Read & Steger 2017](#)). Secondly, we can include proper motions (PMs) to gain additional phase-space information that is independent of the LOS velocities (e.g., [Van der Marel & Anderson 2010](#); [Watkins et al. 2013](#)). Until recently, such PMs have only been available for GCs (e.g., [McNamara et al. 2003](#); [Anderson & Van der Marel 2010](#); [Vitrail et al. 2022](#); [Dickson et al. 2023](#); [Häberle et al. 2024](#); [Bañares-Hernández et al. 2025](#)). However, data are becoming increasingly available for nearby dwarf galaxies, with recent mass models incorporating PMs recently reported for the Draco and Sculptor dSphs ([Vitrail et al. 2024, 2025](#)).

Although Jeans mass modeling methods typically use data binning (e.g. [Read & Steger 2017](#)), this is not actually a requirement. Several authors have explored assuming a local Gaussian distribution function constrained by the second moment to create a bin-free Jeans method (e.g., [Watkins et al. 2013](#); [Mamon et al. 2013](#)). With the widespread availability of individually resolved stellar velocities, this approach is preferable and has all the advantages of having f that distribution function methods enjoy².

In this work, we present the GRAVSPHERE2 mass modeling method. This builds on GRAVSPHERE ([Read & Steger 2017](#); [Collins et al. 2021](#)) with the specific goals of: (i) reducing bias; (ii) maximizing the information content in the data; (iii) avoiding the need for data binning; and (iv) having the capacity able to successfully model, without bias, very low numbers of tracers. To achieve these goals, we self-consistently solved the second- and fourth-order Jeans equations, following the framework in [Richardson & Fairbairn \(2013\)](#), extending it to include a new observable: fourth-order PMs. We show that this approach breaks the degeneracies in the higher order Jeans equations by solving along their three projected components. Unlike previous approaches, which have made more restrictive assumptions (e.g., [Merrifield & Kent 1990](#); [Łokas & Mamon 2003](#); [Łokas et al. 2005](#); [Richardson & Fairbairn 2013](#); [Wardana et al. 2025](#)), we also included fully independent and radially varying anisotropies

¹ This is particularly timely when assessing increasingly large samples with thousands (in some cases approaching millions, [Häberle et al. 2025](#)) of tracers.

² This is true, up to a given order of moments, provided that the Jeans method has those moments self-consistently implemented in the velocity distribution function.

at both orders. We modeled the individual stellar velocities, rather than binning the data, by introducing a local non-Gaussian velocity probability distribution function (PDF) constrained by the second and fourth moments. We also modeled the positions of stars without binning, when the relevant data were available. We tested these improvements using mock data for dSphs and put together the Python package, GRAVSPHERE2, which is now publicly available³.

This paper is organized as follows. In Sect. 2, we introduce the modeling methodology behind GRAVSPHERE2, including our method for solving the fourth-order Jeans equations, which we applied following Richardson & Fairbairn (2013) and extended to include PMs (Sect. 2.1). We also describe its implementation using non-Gaussian local velocity PDFs for individual tracers, following Sanders & Evans 2020 (Sect. 2.2), and the functional forms we assume for our mass model and the second and fourth-order stellar velocity anisotropy profiles (Sect. 2.3). In Sects. 3.1 and 3.2, we introduce the mock data we used to test GRAVSPHERE2. In Sect. 3.3, we describe how we integrate the approach introduced in Sect. 2.1 to fit these data, specifying the form of the likelihood function and our assumed priors. Sect. 4 shows the results of our tests of GRAVSPHERE2 on the mock data. We compared these results with previous methods that were applied to these same data in the literature, including previous versions of GRAVSPHERE. We also explore the limit of low tracer numbers (i.e., corresponding to the faintest or most distant systems) and we compare the performance of GRAVSPHERE2 with simple mass estimators that are widely used in the literature. In Sect. 5, we briefly discuss the significance of this work and add our concluding remarks, highlighting upcoming works that make use of the new GRAVSPHERE2 method.

2. Modeling stellar kinematics with GravSphere2

GRAVSPHERE is a numerical Jeans solver that was originally introduced in Read & Steger (2017). It has been further extended and extensively tested with mock data (Read & Steger 2017; Read et al. 2018; Gregory et al. 2019; Genina et al. 2020; Read et al. 2021; Collins et al. 2021; Genina et al. 2022; De Leo et al. 2024; Nguyen et al. 2025; Tchiorniy & Genina 2025), including simulations deliberately chosen to explore expected bias in dynamical models due to breaking of key assumptions, such as spherical symmetry (triaxial mocks) and dynamical pseudo-equilibrium (tidally stripped mocks). A public version of this GRAVSPHERE1 code, called PYGRAVSPHERE, was introduced in Genina et al. (2020)⁴. Its most recent version was introduced and made publicly available in Collins et al. (2021)⁵. It improved the velocity binning with the BINULATOR feature, improved the mass model and improved the way that the VSPs contributed to the likelihood. Henceforth, we refer to this version as GRAVSPHERE1.5. Júlio et al. (2023) also introduced and made publicly available a version of GRAVSPHERE1.5 modified to test scalar field dark matter models in dwarf galaxies⁶. A version of GRAVSPHERE1.5 with an explicit treatment of rotation has also been presented in Júlio et al. (2025). Lastly, self-consistent inclusion of millisecond pulsar accelerations as independent kinematic tracers to constrain mass-anisotropy models in star clusters and improved photometric modeling with an $\alpha\beta\gamma$ profile were also

introduced in Bañares-Hernández et al. (2025), who used it to probe the presence of an IMBH and stellar remnant populations in Omega Centauri.

GRAVSPHERE has been a popular tool which has been used in the study of the dark matter distribution of nearby dwarf galaxies (e.g., Read et al. 2018, 2019; Gregory et al. 2019; Alvarez et al. 2020; Alvey et al. 2021; Collins et al. 2021; Zoutendijk et al. 2021; Júlio et al. 2023; De Leo et al. 2024; Zimmermann et al. 2025; Pickett et al. 2025; Júlio et al. 2025), galaxy clusters (Li et al. 2023), and to model GCs (Bañares-Hernández et al. 2025). In this section, we introduce the improved methodology of GRAVSPHERE2 for the mass modeling of stellar kinematics. We briefly summarize the key elements of previous incarnations of GRAVSPHERE, focusing here on the novel aspects with respect to these earlier versions (Read & Steger 2017; Read et al. 2018; Genina et al. 2020; Read et al. 2021; Collins et al. 2021). Crucially, these improvements include the removal of binning in favor of realistic self-consistent velocity distribution functions (Sect. 2.2), the incorporation of the fourth-order Jeans equations to treat higher moments, with a more general treatment than previous approaches (Sect. 2.1), and the inclusion of higher order PMs. The details of mass-anisotropy models are addressed in Sect. 2.3.

2.1. An extended Jeans analysis up to fourth order

GRAVSPHERE2 numerically solves the Jeans equations for a spherical, nonrotating system of tracers, with radial density profile, $\nu_\star(r)$, in a dynamical pseudo-equilibrium (e.g., Binney & Tremaine 2008), expressed as

$$\frac{1}{\nu_\star} \frac{\partial \nu_\star \sigma_r^2}{\partial r} + \frac{2\beta \sigma_r^2}{r} = -\frac{GM}{r^2}, \quad (3)$$

where

$$\beta \equiv 1 - \frac{\sigma_t^2}{2\sigma_r^2} \quad (4)$$

is the velocity anisotropy⁷ and we assume Newtonian weak-field gravity with an enclosed mass, $M(r)$, within the 3D radial coordinate, r .

As discussed in Bañares-Hernández et al. (2025), the assumption of nonrotation can be relaxed to some extent by solving for the second velocity moment directly. This is what we aim to do in this study. This is done by subtracting only global averages to the velocity distribution, leaving potential rotation signatures present. These are then reabsorbed when solving the Jeans equation, leading to a quadrature term that is implicitly added to the pure velocity dispersion (see, e.g., the discussion in Fritz et al. 2016).

The solution for the radial velocity dispersion (Mamon & Łokas 2005; van der Marel 1994) is then given by

$$\sigma_r^2(r) = \frac{1}{\nu_\star(r)g(r)} \int_r^\infty \frac{GM(r')\nu_\star(r')}{r'^2} g(r') dr', \quad (5)$$

where

$$g(r) \equiv \exp\left(2 \int \frac{\beta(r)}{r} dr\right). \quad (6)$$

³ Available at: <https://github.com/dadams42/GravSphere2>

⁴ Available at: <https://github.com/AnnaGenina/pyGravSphere>

⁵ Available at: <https://github.com/justinread/gravsphere>

⁶ Available at: https://github.com/marianajulio/alternative_models_for_gravsphere

⁷ Unlike in, e.g., Bañares-Hernández et al. (2025), σ_t corresponds to the total 2D tangential velocity dispersion, rather than the 1D averaged one (thus the factor of two). This is purely notational, but we adopted this change because we found it more convenient to work with this quantity for the subsequent parts of this section.

Projecting along the observable LOS and tangential and radial PM components, respectively, we obtain

$$\sigma_{\text{LOS}}^2(R) = \frac{2}{\Sigma_\star(R)} \int_R^\infty \left(1 - \frac{R^2}{r^2} \beta(r)\right) \frac{\nu_\star(r) \sigma_r^2(r) r}{\sqrt{r^2 - R^2}} dr, \quad (7)$$

$$\sigma_{\text{PM, t}}^2(R) = \frac{2}{\Sigma_\star(R)} \int_R^\infty \left(1 - \beta(r)\right) \frac{\nu_\star(r) \sigma_r^2(r) r}{\sqrt{r^2 - R^2}} dr, \quad (8)$$

and

$$\sigma_{\text{PM, R}}^2(R) = \frac{2}{\Sigma_\star(R)} \int_R^\infty \left(1 - \beta(r) + \frac{R^2}{r^2} \beta(r)\right) \frac{\nu_\star(r) \sigma_r^2(r) r}{\sqrt{r^2 - R^2}} dr, \quad (9)$$

where $\Sigma_\star(R)$ is the projected stellar tracer surface density.

By considering these three separate components, provided that one has enough data, it is possible to alleviate the well-known problem of the degeneracy between the cumulative mass, $M(r)$, and stellar velocity anisotropy, $\beta(r)$ (Merrifield & Kent 1990; Wilkinson et al. 2002; Lokas & Mamon 2003; de Lorenzi et al. 2009; Read & Steger 2017). Further, we can resort to higher moments of the velocity distribution to place additional constraints that have also been found to alleviate this problem. One popular choice is to use VSPs (Merrifield & Kent 1990; Richardson & Fairbairn 2014; Read & Steger 2017), which are two numbers obtained by integrating the luminosity-weighted, fourth-order LOS velocity moment to infinity, thereby marginalizing out any higher order velocity anisotropy. This approach is used in GRAVSPHERE1 and GRAVSPHERE1.5.

The key advantage of using VSPs over explicit modeling of the higher order moment equations is that no higher order analog of $\beta(r)$ is required. However, a key disadvantage is that the VSPs require integrals to infinity over the surface brightness profile and fourth velocity moment, which may both be poorly constrained at large radii. Furthermore, their large radius behavior could be adversely affected by disequilibrium due to tides. Both effects can lead to bias (e.g., Read & Steger 2017; Read et al. 2018; De Leo et al. 2024; Tchiorny & Genina 2025). For this reason, in GRAVSPHERE2 we take a different approach that (as we show in Sect. 4) is less biased. Instead of using VSPs, we explicitly solve the fourth-order Jeans equations, introducing a flexible functional form for the associated higher order anisotropy that we can then marginalize over. This is more similar to the approach introduced in Richardson & Fairbairn (2013). Explicitly, we solve the fourth-order moment Jeans equations (Merrifield & Kent 1990), expressed as

$$\frac{1}{\nu_\star} \frac{\partial \nu_\star \langle v_r^4 \rangle}{\partial r} - \frac{3}{r} \langle v_r^2 v_t^2 \rangle + \frac{2}{r} \langle v_r^4 \rangle = -3\sigma_r^2 \frac{GM}{r^2}, \quad (10)$$

$$\frac{1}{\nu_\star} \frac{\partial \nu_\star \langle v_r^2 v_t^2 \rangle}{\partial r} - \frac{1}{r} \langle v_t^4 \rangle + \frac{4}{r} \langle v_r^2 v_t^2 \rangle = -\sigma_t^2 \frac{GM}{r^2}, \quad (11)$$

where the bra-kets “ $\langle \rangle$ ” are used to denote the different velocity moments of the distribution.

As shown in Richardson & Fairbairn (2013), this system of equations can be further simplified, without loss of generality by first introducing the fourth-order anisotropy analog,

$$\beta' \equiv 1 - \frac{3 \langle v_r^2 v_t^2 \rangle}{2 \langle v_r^4 \rangle}, \quad (12)$$

which allows the first fourth-order Jeans expression in Eq. (10) to be recast in the more familiar form of

$$\frac{1}{\nu_\star} \frac{\partial \nu_\star \langle v_r^4 \rangle}{\partial r} + \frac{2\beta'}{r} \langle v_r^4 \rangle = -3\sigma_r^2 \frac{GM}{r^2}, \quad (13)$$

which, by comparison to the (second-order) Jeans Eq. (3), can be readily solved for the fourth-order radial velocity moment, yielding

$$\langle v_r^4 \rangle(r) = \frac{1}{\nu_\star(r) g'(r)} \int_r^\infty 3\sigma_r^2(r') \frac{GM(r') \nu_\star(r')}{r'^2} g'(r') dr', \quad (14)$$

where

$$g'(r) \equiv \exp\left(2 \int \frac{\beta'(r)}{r} dr\right). \quad (15)$$

Therefore, by first solving Eq. (3) for σ_r^2 and then using this to obtain the solution $\langle v_r^4 \rangle$ for Eq. (13), we can self-consistently solve for the second- and fourth-order Jeans equations. This is the first key change we made in GRAVSPHERE2.

