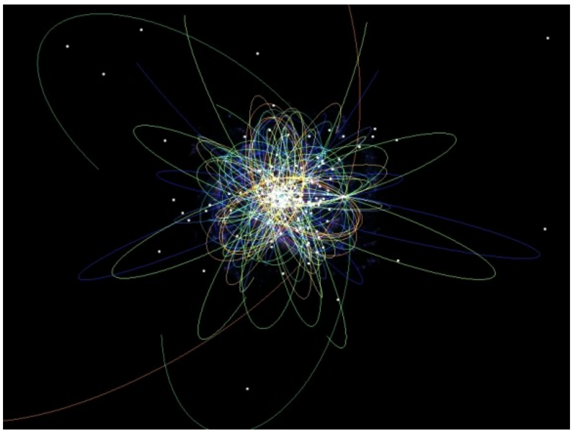
Axisymmetric orbit-based models of a mock dwarf spheroidal galaxy



Author: Jorrit Hagen
Under supervision of: Amina Helmi
Maarten Breddels

Affiliation: Kapteyn Astronomical Institute

October 6, 2015

Abstract

In the current cosmological ΛCDM model, there is six times more dark matter mass in the Universe than visible mass, but we still do not know what this dark matter is. The dwarf spheroidal satellite galaxies of our Milky Way, whose internal dynamics are believed to be dominated by dark matter, can therefore used to put constraints on its nature.

The goal of this work is to establish if it is possible to reliably measure the mass content, shape and internal orbital structure of a Sculptor-like Dwarf Spheroidal Galaxy, using the Schwarzschild's orbital superposition method. Most of the work thus far has assumed that dwarf spheroidal galaxies and their host halos are spherical, although we know from the light distribution that this is not true. This motivates us to use the Schwarzschild method in the axisymmetric regime to model these systems. The Schwarzschild method uses a complete set of orbits as building blocks for the system that is being modeled. The combination of orbits needed to match the observations, results in the distribution function of the system.

We set up a mock galaxy whose properties might resemble those of the Sculptor dwarf spheroidal galaxy. The mock galaxy, whose global potential results in a logarithmic profile, contains flattened luminous and dark matter components and has the advantage that it can be generated from a simple analytic distribution function. We then tested how well our Schwarschild method can constrain the global parameters of the potential and the stellar dynamical properties. We show that our method can reproduce the light distribution and the stellar kinematics of a specific mock galaxy and that we can recover the true characteristic parameters of its potential when pretending to observe 10^5 and 10^4 stars in a known edge-on view. For both cases we find, within 1σ-confidence interval, $q = 0.8^{+0.04}_{-0.04}$ for the flattening and $v_0 = 20^{+3}_{-3}$ km/s for the mass parameter.

Finally, we used the same mock galaxy and show that we can constrain the mass $\log_{10}(M_{1kpc})$, scale radius R_s and flattening c/a of the system by assuming an axisymmetric NFW potential form. Within 1σ-confidence intervals we find for its best-fit parameters that $\log_{10}(M_{1kpc}[M_\odot]) = 7.8 \pm 0.2$, $R_s = 2 \pm 1$ kpc and $c/a = 0.75^{+0.15}_{-0.05}$ when pretending to observe 10^5 stars in a known adge-on view and $\log_{10}(M_{1kpc}[M_\odot]) = 7.8 \pm 0.2$, $R_s = 2^{+2}_{-1}$ kpc and $c/a = 0.80^{+0.05}_{-0.10}$ for 10^4 stars.

Contents

Chapter 1

Introduction

It is more than 80 years ago that Zwicky (1933) inferred the existence of dark matter. The observed kinematics of (systems of) galaxies showed deviations with respect to the expected kinematics. The latter was based only on the contribution of the luminous and therefore visible matter of these systems. Around the same time, Einstein showed that gravitational lensing effects could exist (Einstein 1936; Renn et al. 1997), although it took until 1979 for the first one to be found (Walsh et al. 1979). Strong gravitational lensing effects that can not be explained by the amount of visible matter only, also supports the existence of dark matter. A similar reasoning holds for the observed HI rotation curves of spiral galaxies (Freeman 1970; Rubin et al. 1980).