As for the second-order counterparts, we now want to obtain the projected quantities that are directly related to observations. To express these purely in terms of $\langle v_r^4 \rangle$, following Richardson & Fairbairn (2013), first, it follows from Eq. (12) that

$$\frac{\partial \nu_\star \langle v_r^2 v_t^2 \rangle}{\partial r} = \frac{2}{3} \left[(1 - \beta') \frac{\partial \nu_\star \langle v_r^4 \rangle}{\partial r} - \frac{\partial \beta'}{\partial r} \nu_\star \langle v_r^4 \rangle \right]. \quad (16)$$

Next, we made a substitution using the second fourth-order Jeans Eq. (10) for the first term on the right-hand side,

$$\frac{\partial \nu_\star \langle v_r^2 v_t^2 \rangle}{\partial r} = -\frac{2}{3} (1 - \beta') \left[\frac{2\beta'}{r} \nu_\star \langle v_r^4 \rangle + 3\nu_\star \sigma_r^2 \frac{GM}{r^2} \right] - \frac{\partial \beta'}{\partial r} \nu_\star \langle v_r^4 \rangle, \quad (17)$$

which can then be substituted into the second fourth-order Jeans Eq. (11) to yield an expression for the tangential fourth-order moment in terms of the radial one as⁸

$$\langle v_t^4 \rangle(r) = \frac{4}{3} \left[(1 - \beta)(2 - \beta') - \frac{r}{2} \frac{\partial \beta'}{\partial r} \right] \langle v_r^4 \rangle + 2(\beta - \beta') \sigma_r^2 \frac{GM}{r}. \quad (18)$$

By considering the geometry of the problem, expanding to fourth order, and simplifying in terms of radial moments for the projected LOS component, we find

$$\langle v_{\text{LOS}}^4 \rangle(R) = \frac{2}{\Sigma_\star(R)} \int_R^\infty \frac{F_{\text{LOS}}(r, R) \nu_\star(r) r}{\sqrt{r^2 - R^2}} dr, \quad (19)$$

where

$$F_{\text{LOS}}(r, R) \equiv \left(1 - 2\beta' \frac{R^2}{r^2} + \frac{1}{2} \beta' (1 + \beta') \frac{R^4}{r^4} - \frac{1}{4} \frac{\partial \beta'}{\partial r} \frac{R^4}{r^3} \right) \langle v_r^4 \rangle + \frac{3}{4} (\beta - \beta') \sigma_r^2 \frac{GM}{r^3}. \quad (20)$$

As noted in Richardson & Fairbairn (2013), this approach constitutes a significant theoretical improvement over second-order Jeans methods by ensuring physical consistency of a given mass model with higher order velocity moments. This is not trivially warranted, as Jeans models can in some cases lead to unphysical solutions implying negative distribution functions or be inconsistent with the velocity distribution that is implicitly assumed to map velocity dispersions (Dejonghe & Merritt 1992; Richardson & Fairbairn 2013). We note that this is explicitly addressed in

⁸ It should be understood that $\beta, \beta', M, \langle v_r^4 \rangle, \sigma_r$ are all, in general, functions of r (we suppressed these implicit dependencies for notational convenience).

Sect. 2.2. It also has the promise of serving as a means to further constrain mass models by considering additional higher order moments (Richardson & Fairbairn 2013).

One potential disadvantage, however, which was also noted in Richardson & Fairbairn (2013), is that the constraining power of this approach can be compromised due to the fact that a fully general treatment leads to the introduction of the fourth-order anisotropy analog, β' , meaning that one may not be able to unambiguously constrain solutions to the fourth-order Jeans equation for a given mass model and (second-order) anisotropy, β , leading to a problem that is similar to the original mass-anisotropy degeneracy. One approach to this problem is to impose a prescription that fixes β' . This can be done, for example, by assuming distribution function models which predict these (Łokas 2002; Łokas et al. 2005). More generally, as explored in Dejonghe & Merritt (1992); Richardson & Fairbairn (2013), assuming a particular form for the distribution function amounts to imposing a prescription to solve all the moments of the Jeans equations (beyond the fourth order included). One problem with this approach, however, is that it might restrict the space of solutions unnecessarily and bias the models, unless one is able to be sufficiently judicious and general in specifying a form for the distribution function.

To tackle this potential problem and to make full use of the available kinematic information, similarly to how we can use PM velocity dispersions to constrain the second-order system, we extend this methodology by also considering the fourth-order velocity moments for the tangential and radial PM components, constraining also the fourth-order system. We show that this information can be then used to further constrain the system while allowing for a general treatment of β' . As in Eq. (19), this leads to an expression for the tangential and radial fourth-order moments, which can be obtained (respectively) as

$$\langle v_{\text{PM},t}^4 \rangle(R) = \frac{2}{\Sigma_\star(R)} \int_R^\infty \frac{F_{\text{PM},t}(r, R) v_\star(r) r}{\sqrt{r^2 - R^2}} dr, \quad (21)$$

and

$$\langle v_{\text{PM},R}^4 \rangle(R) = \frac{2}{\Sigma_\star(R)} \int_R^\infty \frac{F_{\text{PM},R}(r, R) v_\star(r) r}{\sqrt{r^2 - R^2}} dr, \quad (22)$$

where

$$F_{\text{PM},t}(r, R) \equiv \frac{1}{2} \left[(1 - \beta')(2 - \beta') - \frac{r}{2} \frac{\partial \beta'}{\partial r} \right] \langle v_r^4 \rangle + \frac{3}{4} (\beta - \beta') \sigma_r^2 \frac{GM}{r}, \quad (23)$$

and

$$F_{\text{PM},R}(r, R) \equiv \left(1 - 2 \frac{R^2}{r^2} + \frac{R^4}{r^4} \right) F_{\text{PM},t}(r, R) + \left(2(1 - \beta') \frac{R^2}{r^2} - (1 - 2\beta') \frac{R^4}{r^4} \right) \langle v_r^4 \rangle. \quad (24)$$

In this way, we can exploit the increasing abundance of PM data found in nearby galactic systems and star clusters to constrain the fourth-order anisotropy, leading to a full three-dimensional and self-consistent modeling of both the second and fourth-order moments along the three observable LOS and PM components.

To our knowledge, this is a new point that has not been published in the literature to date⁹.

Lastly, we emphasize the following point: while PMs constitute an important and novel addition to the higher order treatment of kinematics, this does not mean that our approach will only be useful when these are available. In particular, the fact that quantities such as VSPs exist (i.e., fourth-order anisotropy-independent constraints on the higher moments) demonstrates that the Jeans equations should automatically encode this information, yielding comparable constraints if we marginalize over some general functional forms for β and β' . In fact, given that virial constraints amount to positivity constraints on velocity moments and that the virial theorem can itself be derived from the Jeans equations (Dejonghe & Merritt 1992), any quantities derived from these constraints should already be encoded in the Jeans equations. This means that the Jeans equations, solved at fourth order, should have at least as much information as using the Jeans equations at second order with VSPs. This appears to be an underappreciated fact: without additional assumptions, quantities such as VSPs should not provide any more information than velocity moments themselves from the Jeans equations (at a given order). Thus, any anisotropy-independent constraints derived from these are not inherently more constraining, but rather reflect a property of the Jeans equations themselves, which can be recovered after marginalization over the anisotropy. Therefore, we do not need to resort to these quantities to break the mass-anisotropy degeneracy and it may be possible to obtain better performance by working directly with the fourth-order Jeans equations instead.

In fact, VSPs can be susceptible to significant biases, since they rely on assumptions about the behavior of the light profile and fourth velocity moment at large radii where data constraints are poor. By contrast, in the approach we take here, we managed to avoid these potential biases; however, we had to instead assume some functional form for β' to marginalize over. In Sect. 4, we use mock data for the LOS velocities without PMs to demonstrate that the approach we take in GRAVSPHERE2 is indeed less biased than GRAVSPHERE1 or GRAVSPHERE1.5.

2.2. Velocity distribution function

Given the methodology discussed in Sect. 2.1 to derive the velocity moments from a given mass model, we need a prescription to constrain these moments from observable velocities. A common approach is to bin the data into discrete portions of a certain spatial width. While this has been shown to be sufficient to well-recover mass models from dwarf galaxy mock data (Read & Steger 2017; Read et al. 2018; Gregory et al. 2019; Genina et al. 2020; Read et al. 2021; Collins et al. 2021; De Leo et al. 2024; Nguyen et al. 2025; Tchiorniy & Genina 2025), it can lead to a loss of information and potential biases. It introduces an unavoidable trade-off between spatial and velocity resolution, which may be particularly acute when aiming to extract information of higher order moments, and it makes it challenging to properly marginalize over uncertain membership for the tracers, survey selection effects, binary stars, and so on.

To solve all of the above problems, in GRAVSPHERE2, we modeled each stellar velocity individually, without binning. In

⁹ We note that Evslin & Del Popolo (2017) independently introduced fourth-order PMs to quantify departures from dynamical equilibrium. However, to the authors' knowledge, our work is the first in which these are explicitly solved for from the Jeans equations and used to break the mass-anisotropy degeneracy and it is therefore a novelty in this respect.

the context of Jeans analyses, comparable approaches have been introduced in the past, such as DISCRETEJAM (Cappellari 2008; Watkins et al. 2013; Cappellari 2015), MAMPOSST (Mamon et al. 2013), and, more recently, in Wardana et al. (2025)¹⁰. In our case, rather than assuming Gaussianity for the velocity PDF (projected or otherwise), we used the PDF models introduced in Sanders & Evans (2020). These are based on analytical convolutions of a Gaussian with a uniform distribution kernel for models of negative excess kurtosis and with a Laplace kernel for positive excess kurtosis models. Taken together, they allow for meaningful deviations from Gaussianity with kurtoses $\kappa \equiv \langle v^4 \rangle / \sigma^4$ in the range $1.8 < \kappa < 6$, with $\kappa = 3$ corresponding to a Gaussian (which defines excess kurtosis). For a symmetric distribution of mean zero for the scaled velocity variable, $w \equiv v/\sigma$, and measurement error, $s \equiv \delta v/\sigma$ (assumed to be symmetric and Gaussian), this yields the error-convolved PDF of the form

$$f_{\text{vel}}(w) = \frac{b}{2a} \left[\Phi\left(\frac{bw+a}{t}\right) - \Phi\left(\frac{bw-a}{t}\right) \right], \quad (25)$$

for the uniform kernel ($1.8 < \kappa < 3$), where

$$\Phi(x) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x dy e^{-y^2/2}, \quad (26)$$

$$t^2 = 1 + b^2 s^2, \quad (27)$$

and the parameters a and b are fully specified by the variance and kurtosis (before convolving with errors), respectively, as

$$\langle w^2 \rangle = \frac{1}{b^2} \left(1 + \frac{a^2}{3} \right), \quad \kappa = 3 - \frac{2a^4}{15} \left(1 + \frac{a^2}{3} \right)^{-2}. \quad (28)$$

Setting $\langle w^2 \rangle = 1$, so that $\langle v^2 \rangle = \sigma^2$, implies

$$b^2 = 1 + \frac{a^2}{3}. \quad (29)$$

For the Laplacian kernel ($3 < \kappa < 6$), the error-convolved PDF takes the form

$$f_{\text{vel}}(w) = \frac{b}{4a} \left[\exp\left(\frac{t^2 - 2abw}{2a^2}\right) \text{erfc}\left(\frac{t^2 - abw}{\sqrt{2}ta}\right) + \exp\left(\frac{t^2 + 2abw}{2a^2}\right) \text{erfc}\left(\frac{t^2 + abw}{\sqrt{2}ta}\right) \right], \quad (30)$$

with variance and kurtosis

$$\langle w^2 \rangle = \frac{1}{b^2} (2a^2 + 1), \quad \kappa = 3 + 12a^4 (2a^2 + 1)^{-2}, \quad (31)$$

so that, again, setting unit variance yields

$$b^2 = 2a^2 + 1. \quad (32)$$

This allows for a fully self-consistent treatment between the assumed PDF and the stellar kinematic model¹¹.

¹⁰ The approach from Wardana et al. (2025), which was recently and independently developed, is the one most similar to the one we use in GRAVSPHERE2. Some of the main differences between these include the improvements which we introduce, such as a fully general treatment of anisotropies (we marginalized β and β' as two independent and radially dependent, i.e., nonconstant, quantities) and the use of PMs to constrain both the second and fourth-order Jeans equations. We also implement uncertainties differently in our PDF modeling, which may explain why these authors find greater susceptibility to error-driven biases in their results.

¹¹ Our numerical implementation of these functions is based on the code introduced in Sanders & Evans (2020), which is publicly available at: https://github.com/jls713/gh_alternative

2.3. Mass-anisotropy modeling

As in Bañares-Hernández et al. (2025), for the photometric tracer density profile, we use the generalized $\alpha\beta\gamma$ profile (e.g., Hernquist 1990; Zhao 1996). This is given by a double power-law model,

$$v_{\star}(r) = \frac{\rho_p}{(r/r_p)^\gamma (1 + (r/r_p)^\alpha)^{(\beta-\gamma)/\alpha}}, \quad (33)$$

where we introduced the three exponent variables α, β, γ , and the scale radius and density r_p, ρ_p , respectively.

Following recent versions of GRAVSPHERE (e.g., Collins et al. 2021), we used the CORENFWTIDES profile (Read et al. 2018) to model dark matter profiles of tidally stripped galactic halos expected from hydrodynamical simulations simulating Local Group environments. Previous versions of GRAVSPHERE have used nonparametric modeling with sequential power laws Read & Steger (2017); Genina et al. (2020). As compared to these earlier choices, the CORENFWTIDES profile has number of advantages: it does not suffer from biases due to monotonicity priors inherent to the previous method (Tchiorniy & Genina 2025); rather, it is designed to model mass profiles realistic galactic systems, and it makes recovering cosmologically valuable parameters such as the halo mass and concentration straightforward. The model has an enclosed mass profile given by

$$M_{\text{cNFWt}}(r) = \begin{cases} M_{\text{cNFW}}(r) & r \leq r_t, \\ M_{\text{cNFW}}(r) + 4\pi\rho_{\text{cNFW}}(r_t) \frac{r_t^3}{3-\delta} \left[\left(\frac{r}{r_t} \right)^{3-\delta} - 1 \right] & r > r_t, \end{cases} \quad (34)$$

where

$$M_{\text{cNFW}}(r) = M_{\text{NFW}}(r) f^n \quad (35)$$

is the CORENFW profile (Read et al. 2016) and we have the original Navarro-Frenk-White (NFW) profile (Navarro et al. 1996), expressed as

$$M_{\text{NFW}}(r) = M_{200} g_c \left[\ln\left(1 + \frac{r}{r_s}\right) - \frac{r}{r_s} \left(1 + \frac{r}{r_s}\right)^{-1} \right], \quad (36)$$

with a function that modulates the degree of core formation,

$$f^n = \left[\tanh\left(\frac{r}{r_c}\right) \right]^n, \quad (37)$$

and the densities given by

$$\rho_{\text{cNFW}}(r) = f^n \rho_{\text{NFW}} + \frac{n f^{n-1} (1 - f^2)}{4\pi r^2 r_c} M_{\text{NFW}}, \quad (38)$$

$$\rho_{\text{NFW}}(r) = \rho_s \left(\frac{r}{r_s} \right) \left(1 + \frac{r}{r_s} \right)^{-2}, \quad (39)$$

with the scale density and length given (respectively) as

$$\rho_s = \rho_{\text{crit}} \Delta c_{200}^3 g_c / 3, \quad r_s = r_{200} / c_{200}, \quad (40)$$

with:

$$g_c = \frac{1}{\ln(1 + c_{200}) - \frac{c_{200}}{1 + c_{200}}}, \quad (41)$$

and

$$r_{200} = \left(\frac{3M_{200}}{4\pi\Delta\rho_{\text{crit}}} \right)^{1/3}, \quad (42)$$

with c_{200} being the dimensionless concentration parameter and r_{200} , M_{200} as the virial radius and mass, respectively, defined at a region where the average density exceeds the present-day cosmological critical density $\rho_{\text{crit}} = 135.05 \text{ M}_\odot \text{ kpc}^{-3}$ by the critical overdensity factor $\Delta = 200$ ¹². The core formation parameter n yields full core formation when $n = 1$, retrieves an NFW cusp when $n = 0$, and can model potentially cuspy distributions for negative values. The core radius, r_c , determines the extent of the inner core region. The tidal radius, r_t , and slope exponent, δ , regulate the amount of stripping in the outer region.

Following previous versions of GRAVSPHERE, we use the anisotropy profile assuming the following the generalized form from Baes & van Hese (2007),

$$\beta(r) = \beta_0 + (\beta_\infty - \beta_0) \frac{1}{1 + \left(\frac{r_\beta}{r} \right)^\eta}, \quad (43)$$

where β_0 is the anisotropy at the center, β_∞ its limit as it asymptotically approaches infinity, r_β is the transition scale between these two regimes, and η is the exponent which modulates the steepness of the transition. We also assumed the same functional form for the fourth-order anisotropy, β' , allowing for a fully general treatment of this quantity, unlike previous works that have made more restrictive assumptions about its form (e.g., Merrifield & Kent 1990; Lokas & Mamon 2003; Lokas et al. 2005; Richardson & Fairbairn 2013; Wardana et al. 2025).

3. Data, likelihood, and priors

3.1. Gaia Challenge mock data

To test the overall performance of GRAVSPHERE2, we started with a subset of the GAIA CHALLENGE catalog of mock galaxies¹³, constructed from spherical dynamical equilibrium models. The results are shown in Sect. 4.1. They have the advantage of having been extensively used for prior testing in the literature, allowing for a straightforward comparison with previous studies, including previous versions of GRAVSPHERE (e.g., Gregory et al. 2019; Read et al. 2021; Collins et al. 2021). As in Collins et al. (2021), we focus on a subset of two mock galaxies originally from Walker & Peñarrubia (2011) which have been the ones for which several previous methods (including GRAVSPHERE) have been affected by significant bias (e.g., see Read et al. 2021). These correspond to a dark matter profile with the $\alpha\beta\gamma$ functional form (same as Eq. (33)) embedded in a Plummer-like¹⁴ stellar tracer distribution, with a varying Osipkov-Merritt anisotropy profile (Osipkov 1979; Merritt 1985). These two mock galaxies consist of a cored dark matter model (PlumCoreOM) and a cuspy one (PlumCuspOM). We consider cases with 100, 1000,

and 10 000 tracers, with and without including PMs. In all cases we simultaneously fit the binned surface density profile using 10 000 tracers, assuming also velocities with 2 km/s errors.

3.2. Simulated dwarf galaxy data

The second case we considered is the recent suite of simulated dwarf galaxies from Tchiorniy & Genina (2025). These are initialized with cuspy NFW profiles with massive Plummer-like stellar components and are evolved in a Milky Way-like potential to simulate the properties of nearby present-day dSphs. This allows us to probe more realistic galactic environments and aspects such as departures of equilibrium and departures from spherical symmetry due to tidal effects, or induced rotation, which may be present in simulations.

For the purposes of this study, we focus on the Fornax analog of this catalog which was extensively studied, has a large number of tracer particles and was found to produce significant biases when applying PYGRAVSPHERE (GRAVSPHERE1)¹⁵. We studied the case with all the 34 158 bound tracers¹⁶, using only the LOS velocities used in Tchiorniy & Genina (2025). We also considered a reduced samples of 1000 and 100 tracers. Additionally, we considered a sample of 100 and 1000 tracers where PMs are included. We also considered LOS samples with 10, 25, and 50 tracers for our comparisons with mass estimators. For the tracer profile fit, we used 10 000 tracer positions in all cases, except for the fit with all tracers from the simulation. Velocities were perturbed with 2 km/s errors in all cases.

3.3. Likelihood and priors

The full log-likelihood we adopt for the fitting process is given by

$$\ln \mathcal{L}_{\text{tot}}(\theta) = \ln \mathcal{L}_{\text{LOS}} + \ln \mathcal{L}_{\text{PM,t}} + \ln \mathcal{L}_{\text{PM,R}} + \ln \mathcal{L}_{\Sigma_*}, \quad (44)$$

where θ includes all the model parameters, which were simultaneously fitted. This includes the anisotropy parameters (that are distinct for the second and fourth-order anisotropy profiles), the mass model parameters, and the photometric tracer profile parameters. The first three terms correspond to the likelihood due to individual velocity PDFs (Eqs. (25) and (30)) for the LOS and tangential and radial PM components for the data we considered, whilst the last term corresponds to the term from the photometric tracer profile. GRAVSPHERE2 incorporates two options in this respect: one for a binned photometric light profile with Gaussian errors and a completely ‘bin-free’ option where individual stellar tracer positions are used instead. We note that GRAVSPHERE2 does not bin any data. For the case of the binned profile it is assumed that this is provided as a data constraint from observations. Unlike previous versions of GRAVSPHERE, it is not necessary to derive a binned profile from individual stars as we can simply use the ‘bin-free’ approach instead. In the latter case, similarly to the case of kinematic tracers, we can write the total likelihood as the product of those of individual photometric

¹² The precise value of the adopted cosmological critical density will vary depending on the adopted cosmology. Our value corresponds to the ‘canonical value’ with Hubble parameter $H_0 \sim 70 \text{ km s}^{-1} \text{ Mpc}$, so that $\rho_{\text{crit}} = 3H_0^2/8\pi G$. It should be noted that various authors use different values of Δ to define the virial radius.

¹³ See the GAIA CHALLENGE website for full access to the data and further information: <https://astrowiki.surrey.ac.uk/>

¹⁴ In practice, these profiles closely resemble a Plummer sphere but show minor deviations in the form of an $\alpha\beta\gamma$ profile (Eq. (33)), which ensures positivity of the phase-space distribution function.

¹⁵ In particular, we are referring to variant with a light Milky Way potential, as opposed to the heavy one, chosen due to the properties we outlined and the fact that it was also more extensively studied by these authors in this context.

¹⁶ As per the subhalo finder, and applying also a 3σ -velocity clipping to the complete sample. For datasets where we use PMs, we require tracers to pass a 3.5σ -clipping test along each direction, reducing the sample to 34 081 tracers.

tracers, whose (positional) PDFs take the form

$$f_{\text{pos}}(R) = \frac{R\Sigma_{\star}(R)}{\int_0^{\infty} R'\Sigma_{\star}(R') dR'}. \quad (45)$$

In all cases, we verified that our fits were able to reproduce the photometric tracer profiles well and recover the true solution. We show a representative example fit in Appendix B.1 (Fig. B.2), showing excellent recovery which is far in excess of the range allowed by our priors. The advantages of the $\alpha\beta\gamma$ functional form will be more evident when reproducing more complex distributions which cannot be reduced to Plummer profiles (or a summation of these), this includes cusper distributions or those with shallow slopes at higher radii. One example of this is the GC Omega Centauri, where this functional form was introduced for GRAVSPHERE (Bañares-Hernández et al. 2025).

As in Bañares-Hernández et al. (2025), we worked with the symmetrized anisotropy (for both the second- and fourth-order anisotropy profiles), expressed as

$$\tilde{\beta} \equiv \frac{\beta}{2 - \beta}, \quad (46)$$

which has the property that is bounded as $-1 \leq \tilde{\beta} \leq 1$, to give an unbiased sampling of the full parameter space.

For the GAIA CHALLENGE data (Sect. 3.1), the priors we adopt for the mass and anisotropy profile are similar (albeit broader for some quantities) to those Section 4.1.2 of Collins et al. (2021), whilst for the photometric component we allow a flat prior with limits set with 50% scatter around the true solution, which we fit as a binned surface density profile using 10 000 tracers. A summary of all priors adopted is shown in Table 1.

For the Fornax-like simulated dwarf galaxy (Sect. 3.2), our mass and anisotropy priors were also the same as for the Gaia Challenge mocks, except for the fact the we used unbounded values for the limiting anisotropies ($-1 < \tilde{\beta}_{0,\infty} < 1$) and somewhat less broad priors for the concentration parameter ($1 < c_{200} < 50$). For the stellar tracer model, we did not use binned photometry and instead use the full “bin-free” method, using Eq. (45). We allowed for a scatter of 1 dex from the 2D half-light radius ($R_{1/2}$) of ~ 770 pc for the length scale in the photometric light profile (which we sample logarithmically as a flat prior)¹⁷ and $0.1 < \alpha < 4$, $3.1 < \beta < 7$, $0 < \gamma < 2.9$. These were chosen to allow for convergence (needed to set a proper normalization) while allowing a wide range of profiles of varying steepness. We recall here that γ and β are the negative inner and outer logarithmic slopes of the light profile (respectively), where a Plummer profile corresponds to $\alpha, \beta, \gamma = 2, 5, 0$. Lastly, we sampled the mass of the stellar component with a flat prior with 25% scatter around $4.3 \times 10^7 M_{\odot}$, as in Tchiorniy & Genina (2025), assuming a constant mass-to-light ratio and that the stellar mass distribution has the same radial density profile as the tracer stars. This is usually a good approximation provided that the tracer profile is complete and that there is no meaningful mass segregation or observational bias in the photometric selection. The latter is typically a good approximation for nearby dwarf galaxies, which have relaxation times longer than a Hubble time due to the presence of dark matter (e.g., Baumgardt et al. 2022). If this is not the case (as in star clusters), we need to include the presence of

Table 1. Priors implemented in DYNESTY for our analysis.

Parameter	Units	Prior	Range
$\tilde{\beta}_0, \tilde{\beta}'_0$	None	Flat	$[-0.01, 0.01]$ (G) $[-1, 1]$ (F)
$\tilde{\beta}_{\infty}, \tilde{\beta}'_{\infty}$	None	Flat	$[-0.1, 1]$ (G) $[-1, 1]$ (F)
r_{β}, r'_{β}	$\log_{10}$ kpc	Flat	$[-2, 1]$
η, η'	None	Flat	$[1, 3]$
M_{200}	$\log_{10} M_{\odot}$	Flat	$[7.5, 11.5]$
c_{200}	None	Flat	$[1, 100]$ (G) $[1, 50]$ (F)
r_c	$\log_{10}$ kpc	Flat	$[-2, 1]$
n	None	Flat	$[0, 1]$
r_i	$\log_{10}$ kpc	Flat	$[0, \log_{10}(20)]$
δ	None	Flat	$[3.01, 5]$
$M_{\star}$	$10^7 M_{\odot}$	Flat	$4.3 \pm 25\%$ (F)
ρ_p	kpc^{-3}	Flat	Truth $\pm 50\%$ (G)
r_p	kpc (G) $\log_{10}$ kpc (F)	Flat	Truth $\pm 50\%$ (G) $R_{1/2} \pm 1$ dex (F)
α	None	Flat	Truth $\pm 50\%$ (G) $[0.1, 4]$ (F)
β	None	Flat	Truth $\pm 50\%$ (G) $[3.1, 7]$ (F)
γ	None	Flat	Truth $\pm 50\%$ (G) $[0, 2.9]$ (F)

Notes. The parenthesized “G” denotes priors specific to the *Gaia* Challenge mocks, while “F” corresponds to the simulated Fornax dwarf analog. For both cases, there are 19 fitting parameters in total (for Fornax, we marginalize over the additional stellar mass $M_{\star}$, but do not use a density scale for the photometry, since we fit stars individually, rather than from a profile). Primed quantities (rows 1–4) denote the parameters of the fourth-order counterpart of the anisotropy. For the *Gaia* Challenge mocks, the exponents α, β, γ are also ensured not to exceed the broad convergence limits used for the Fornax case. The full implementation of these priors, with initialization files and data, are included in the GRAVSPHERE2 public release repository on [GitHub](https://github.com/dadams42/GS2-results).

additional components of varying segregation (e.g., see Dickson et al. 2023; Bañares-Hernández et al. 2025) in the mass model. The full implementation of our likelihood modeling with the priors and data we used; the GRAVSPHERE2 code itself has been made publicly available, as noted in the introduction (Sect. 1).

We also note that our priors are generally very broad and have little effect on our results. Some, such as imposing isotropy in the center and barring large negative anisotropies, which we do for the *Gaia* Challenge sample following Collins et al. (2021), are physically motivated constraints. We show explicitly in Appendix B that the space of mass and tracer density models spanned by our priors is much broader than those favored by the data, demonstrating that our priors are sufficiently broad so as to have little effect on our results. For completeness, we also included marginalized parameter constraints (corner plots) of each of the 19 parameters for all our models in a supplementary repository¹⁸.

All fits were performed using the nested sampling package DYNESTY (Speagle 2020; Koposov et al. 2022). This is based on nested sampling techniques (Skilling 2004, 2006) and the ‘dynamic nested sampling technique’ (Higson et al. 2019) (which we use here). DYNESTY has been optimized for the inference of

¹⁷ This can be obtained directly from the photometric tracers (either from the profile or their median position), but is only meant to set a relevant scale and its precise value is not needed as the results are insensitive to it.

¹⁸ <https://github.com/dadams42/GS2-results>

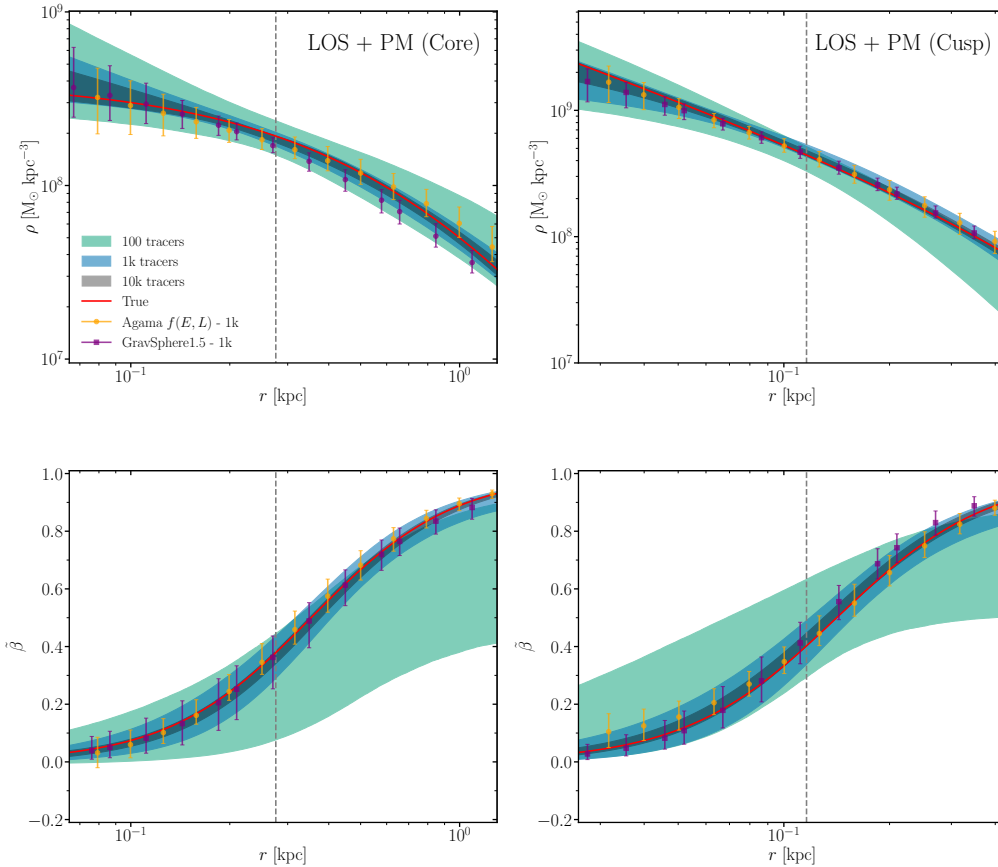


Fig. 1. *Top-left:* 3D dark matter density profile for the PlumCoreOM mock galaxy, bands denote the 95% CL regions from the median for different numbers of stellar tracers including both LOS and PM velocities. The results from Read et al. (2021) for AGAMA $f(E, L)$ models, and from Collins et al. (2021) for GRAVSPHERE1.5 (including LOS VSPs) for 1000 tracers with their respective median values with 95% CL errors are shown for comparison. The red line denotes the true analytical solution from which the mock was generated. The gray, dashed line corresponds to the projected half-light radius ($R_{1/2}$). *Top-right:* same as the left hand side but for the PlumCuspOM model. *Bottom-left:* symmetrized anisotropy profile for the PlumCoreOM model. The red line again denotes the true solution from which the mock was generated, corresponding to an Osipkov-Merritt profile. *Lower right:* same as the left hand side but for the PlumCuspOM model.

posterior distributions (as opposed to evidence estimation), using the bounding method from Feroz et al. (2009). The sampling method used is based on Neal (2003); Handley et al. (2015a,b)¹⁹. An example with the specific implementation used for GRAVSPHERE2 can be found in the publicly available code release. This sampling method has been shown to outperform Markov chain Monte Carlo Methods (MCMC), particularly when dealing with complex, high-dimensional, and multimodal posterior spaces (e.g., Speagle 2020). This is consistent with our own experiments comparing with previous versions of GRAVSPHERE that used EMCEE instead (Foreman-Mackey et al. 2013). Following Bañares-Hernández et al. (2025), we thus opted to use this improved sampling technique for GRAVSPHERE2.

Run times will vary depending on the number of tracer velocities used (including PMs), whether fitting the photometry star-by-star, binned, or fixing it, and also on the sampling resolution used. For our cases, we found that the default value of 500 live points sufficed. As a rule of thumb, a few times the number of dimensions squared usually suffices, but this will depend on how multimodal the posterior is, as well as how much volume the modes occupy (smaller volumes require higher sampling resolution). In all cases, models could be run within approximately 1–3 days on a normal laptop. The dominant source of computational cost in all versions of GRAVSPHERE is the evaluation of numerical integrals for solving the Jeans equations. For a given likelihood evaluation, GRAVSPHERE2 takes roughly twice as long as GRAVSPHERE1, since one solves the Jeans equations twice, once for second-order moments (as in previous versions), and another time for fourth-order ones. This still makes

GRAVSPHERE2 a highly competitive method in terms of run-time (cf. Read et al. 2021). We note that the run-times reported for GRAVSPHERE1 in Read et al. (2021) appear substantially less than what one might expect given the above. This is at least partly due to the fact that we use a more sophisticated and robust sampling method (DYNESTY rather than EMCEE). As noted earlier, runs can also be accelerated by reducing the number of live points, and we opted for a conservative choice in this respect. Likelihood evaluation times will also scale roughly as $\sim N_{\text{int}}^2$, where N_{int} is the number of points used for each numerical integral when solving the Jeans equations.

4. Results

4.1. Gaia Challenge mocks

Figs. 1 and 2 show the results for the mass density and anisotropy recovery of the two models considered, with and without the inclusion of PMs. These are shown as the regions in the posterior spanning 95% of the distribution centered at the median values of the profiles. For reference, we included the results of GRAVSPHERE1.5 (originally from Collins et al. 2021)²⁰, the most recent version of GRAVSPHERE until now, and the results presented in Read et al. (2021) for the AGAMA $f(E, L)$ distribution function-based models, which was in most cases the best-performing method (or amongst the best-performing ones) of the multiple ones considered in Read et al. (2021). This distribution function method is part of the widely-used AGAMA

¹⁹ For further information, see: <https://dynesty.readthedocs.io/en/latest/index.html>

²⁰ Note that Collins et al. (2021) only present their results for LOS tracers. We have, however, also included their (until now unpublished) results including also PMs.

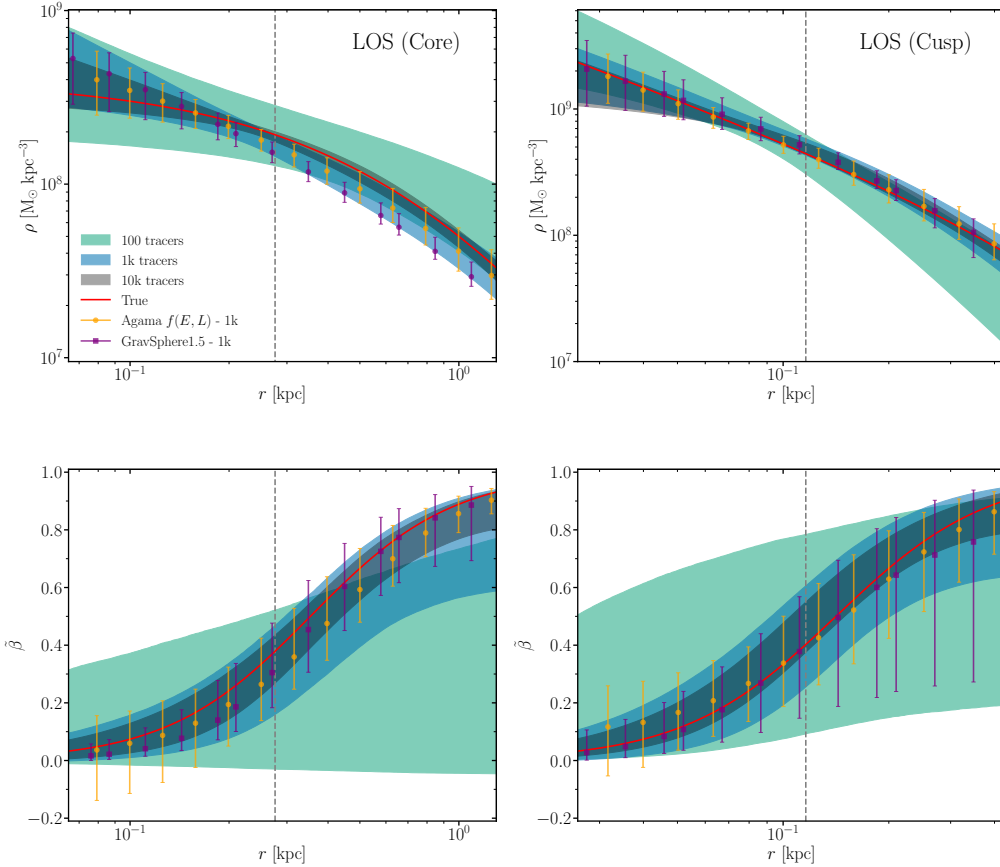


Fig. 2. Same as Fig. 1 but only with LOS tracers. *Top-left:* GRAVSPHERE2 recovery for the PlumCoreOM mock 3D mass density. *Top-right:* same as the left but for the PlumCuspOM model. *Bottom-left:* symmetrized anisotropy profile for the PlumCoreOM model. *Bottom-right:* same as the left but for the PlumCuspOM model.

library (Vasiliev 2019). These cases shown for comparison use 1000 tracers.

While GRAVSPHERE1.5 shows better performance than GRAVSPHERE1 in most cases for this mock sample (cf. Read et al. 2021), it still suffers from significant bias, particularly for the cored model, where densities are systematically underestimated beyond $\sim R_{1/2}$, even when PMs are included (Figs. 1 and 2). GRAVSPHERE2, on the other hand, is able to recover the true solutions within the 95% confidence-level (CL) region from the median within $0.1 \lesssim r/R_{1/2} \lesssim 10$ that encloses the vast majority of the tracers. The only exception to this concerns cases with limited data, where we can see that for the core model with 100 LOS tracers (Fig. 2), the model underestimates the anisotropy beyond $\sim 2R_{1/2}$, but is nonetheless able to recover the density everywhere at the specified confidence. For the 1000-tracer case, the true solution also lies at the edge of the 95% CL region beyond $R_{1/2}$, though these mild biases disappear when the number of tracers is increased or PMs are included, which is not the case with GRAVSPHERE1.5. The performance is otherwise excellent and in all cases competitive with sophisticated modeling approaches such as AGAMA, which shows a similar performance in most cases. We even see some instances where GRAVSPHERE2 shows less bias than AGAMA (e.g., Fig. 1, upper left) at larger radii, and reduced errors overall²¹.

Another illustrative case is shown in Fig. 3, where we compare the core model with 10 000 LOS tracers. Here all the models

considered by Read et al. (2021) were biased in either the density or anisotropy recovery, with GRAVSPHERE1.5, GRAVSPHERE1, and AGAMA showing a degree of bias in both. GRAVSPHERE2, on the other hand, is able to obtain a 95% CL recovery everywhere throughout. Models that do not capture deviations of Gaussianity in this case suffer particularly strong bias. To illustrate this, we included the results of DISCRETEJAM axisymmetric Jeans models, which assume Gaussian velocity PDFs. We then compared this with the result of imposing a Gaussian PDF in our models and not considering higher moments. This yields a similar result in both cases with a clearly biased model showing significantly greater slopes and uncertainty. GRAVSPHERE1.5 and GRAVSPHERE1, with VSPs, and AGAMA, which also allows more generality in the velocity PDF modeling, are able to achieve better recovery than the other models considered in Read et al. (2021), but still suffer from a degree of bias across much of the radial range shown ($0.25 \lesssim r/R_{1/2} \lesssim 4$). We also note that the anisotropy priors imposed by Collins et al. (2021) on GRAVSPHERE1.5, which we also impose, mean that our models are barred from large negative anisotropies and are close to isotropic at the center. Similarly, AGAMA models are fixed to be purely radially anisotropic at large radii. This partly explains why these models tend to be more tightly constrained toward the edges of our anisotropy plots, but, at least for GRAVSPHERE2, our results are otherwise insensitive to these priors.

Our results show that the biases we observe can be overcome in GRAVSPHERE2 by including a full treatment of higher moments via the extended Jeans equation (rather than using VSPs, which can be prone to bias themselves when determining them; see the discussion at the end of Sect. 2.1). The inclusion of these higher moments also removes negative-moment models

²¹ We emphasize, however, that we are referring to the AGAMA model applied in Read et al. (2021). This means that the biases that we see strictly speaking apply only to the phase-space distribution function parametrization adopted by these authors.

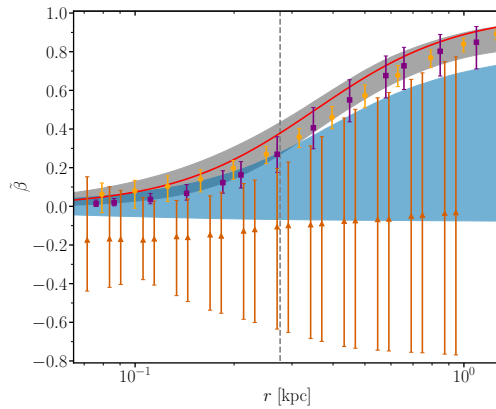
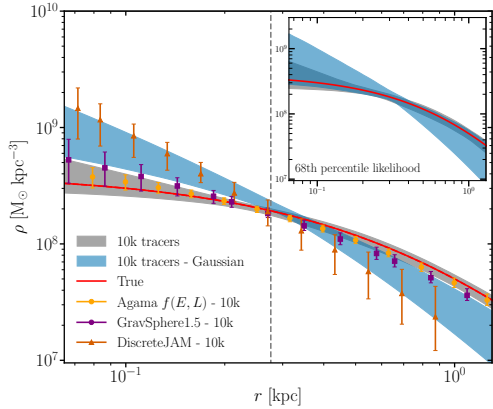


Fig. 3. Same as Fig. 2 but for 10000 tracers in the Plum-CoreOM mock, including in the comparison with other models. We have included the DISCRETEJAM results from Read et al. (2021). *Left:* mass density profiles. Blue band shows our results when Gaussianity is assumed and fourth-order moments are excluded. For comparison, the inset shows instead bands obtained from the posterior distribution of likelihoods when computing extrema of all models falling within the 68th percentile from the maximum likelihood (see text for details). Using this definition, there is a wide range of equally good-fitting models that encompasses the true solution, while the posterior becomes biased due to a relative abundance of cuspy models in the allowable hypervolume. *Right:* same as left, but for symmetrized anisotropy profile.

at fourth order, while we extract more information from the data than GRAVSPHERE1.5 by fitting the individual velocities with our generalized PDF modeling. For distribution function-based models, the challenge consists in creating sufficiently generalized functional forms so as to eliminate any biases. On the other hand, it is apparent that models which assume Gaussian velocity PDFs will be prone to biased recovery if the velocity distributions of the system, as in the case of Fig. 3, are intrinsically non-Gaussian.

In Fig. 4, we show the recovery of the logarithmic density slope for the *Gaia* Challenge mocks. In all cases, GRAVSPHERE2 is able to recover these within its 95% confidence intervals within $2R_{1/2}$, and in most cases everywhere in the plot and beyond ($0.1 \lesssim r/R_{1/2} \lesssim 10$). For the 1000-tracer case, only GRAVSPHERE2 recovers the true solution everywhere within the plots in all mocks. GRAVSPHERE1.5 systematically overestimates the steepness of the slope below $R_{1/2}$ for the LOS core model. AGAMA, on the other hand, is able to recover the slope profile well for all cases except for the core model with PM data beyond $\gtrsim R_{1/2}$, where moderate bias is observed (and also with higher uncertainties toward the center at $r \lesssim R_{1/2}$). The only case where a mild degree of bias is observed for GRAVSPHERE2 is for the LOS-only cusp model beyond $2R_{1/2}$, where the true slope profile narrowly falls outside the 95% CL region. In this case, we verified that this was due to the CORENFWTIDES model slightly overestimating slopes in these regions. If we adopt instead an $\alpha\beta\gamma$ model (Eq. (33)), which was used to generate the mocks (and also adopted for the AGAMA results), the profile was recovered at all radii without any bias. This test is shown in Appendix A²².

To better understand the reason for the biases observed with other methods, we also show in Fig. 3 (inset) the results of taking the maxima and minima of models falling within the 68th

percentile from the maximum likelihood taken from the posterior distribution. These results were achieved as follows: (i) we compute the posterior of the likelihoods; (ii) determine the limiting value corresponding to the 32nd percentile, so that 68% of models are equal or above this likelihood (the 100th percentile corresponding to the maximum likelihood); (iii) for each radius, we calculate the minima and maxima of all models with likelihoods greater or equal than our threshold value. This approach differs from the classical Bayesian posterior and we have chosen it as a measure to define the space of models which can be said to demonstrate an “equally good fit”. It is more reminiscent of frequentist statistics and we intend to explore it in more detail in upcoming works as a means of testing prior and posterior hypervolume shape dependence in models.

For the Gaussian model, this approach significantly increases the spread and encompasses the true solution, while for the default fourth-order GRAVSPHERE2 model, the effect is more moderate. This indicates that the Gaussian model, rather than unable to fit the correct solution, is not able to discriminate between the true model and those that exhibit substantial deviations from it, while the GRAVSPHERE2 model is. Since the posterior hypervolume of the Gaussian model contains large regions of model configurations with comparable likelihood that are skewed away from the true model, this biases the posterior distribution. This suggests that a similar effect may be taking place in other cases, such as with GRAVSPHERE1.5.

The latter result leads us to propose a diagnostic that may be important to better understand differences between quoted results for, for instance, the slopes and density profiles of dwarf galaxies, an active topic of debate and discussion in the literature (e.g., Read et al. 2018; Hayashi et al. 2020; Vitral et al. 2024; Arroyo-Polonio et al. 2025). If bounds from the posterior confidence intervals show significantly less spread (i.e., appear more aggressive) than the ones resulting from the extremal values of models of comparable likelihoods (with our reference test being the 68th percentile from the maximum), this indicates that the results are dominated by the shape of the posterior hypervolume, which is sensitive to priors and model parameterizations.

²² Note: it is not very surprising that biases will be naturally reduced when assuming the same mass model as the one used to generate the mock data. For real data, of course, we cannot know what the true underlying model is, so it is prudent to explore a couple of different options where possible.

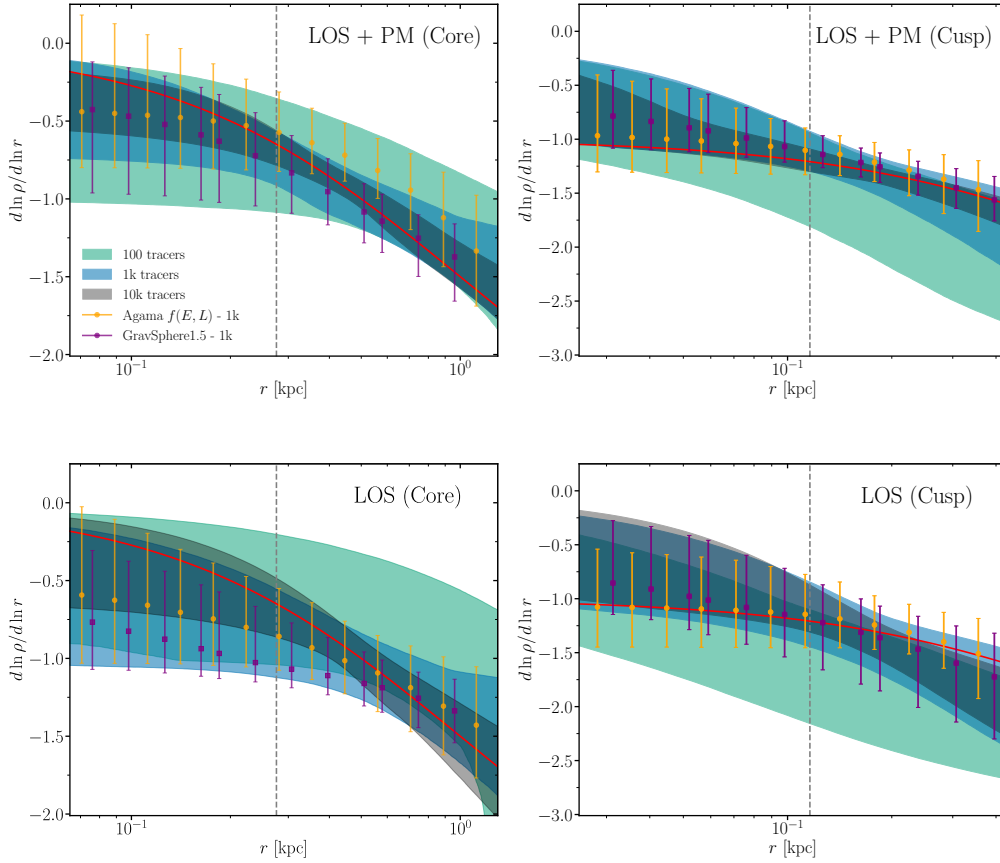


Fig. 4. GRAVSPHERE2 recovery of the logarithmic density slope profile for the GAIA CHALLENGE mock galaxies. Upper panels include LOS and PM tracers, whilst lower panels include only LOS ones. The left panels correspond to the Plum-CoreOM case, whilst the right ones to the PlumCuspOM case. The bands, errors and results presented follow the same notation as Fig. 1.

Consequently, one should on these occasions proceed with caution when interpreting the results, which might (without careful consideration) overstate the capacity of the analysis to truly distinguish among competing hypotheses. Conversely, if one does not find such a discrepancy, then it is indicative of the fact that the favored models are indeed better-fitting than the disfavored ones and not merely more abundant in the posterior hypervolume.

As noted in Sect. 3.3, however, the priors we adopt in our models are generally very broad compared to the results favored by the data, which we show explicitly in Appendix B. This shows that the posterior hypervolume dependence that we observe with our diagnostic is not limited to cases where the posterior parameter space spans the entire prior volume, but more generally to degenerate regions allowed by the data within sub-volumes of the prior space, especially when the true solution lies at the edge of the allowable hypervolume. This can be conditioned by a biased likelihood, which is partly at play in our example assuming Gaussianity, and also by the prior itself (noting the hypervolume region is the intersection of the models with allowable likelihoods that lie in the prior space). Also, even in cases where the model correctly recovers the true solution, meaning that the hypervolume is not skewed, a large spread when comparing with our diagnostic can indicate that the confidence of the model is overstated by posterior statistics (i.e., that the posterior hypervolume happens to have more models clustering around the true solution). In this way, our diagnostic based purely on the likelihood does not attempt to compute posterior statistics based on the relative abundance of models, as with typical Bayesian approaches, but merely computes the spread on a given quantity

allowable within a likelihood threshold, and as a result will not be conditioned by these potential biases.

As noted in Sect. 3.3, we also included marginalized plots of parameters in a public repository for all mocks considered in this study. We note that individual parameters are well constrained in most cases with respect to the prior limits, and even in cases where this is not the case, physical quantities of interest, such as the mass density in regions of interest, are insensitive to these degeneracies (cf. Appendix B).

Lastly, we included binned dispersion and kurtosis profiles for reference in Appendix D. The binned profiles are generally consistent with those derived from our GRAVSPHERE2 mass models and show meaningful deviations from Gaussianity. This, once again, stresses the importance of using models which are able to capture these deviations.

4.2. Simulated dwarf galaxies

Fig. 5 shows the GRAVSPHERE2 recovery of the mass density and anisotropy profiles for the Fornax-like simulated dwarf galaxy with and without PMs in addition to the LOS velocities, whilst Fig. 6 shows the recovery of the logarithmic density slope profiles for these models. The recovery of the density, anisotropy and slope using all available bound tracers with LOS velocities is excellent. Using only a small subset of these still yields very good recovery that is further improved by including PMs, leading to a comparable result to the GAIA CHALLENGE equilibrium mocks from Sect. 4.1, and with true values falling everywhere ($0.1 < r/R_{1/2} < 10$) within the 95% CL regions. This, on the whole, suggests that these simulated

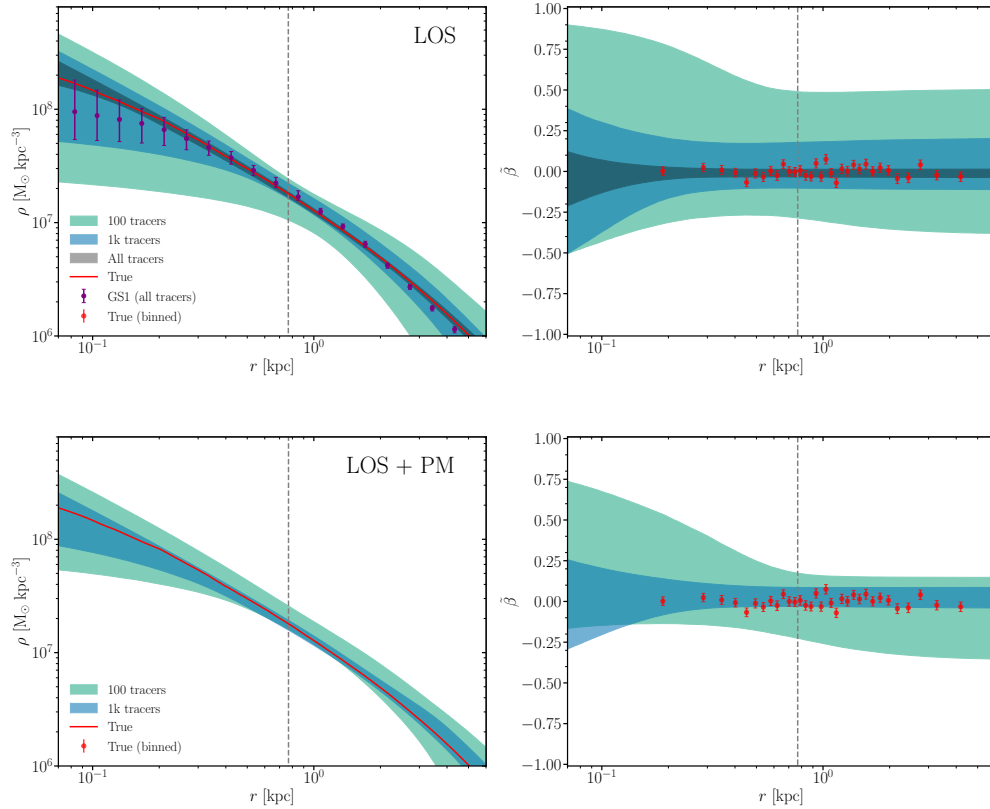


Fig. 5. *Top-left:* GRAVSPHERE2 recovery of the 3D dark matter density profile for the Fornax-like simulated galaxy, where the bands denote the 95% CL regions for different numbers of stellar tracers, as marked, including only LOS velocities. The total sample of bound tracers is 34 158 (gray band). The purple points correspond to the median values with 95% CL errors from Tchiorniy & Genina (2025) using all LOS tracers with GRAVSPHERE1. The red, solid line corresponds to the true density profile from the simulation (smoothened with a Savitzky-Golay filter). The gray, dashed line corresponds to the projected half-light radius ($R_{1/2}$). *Top-right:* symmetrized anisotropy profile results. The red points denote the binned 3D anisotropies obtained from bins of equal numbers of tracers and directly fitting the generalized velocity PDF from Eqs. (25) and (30) along the three spherical-coordinate directions. *Lower left:* same as Upper Left but including both LOS and PM velocities. *Lower right:* same as Upper Right but with both LOS and PM velocities.

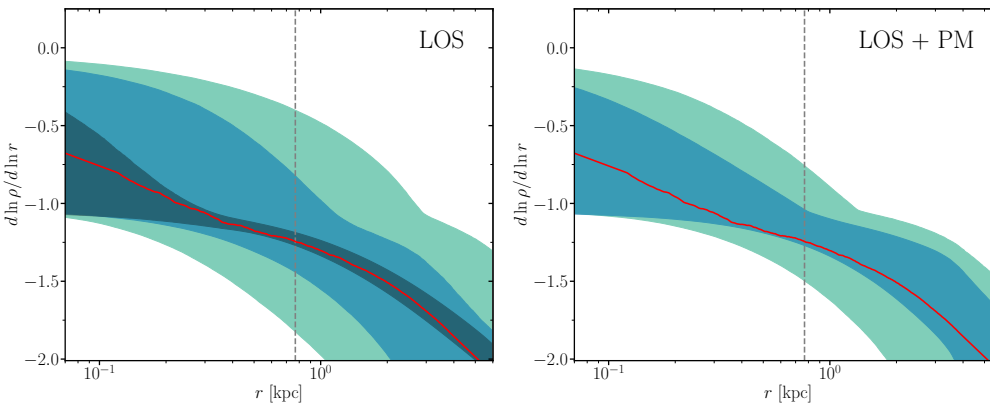


Fig. 6. *Left:* GRAVSPHERE2 recovery of the logarithmic density slope profile for the simulated Fornax-like dwarf. The bands denote the 95% CL regions for different numbers of stellar tracers (100, 1000, and 34 518, same as in Fig. 5) including only LOS velocities, as marked. The red, solid line corresponds to the slope from the smoothed density profile from the simulation. *Right:* same as the left panel but including both LOS and PM velocities.

dwarf galaxies do not show significant departures from dynamical equilibrium, spherical symmetry, or the assumption of no rotation. Comparing our results for the all-tracer LOS model with GRAVSPHERE1 from Tchiorniy & Genina (2025) (purple points in Fig. 5, upper left), GRAVSPHERE2 is significantly less biased than GRAVSPHERE1 for the density recovery, whilst also showing substantially smaller uncertainties overall (particularly toward the central region). Whilst Tchiorniy & Genina (2025) show that their biases can be substantially reduced by removing VSPs from their fits, this is done at the expense of higher uncertainties. They also obtained less biased results when weakening the monotonicity prior constraint on their nonparametric mass model, though some bias still remained (see their Fig. D1, also Fig. 8, where their biases are most apparent).

Lastly, we included the projected binned velocity dispersion and kurtosis profiles in Appendix E. These are consistent with the ones inferred from our models using individual velocities

and are close to isotropic, not showing large deviations from Gaussianity. The isotropy of the distribution is most evident from the right hand side of Fig. 5. This is less evident in Tchiorniy & Genina (2025), but our binning procedure effectively removes the stochastic noise by averaging over large numbers of tracers, revealing a profile which is close to isotropic.

4.3. Low tracer numbers and comparisons with mass estimators

Some of the most interesting objects to study at any given time are either extremely intrinsically faint or distant, meaning that they are (and perhaps always will be) scarce in tracer numbers²³.

²³ This should not be interpreted as prognosticating pessimism for future observations, but rather that with increasing observations it appears likely that interest will shift toward ever fainter systems that lie at the limit of the capabilities at any given time.

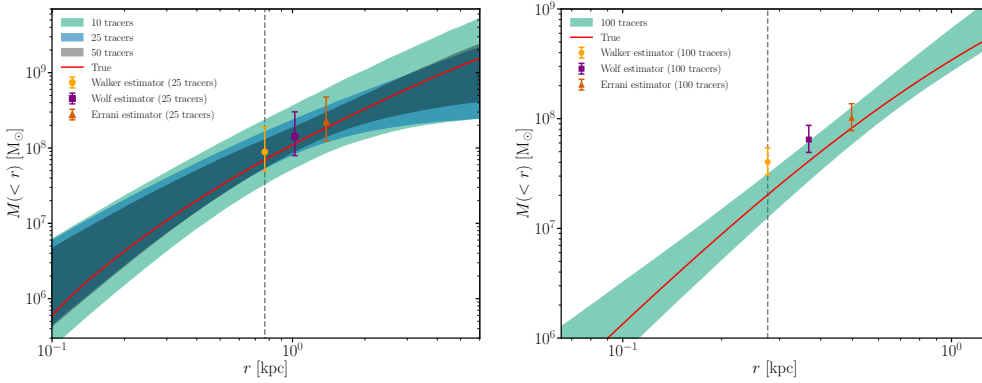


Fig. 7. *Left:* GRAVSPHERE2 recovery of the mass profile for the Fornax simulated galaxy with only LOS velocities for low stellar tracer numbers (10, 25, 50). The bands display the 95% CL values. We show the results obtained using the mass estimators from Walker et al. (2009); Wolf et al. (2010); Errani et al. (2018). *Right:* same, but for the Plummer CoreOM mock with 100 tracers and Gaussian PDF binning.

Some current examples include the discovery of the faintest known Milky Way satellite (Smith et al. 2024b) or the dwarf galaxies in Andromeda (Collins et al. 2021; Pickett et al. 2025). In these cases, one is limited to, at most, a few tens of kinematic tracers. It is therefore of considerable interest to study the behavior of GRAVSPHERE2 in this low tracer limit, comparing it with other commonly-applied methods such as the mass estimators from Walker et al. (2009); Wolf et al. (2010); Campbell et al. (2017); Errani et al. (2018). We have done this for the case of the Fornax simulated dwarf, showing our results in Fig. 7 (left)²⁴. For this case, we computed the global velocity dispersion by fitting GRAVSPHERE2’s non-Gaussian PDFs from Sect. 2.2 (i.e., creating one bin for all the tracers), although Gaussian PDF instead did not change our results (in this case). The errors on the estimator were then propagated from the posterior distribution of the global dispersion directly, assuming no other sources of error (systematic or otherwise). In this case, all estimators show good agreement with GRAVSPHERE2, albeit with somewhat larger errors, with all recovering the true solution within the specified 95% CL regions.

Fig. 7 (right), on the other hand, shows a different case, where all estimators show a degree of bias that is not present in GRAVSPHERE2. A major source of bias in this case stems from binning velocities with a Gaussian PDF, as commonly done in the literature. In Appendix C we show how using non-Gaussian PDFs instead substantially improves these biases. In this case, all estimators still show a degree of bias, but for the Errani estimator this only becomes apparent in the mock with 10 000 tracers.

Therefore, while the nominal errors of mass estimators are comparable to those of GRAVSPHERE2, the simple estimators suffer from potentially significant systematic biases that are not present in GRAVSPHERE2. We conclude that even for very low numbers of tracers (on the order of 10) it is still worth using GRAVSPHERE2 rather than these estimators, which has the additional advantage of giving constraints on the mass profile at all radii, rather than just one point at $\sim R_{1/2}$. Alternatively, particularly for more robust choices such as the Errani estimator, biases can be substantially reduced when adopting non-Gaussian PDFs, but these can still creep in as the number of tracers is increased.

We note that in our main analysis, we assumed a good knowledge of the photometric tracer profile, using our default value of 10 000 tracers for most tests so far. While it is often the case that the photometric light profile is far better constrained than the kinematic one, one may ask what the effects of restricting ourselves to a low number of photometric tracers might

be on the mass estimates. Interestingly, repeating our analysis using the kinematic tracers as the only source of photometric information yielded essentially identical results. This indicates that the advantage of GRAVSPHERE2 over the mass estimators does not stem from additional information provided by a good knowledge of the tracer profile and that the uncertainty in the kinematic information dominates the uncertainty in our mass estimates.

5. Discussion and conclusion

In this work, we introduced and tested GRAVSPHERE2, an extended Jeans modeling tool that includes full treatment of the higher order Jeans equations of fourth-order moments. We introduced fourth-order PMs as a new observable to break model degeneracies, using a general radially dependent fourth-order anisotropy term as part of our analysis. This approach builds on previous works, which have only considered higher moments for LO velocities and/or made restrictive assumptions about the form of the higher order counterpart of the anisotropy.

In contrast to previous versions of GRAVSPHERE, our methodology does not rely on VSPs which can become biased (see the discussion at the end of Sect. 2.1). It is also completely “bin-free”, with a generalized velocity PDF that we used to fit discrete data star-by-star. GRAVSPHERE2 includes an improved tracer density model that has the option of fitting tracer positions individually, along with an improved sampling technique. GRAVSPHERE2 does not rely on assumptions about the form of the phase-space distribution function, as it implicitly constrains it through its velocity moments via the second- and fourth-order Jeans equations. In this way, our method is able to retain the flexibility and bias-free nature of usual Jeans analyses by not needing to specify any particular form for the distribution function, while gaining constraining power from higher order velocity information, avoiding negative moments that are unphysical (at fourth order), and offering all the other general advantages of having a distribution function.

We tested these improvements using equilibrium mock data from the GAIA CHALLENGE suite, comparing our results to those from Read et al. (2021) and Collins et al. (2021), where multiple methods were tested, including GRAVSPHERE1.5, the (until now) latest version of GRAVSPHERE, AGAMA, a popular distribution function-based method, and DISCRETEJAM, an axisymmetric Jeans-based method. We also tested our method on simulated dwarf galaxy data recently presented in Tchiorniy & Genina (2025), which were also used to test PYGRAVSPHERE (referred to in the present paper as GRAVSPHERE1).

²⁴ We have omitted showing the mass estimator result from Campbell et al. (2017) as it produced a very similar result that overlapped with the one from Errani et al. (2018).

We also tested GRAVSPHERE2 on mocks with a small number of tracers (10–100). We compared its performance with simple mass estimators, finding that the GRAVSPHERE2 recovery is not only good, but generally superior. In particular, mass estimators can become significantly biased, while GRAVSPHERE2 not only remains unbiased, but provides constraints on the mass profile at all radii, rather than just one point at $\sim R_{1/2}$ with equal or lower errors. This suggests that even with just ten tracer velocities, it is still preferable to perform full GRAVSPHERE2 modeling (or at least include more flexible, non-Gaussian PDFs such as the ones we use here) over simpler, more commonly used approaches.

While previous studies have suggested that the effectiveness of higher order Jeans analyses is compromised without assumptions about the fourth-order anisotropy counterpart (e.g., Richardson & Fairbairn 2013, 2014), we find that, not only does a general fourth-order analysis yield comparable (if not improved) constraining power to previous methods for the same data, but, in many cases, a substantially less biased recovery of the anisotropy, mass density, and slope profiles. These are all recovered at 95% confidence across almost all mocks over a wide range ($0.1 \lesssim r/R_{1/2} \lesssim 10$), regardless of tracer numbers. This stands in contrast to previous methods, including GRAVSPHERE1.5 and PYGRAVSPHERE, which are prone to notable biases, particularly at small ($R \lesssim 0.25R_{1/2}$) and large ($R \gtrsim 2R_{1/2}$) radii.

Using 1000 tracers without PMs, we recovered the logarithmic density slope at $R_{1/2}$, with 12% (25%) statistical errors for cuspy (cored) mock data and with comparable absolute errors at lower radii. When including PMs, these can be distinguished with 8% (12%) errors. With only 100 tracers and no PMs, we were still able to recover slopes with $\sim 30\%$ (20%) errors. This indicates that we are able to meaningfully distinguish between cusped and cored models even with limited data, which is particularly timely in light of the present interest for such searches in the literature.

In summary, while PMs lead to substantially improved constraints on our mock data, LOS velocities alone are able to break the mass-anisotropy degeneracy and significantly reduce biases using our framework. Thus, we have established higher order Jeans analysis as a viable alternative to other methods, such as VSPs, which show significantly greater bias across our mock data, irrespective of tracer numbers. Improvements to our method over VSPs likely stem from some combination of biases in the VSPs themselves (see Sect. 2.1) and a full star-by-star, fourth-order treatment leading to stronger constraints that disfavor biased models (cf. Fig. 3).

Finally, we contrasted GRAVSPHERE2 with classic second-order Jeans modeling assuming Gaussianity. We demonstrated that due to the mass-anisotropy degeneracy, classic Jeans models are unable to distinguish cusps from cores even with 10 000 tracers. Indeed, as reported previously in the literature (e.g., Read et al. 2021), classic Jeans can become seemingly highly biased in its recovery of the density profile, appearing to strongly favor a cusp when the mock is actually cored. We show that this bias owes to the presence of many more cuspy models than cored models in the solution hypervolume, rather than being due to cored models giving a poorer fit to the data. As such, apparent biases on this level can be misleading, owing to the shape of the prior volume and unphysical model degeneracies (which GRAVSPHERE2 is able to resolve), rather than the data. We suggest a new diagnostic test to determine when posterior intervals are biased in this way: plotting the spread of models of comparable likelihood, using the 68th percentile from the maximum as a

reference threshold. We showed that this diagnostic test removes the apparent bias, giving a more faithful representation of the span of models that well-represent the data.

Since GRAVSPHERE2 has a distribution function, it can self-consistently model uncertain membership probabilities of individual tracers, robustly including a foreground model, which we will include in an upcoming publication Bañares-Hernández et al. (in prep.). Further additions to GRAVSPHERE2 are in preparation and will be presented in upcoming papers, including a mixture model for binary stars (following Gratton et al. 2025) and a validation of GRAVSPHERE2 on rotating mocks. Additionally, we have already incorporated in the public version of GRAVSPHERE2 the ability to include self-consistent escape velocities of stars as an additional constraint to the potential (cf. Häberle et al. 2024).

Given the present challenges concerning the structure of nearby dwarf galaxies and star clusters, GRAVSPHERE2 will be an excellent new tool for studying these systems and making the most of their ever-growing data sets. This will help us to address key outstanding questions in the field, such as whether nearby dSphs have dark matter cusps or cores (e.g., Oman et al. 2015; Bullock & Boylan-Kolchin 2017; Read et al. 2019; Kaplinghat et al. 2020; Santos-Santos et al. 2020; De Leo et al. 2024) and whether nearby GCs host central IMBHs (e.g., Mezcua 2017; Greene et al. 2020; Neumayer et al. 2020; Vitral et al. 2023; Askar et al. 2023; Häberle et al. 2024; Bañares-Hernández et al. 2025). We will apply GRAVSPHERE2 to data for nearby dSphs and GCs in forthcoming publications to address these questions.

Acknowledgements. We thank the anonymous referee for prompt and helpful comments that have helped improving this work. We are grateful to Jason Sanders for helpful comments on the use of velocity distribution functions used for this work and on the manuscript. We are grateful to Anna Genina for facilitating us access to simulated data and PYGRAVSPHERE mock data results used in this work. We would like to thank Eugene Vasiliev, Jorge Martín Camalich, Giuseppina Battaglia, José María Arroyo Polonio, Raffaele Pascale, Eduardo Vitral, and Amery Gratton for helpful discussions and their interest in our work. We also thank the organizers of the “Valencia Workshop on the Small-Scale Structure of the Universe and Self-Interacting Dark Matter”, where this work was presented. The authors wish to acknowledge the contribution of the IAC High-Performance Computing support team and hardware facilities to the results of this research. A.B.H. acknowledges funding received from the European Union through the grant “UNDARK” of the Widening participation and spreading excellence program (project number 101159929). A.B.H. acknowledges support from the MICINN through the grant “DarkMaps” PID2022-142142NB-I00. J.I.R. would like to acknowledge support from STFC grants ST/Y002865/1 and ST/Y002857/1. This work has made use of the following software packages: DYNASTY (Speagle 2020), CORNER.PY (Foreman-Mackey 2016), NUMPY (Harris et al. 2020), SCIPY (Virtanen et al. 2020), MATPLOTLIB (Hunter 2007), JUPYTER NOTEBOOK (Kluyver et al. 2016), GH_ALTERNATIVE (Sanders & Evans 2020).

References

- Alvarez, A., Calore, F., Genina, A., et al. 2020, *JCAP*, 2020, 004
- Alvey, J., Sabti, N., Tiki, V., et al. 2021, *MNRAS*, 501, 1188
- An, J. H., & Evans, N. W. 2006, *ApJ*, 642, 752
- Anderson, J., & Van der Marel, R. P. 2010, *ApJ*, 710, 1032
- Arroyo-Polonio, J. M., Pascale, R., Battaglia, G., et al. 2025, *A&A*, 699, A347
- Askar, A., Baldassare, V. F., & Mezcua, M. 2023, arXiv e-prints [arXiv:2311.12118]
- Bañares-Hernández, A., Castillo, A., Martín Camalich, J., & Iorio, G. 2023, *A&A*, 676, A63
- Bañares-Hernández, A., Calore, F., Martín Camalich, J., & Read, J. I. 2025, *A&A*, 693, A104
- Baes, M., & van Hese, E. 2007, *A&A*, 471, 419
- Battaglia, G., & Nipoti, C. 2022, *Nat. Astron.*, 6, 659
- Battaglia, G., Helmi, A., & Breddels, M. 2013, *New Astron. Rev.*, 57, 52
- Baumgardt, H., Faller, J., Meinhold, N., McGovern-Greco, C., & Hilker, M. 2022, *MNRAS*, 510, 3531

- Binney, J., & Tremaine, S. 2008, *Galactic Dynamics* (Berlin: Springer), 2nd edn.
- Boehm, C., Krone-Martins, A., Amorim, A., et al. 2017, arXiv e-prints [arXiv:1707.01348]
- Bosma, A. 1978, PhD thesis, University of Groningen, The Netherlands
- Breddels, M. A., Helmi, A., van den Bosch, R. C. E., van de Ven, G., & Battaglia, G. 2013, *MNRAS*, **433**, 3173
- Bullock, J. S., & Boylan-Kolchin, M. 2017, *ARA&A*, **55**, 343
- Campbell, D. J. R., Frenk, C. S., Jenkins, A., et al. 2017, *MNRAS*, **469**, 2335
- Cappellari, M. 2008, *MNRAS*, **390**, 71
- Cappellari, M. 2015, arXiv e-prints [arXiv:1504.05533]
- Chanamé, J., Kleyna, J., & van der Marel, R. 2008, *ApJ*, **682**, 841
- Collins, M. L. M., Read, J. I., Ibata, R. A., et al. 2021, *MNRAS*, **505**, 5686
- De Leo, M., Read, J. I., Noël, N. E. D., et al. 2024, *MNRAS*, **535**, 1015
- de Lorenzi, F., Debattista, V. P., Gerhard, O., & Sambhus, N. 2007, *MNRAS*, **376**, 71
- de Lorenzi, F., Gerhard, O., Coccato, L., et al. 2009, *MNRAS*, **395**, 76
- Dehnen, W. 2009, *MNRAS*, **395**, 1079
- Dejonghe, H. 1987, *MNRAS*, **224**, 13
- Dejonghe, H., & Merritt, D. 1992, *ApJ*, **391**, 531
- Dickson, N., Hénault-Brunet, V., Baumgardt, H., Gieles, M., & Smith, P. J. 2023, *MNRAS*, **522**, 5320
- Dickson, N., Smith, P. J., Hénault-Brunet, V., Gieles, M., & Baumgardt, H. 2024, *MNRAS*, **529**, 331
- Errani, R., Peñarrubia, J., & Walker, M. G. 2018, *MNRAS*, **481**, 5073
- Evans, A. J., Strigari, L. E., & Zivick, P. 2022, *MNRAS*, **511**, 4251
- Evislin, J., & Del Popolo, A. 2017, *ApJ*, **841**, 90
- Expósito-Márquez, J., Brook, C. B., Huertas-Company, M., et al. 2023, *MNRAS*, **519**, 4384
- Feroz, F., Hobson, M. P., & Bridges, M. 2009, *MNRAS*, **398**, 1601
- Foreman-Mackey, D. 2016, *J. Open Source Softw.*, **1**, 24
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, **125**, 306
- Fritz, T. K., Chatzopoulos, S., Gerhard, O., et al. 2016, *ApJ*, **821**, 44
- Genina, A., Read, J. I., Frenk, C. S., et al. 2020, *MNRAS*, **498**, 144
- Genina, A., Read, J. I., Fattahi, A., & Frenk, C. S. 2022, *MNRAS*, **510**, 2186
- Gieles, M., & Zocchi, A. 2015, *MNRAS*, **454**, 576
- Gratton, A., Hendriks, D. D., Das, P., Heber, D., & Izzard, R. G. 2025, arXiv e-prints [arXiv:2509.14316]
- Greene, J. E., Strader, J., & Ho, L. C. 2020, *ARA&A*, **58**, 257
- Gregory, A. L., Collins, M. L. M., Read, J. I., et al. 2019, *MNRAS*, **485**, 2010
- Häberle, M., Neumayer, N., Seth, A., et al. 2024, *Nature*, **631**, 285
- Häberle, M., Neumayer, N., Clontz, C., et al. 2025, *ApJ*, **983**, 95
- Handley, W. J., Hobson, M. P., & Lasenby, A. N. 2015a, *MNRAS*, **450**, L61
- Handley, W. J., Hobson, M. P., & Lasenby, A. N. 2015b, *MNRAS*, **453**, 4384
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, **585**, 357
- Hayashi, K., Chiba, M., & Ishiyama, T. 2020, *ApJ*, **904**, 45
- Hernquist, L. 1990, *ApJ*, **356**, 359
- Higson, E., Handley, W., Hobson, M., & Lasenby, A. 2019, *SC*, **29**, 891
- Hunt, J. A. S., & Kawata, D. 2013, *MNRAS*, **430**, 1928
- Hunter, J. D. 2007, *Comp. Sci. Eng.*, **9**, 90
- Jeans, J. H. 1922, *MNRAS*, **82**, 122
- Júlio, M. P., Brinchmann, J., Zoutendijk, S. L., et al. 2023, *A&A*, **678**, A38
- Júlio, M. P., Read, J. I., Pawłowski, M. S., et al. 2025, *A&A*, **704**, A330
- Kaplinghat, M., Ren, T., & Yu, H.-B. 2020, *JCAP*, **2020**, 027
- Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, in *Jupyter Notebooks - a publishing format for reproducible computational workflows* (IOS Press) 87
- Koposov, S., Speagle, J., Barbary, K., et al. 2022, <https://doi.org/10.5281/zenodo.6609296>
- Laporte, C. F. P., Agnello, A., & Navarro, J. F. 2019, *MNRAS*, **484**, 245
- Li, P., Tian, Y., Júlio, M. P., et al. 2023, *A&A*, **677**, A24
- Li, Z., Han, J., Wang, W., et al. 2025, *MNRAS*, **538**, 1442
- Lim, S. H., Hayashi, K., Horigome, S., Matsumoto, S., & Nojiri, M. M. 2025, arXiv e-prints [arXiv:2505.00763]
- Łokas, E. L. 2002, *MNRAS*, **333**, 697
- Łokas, E. L., & Mamon, G. A. 2003, *MNRAS*, **343**, 401
- Łokas, E. L., Mamon, G. A., & Prada, F. 2005, *MNRAS*, **363**, 918
- Mamon, G. A., & Łokas, E. L. 2005, *MNRAS*, **363**, 705
- Mamon, G. A., Biviano, A., & Boué, G. 2013, *MNRAS*, **429**, 3079
- McNamara, B. J., Harrison, T. E., & Anderson, J. 2003, *ApJ*, **595**, 187
- Merrifield, M. R., & Kent, S. M. 1990, *AJ*, **99**, 1548
- Merritt, D. 1985, *MNRAS*, **214**, 25P
- Merritt, D. 1987, *ApJ*, **313**, 121
- Mezcua, M. 2017, *Int. J. Mod. Phys. D*, **26**, 1730021
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, *ApJ*, **462**, 563
- Neal, R. M. 2003, *Ann. Stat.*, **31**, 705
- Neumayer, N., Seth, A., & Böker, T. 2020, *A&A Rev.*, **28**, 4
- Nguyen, T., Read, J., Necib, L., et al. 2025, *MNRAS*, **541**, 2707
- Oman, K. A., Navarro, J. F., Fattahi, A., et al. 2015, *MNRAS*, **452**, 3650
- Osipkov, L. P. 1979, *Pisma v Astronomicheskii Zhurnal*, **5**, 77
- Pascale, R., Nipoti, C., Calura, F., & Della Croce, A. 2025, *A&A*, **700**, A77
- Pickett, C. S., Collins, M. L. M., Rich, R. M., et al. 2025, *MNRAS*, **540**, 1701
- Read, J. I., & Steger, P. 2017, *MNRAS*, **471**, 4541
- Read, J. I., Agertz, O., & Collins, M. L. M. 2016, *MNRAS*, **459**, 2573
- Read, J. I., Walker, M. G., & Steger, P. 2018, *MNRAS*, **481**, 860
- Read, J. I., Walker, M. G., & Steger, P. 2019, *MNRAS*, **484**, 1401
- Read, J. I., Mamon, G. A., Vasiliev, E., et al. 2021, *MNRAS*, **501**, 978
- Richardson, T., & Fairbairn, M. 2013, *MNRAS*, **432**, 3361
- Richardson, T., & Fairbairn, M. 2014, *MNRAS*, **441**, 1584
- Rubin, V. C., Ford, W. K., J., & Thonnard, N. 1980, *ApJ*, **238**, 471
- Sanders, J. L., & Evans, N. W. 2020, *MNRAS*, **499**, 5806
- Santos-Santos, I. M. E., Navarro, J. F., Robertson, A., et al. 2020, *MNRAS*, **495**, 58
- Sarrato-Alós, J., Brook, C., Di Cintio, A., et al. 2025, *A&A*, **703**, A140
- Schwarzschild, M. 1979, *ApJ*, **232**, 236
- Skilling, J. 2004, *AIP Conf. Ser.*, **735**, 395
- Skilling, J. 2006, *Bayesian Anal.*, **1**, 833
- Smith, P. J., Hénault-Brunet, V., Dickson, N., Gieles, M., & Baumgardt, H. 2024a, *ApJ*, **975**, 268
- Smith, S. E. T., Cerny, W., Hayes, C. R., et al. 2024b, *ApJ*, **961**, 92
- Speagle, J. S. 2020, *MNRAS*, **493**, 3132
- Syer, D., & Tremaine, S. 1996, *MNRAS*, **282**, 223
- Tchiorniy, K., & Genina, A. 2025, arXiv e-prints [arXiv:2504.18617]
- Trimble, V. 1987, *ARA&A*, **25**, 425
- 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. 1994, *MNRAS*, **270**, 271
- Van der Marel, R. P., & Anderson, J. 2010, *ApJ*, **710**, 1063
- Vasiliev, E. 2019, *MNRAS*, **482**, 1525
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nat. Methods*, **17**, 261
- Vitral, E., Kremer, K., Libralato, M., Mamon, G. A., & Bellini, A. 2022, *MNRAS*, **514**, 806
- Vitral, E., Libralato, M., Kremer, K., et al. 2023, *MNRAS*, **522**, 5740
- Vitral, E., van der Marel, R. P., Sohn, S. T., et al. 2024, *ApJ*, **970**, 1
- Vitral, E., van der Marel, R. P., Sohn, S. T., et al. 2025, arXiv e-prints [arXiv:2508.20711]
- Walker, M. G., & Peñarrubia, J. 2011, *ApJ*, **742**, 20
- Walker, M. G., Mateo, M., Olszewski, E. W., et al. 2009, *ApJ*, **704**, 1274
- Wardana, D., Chiba, M., & Hayashi, K. 2025, *ApJ*, **982**, 167
- Watkins, L. L., van de Ven, G., den Brok, M., & van den Bosch, R. C. E. 2013, *MNRAS*, **436**, 2598
- Wilkinson, M. I., Kleyna, J., Evans, N. W., & Gilmore, G. 2002, *MNRAS*, **330**, 778
- Wolf, J., Martinez, G. D., Bullock, J. S., et al. 2010, *MNRAS*, **406**, 1220
- Zhao, H. 1996, *MNRAS*, **278**, 488
- Zimmermann, T., Alvey, J., Marsh, D. J. E., Fairbairn, M., & Read, J. I. 2025, *Phys. Rev. Lett.*, **134**, 151001
- Zoutendijk, S. L., Júlio, M. P., Brinchmann, J., et al. 2021, arXiv e-prints [arXiv:2112.09374]

Appendix A: Gaia Challenge logarithmic density slope profile recovery with an $\alpha\beta\gamma$ mass model

In this appendix, we consider how well GRAVSPHERE2 can recover the PlumCuspOM mock assuming an $\alpha\beta\gamma$ functional form for the enclosed mass profile rather than the default CORENFWTIDES model. The $\alpha\beta\gamma$ profile was used to generate the mock, and so should yield less bias. The results are shown in Fig. A.1. We note that this does indeed yield less bias as compared to using the CORENFWTIDES model (cf. with Fig. 4 (lower right), where the recovery is very slightly biased outside its 95% CL region).

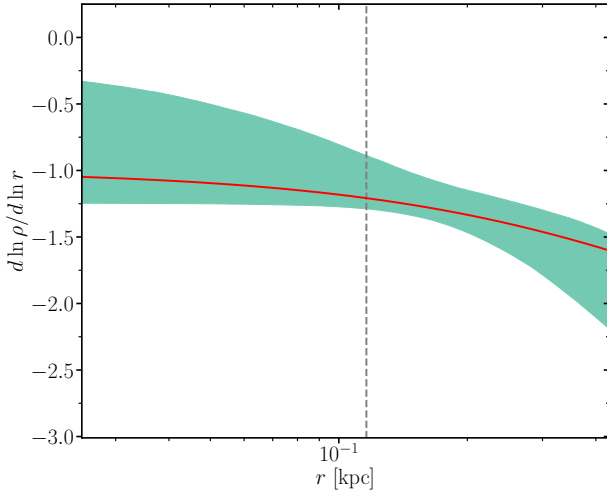


Fig. A.1. GRAVSPHERE2 recovery of the logarithmic density slope profile for the PlumCuspOM model with 10,000 LOS tracers, using an $\alpha\beta\gamma$ model (Eq. (33)) for the mass density profile instead of the default CORENFWTIDES model. The true profile (red) is recovered everywhere in the plot within the 95% CL region (green), yielding less bias than the CORENFWTIDES model, as expected since the mock was set up using an $\alpha\beta\gamma$ model.

Appendix B: Recovery comparison with priors

In Fig. B.1 we show as an example how, even for our least constraining model of the Gaia Challenge sample (just the LOS velocities of 100 tracers), the range of densities allowed by our priors (which is very similar across all the galaxies considered in our mock samples) spans several orders of magnitude and therefore has little effect on the result when using the data to constrain the mass model. We showed this by running a fit without any likelihood (just sampling the prior space) and plotting the range of densities spanned by the model in this case. Fig. B.2 shows the recovery of the tracer surface density for the same mock galaxy using a binned profile with 10,000 photometric tracers. It is clear that the recovery is excellent and far in excess of the allowable range by the prior. In the cases we considered, results will be essentially unchanged if one instead fixes the photometric profile to its best-fit value without additional marginalization. This can be used to accelerate run-times in GRAVSPHERE2.

Appendix C: Mass estimators with non-Gaussian PDFs and multiple tracers

Here, we explore the behavior of the simple mass estimators using GRAVSPHERE2's non-Gaussian PDFs (Sect. 2.2) for both

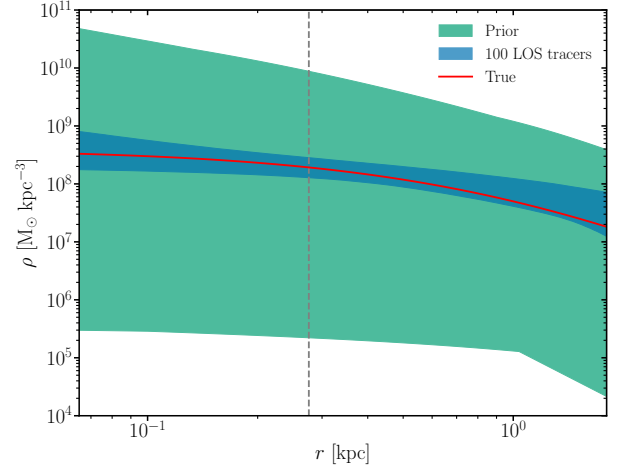


Fig. B.1. Comparison of GRAVSPHERE2 recovery (95% CL from the median) with just 100 LOS tracers and the range spanned by the priors on the mass model for the PlumCoreOM mock galaxy from the Gaia Challenge suite (same as in Fig. 3).

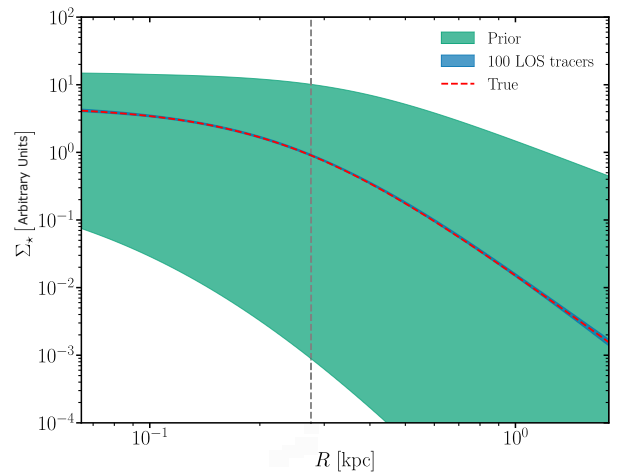


Fig. B.2. Comparison of GRAVSPHERE2 recovery (95% CL from the median) with just 100 LOS tracers (10,000 for the binned photometric component) and the range spanned by the priors on the tracer surface density model for the PlumCoreOM mock galaxy from the Gaia Challenge suite (same as in Fig. 3).

100 tracers and a much larger number of tracer stars (10,000). The results are shown in Fig. C.1, which compares the GRAVSPHERE2 recovery of the enclosed mass for the PlumCoreOM GAIA CHALLENGE mock with both 100 and 10,000 tracers with the same for simple mass estimators. Notice that while biases improve following these PDFs, they still persist for the Wolf and Walker estimators for 100 tracers and (to a varying extent) in all of them for the 10,000 tracer case, while GRAVSPHERE2 remains unbiased.

Appendix D: Gaia Challenge binned profiles

Figs. D.1 and D.2 show binned projected velocity dispersion and kurtosis profiles for the Gaia Challenge mock sample.

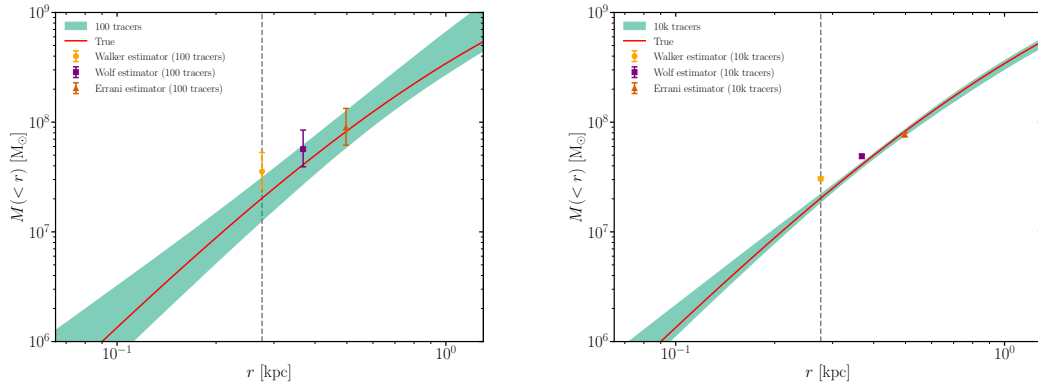


Fig. C.1. *Left:* GRAVSPHERE2 recovery, using GRAVSPHERE2’s non-Gaussian PDFs (Sect. 2.2), of the enclosed mass profile for the PlumCoreOM mock galaxy with only LOS velocities for 100 tracers, as compared to the same for simple mass estimators (as marked). The bands show the 95% CL values. *Right:* Same as left but for 10,000 tracers. While non-Gaussian PDFs improve results, a degree of bias is present in all estimators, though to a lesser extent in the Errani estimator, where the bias becomes negligible (i.e., relative to statistical errors) for 100 tracers.

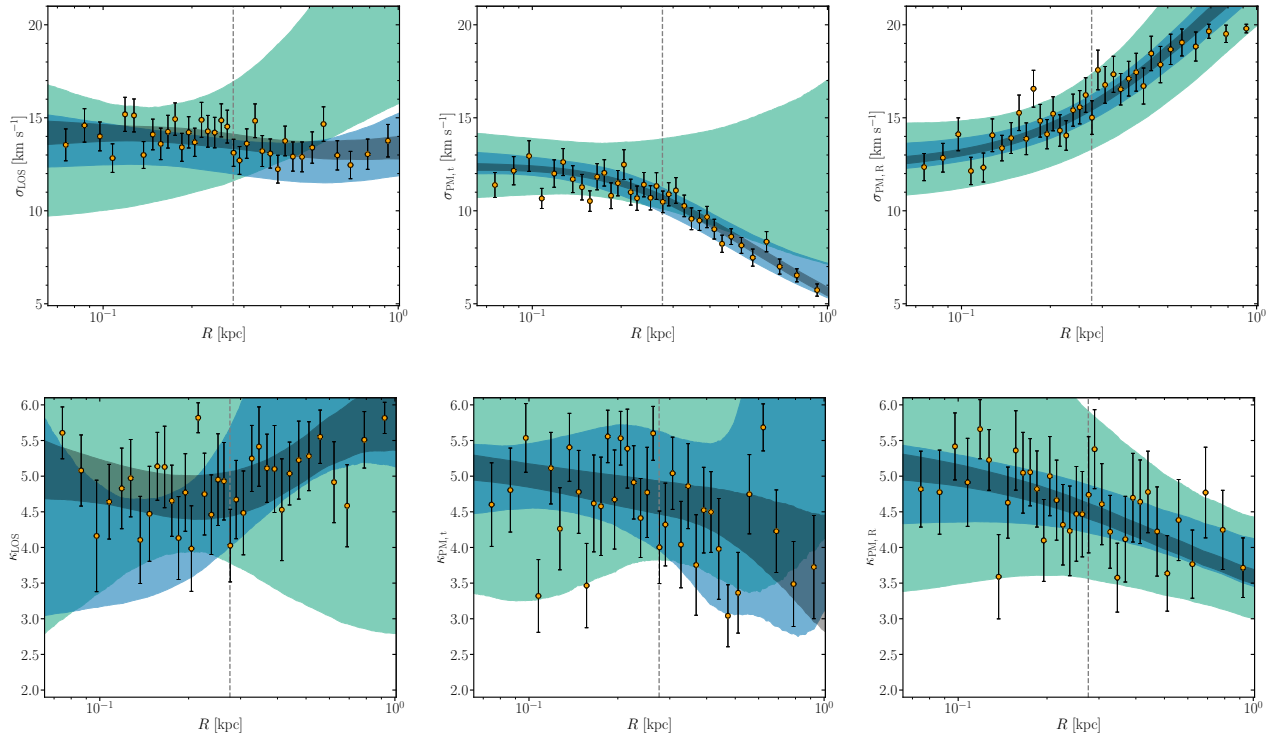


Fig. D.1. Projected binned velocity dispersion and kurtosis profiles for the PlumCoreOM GAIA CHALLENGE mock galaxy. LOS and PM velocity dispersion profiles are shown in the upper panels, whilst kurtosis profiles are shown in the bottom ones. The error bars denote 68 % CL regions from the median, whilst the colored bands denote the 95 % CL regions for 100 (green), 1,000 (blue), and 10,000 (gray) tracers. The LOS profile results are obtained from the LOS-only model, whilst the PM ones combine both data. Binned profiles are for the 10,000 tracer model in bins with an equal number of stars obtained by fitting the generalized PDFs (Eqs. (25) and (30)) for a given variance and kurtosis for each bin. The gray, dashed line denotes the projected half-light radius ($R_{1/2}$).

Appendix E: Simulated dwarf galaxy binned profiles

Figure E.1 shows the projected binned velocity dispersion and kurtosis profiles for the Fornax simulated galaxy.

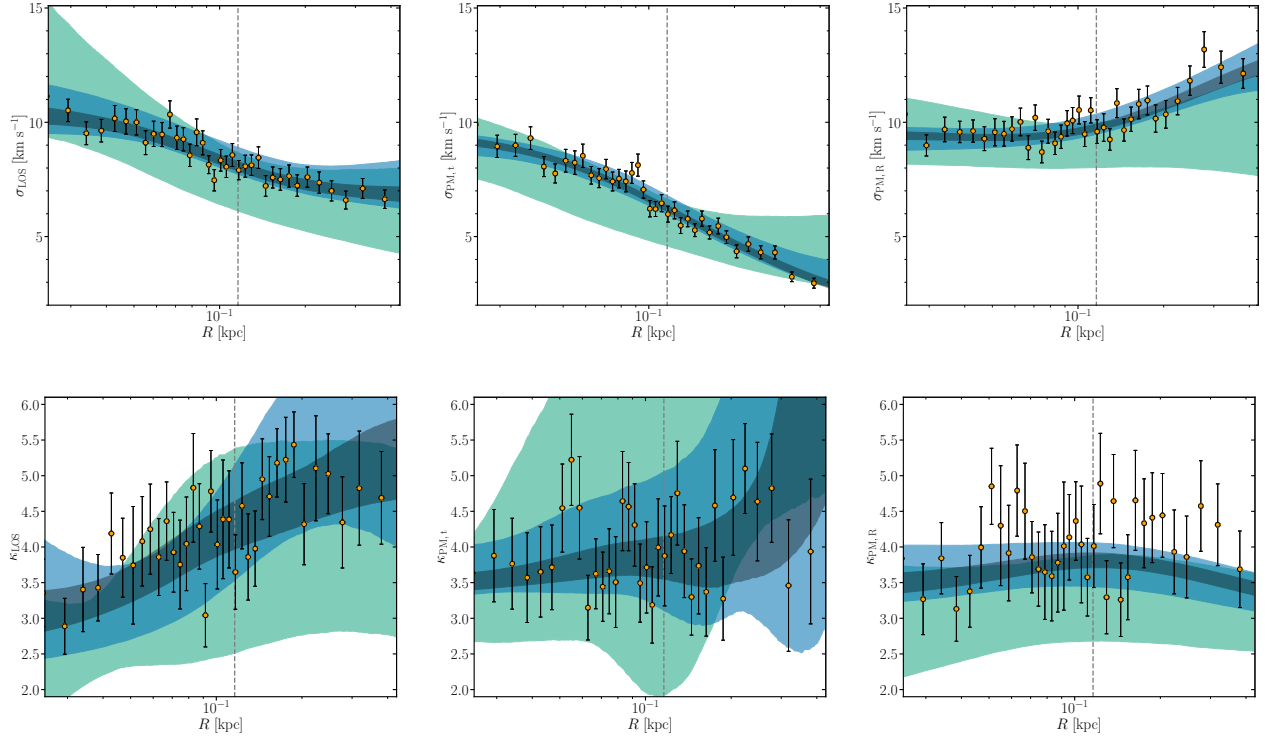


Fig. D.2. Same as Fig. D.1 but for the PlumCuspOM GAIA CHALLENGE mock galaxy.

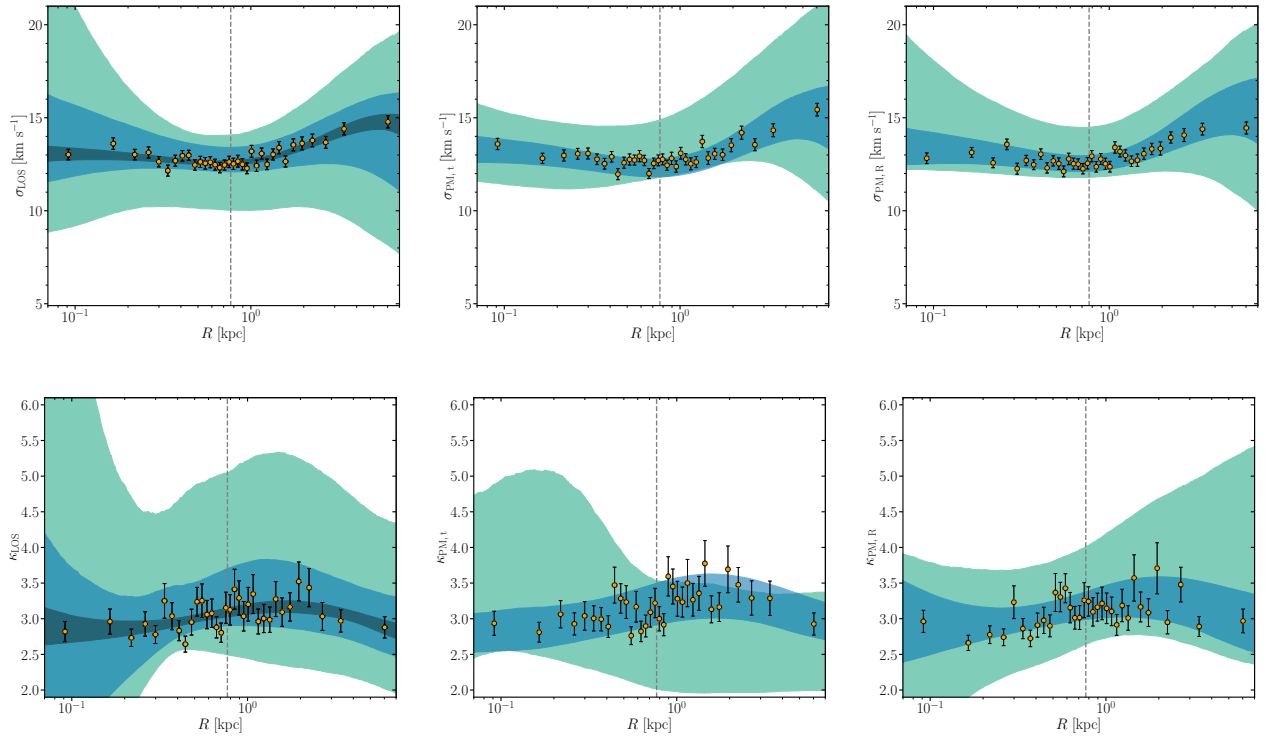


Fig. E.1. Projected binned projected velocity dispersion and kurtosis profiles for the Fornax-like simulated dwarf galaxy. The LOS and PM velocity dispersion profiles are shown in the upper panels, whilst kurtosis profiles are shown in the bottom ones. The error bars denote 68 % CL regions from the median, whilst the colored bands denote the 95 % CL regions for 100 (green), 1,000 (blue), and all tracers (gray). The LOS profile results are obtained from the LOS-only model, whilst the PM ones combine both data. Binned profiles are derived using all the bound and velocity-clipped tracers from the simulation in bins with of equal number fitting the generalized PDFs (Eqs. (25) and (30)) for a given variance and kurtosis for each bin. The gray, dashed line denotes the projected half-light radius ($R_{1/2}$).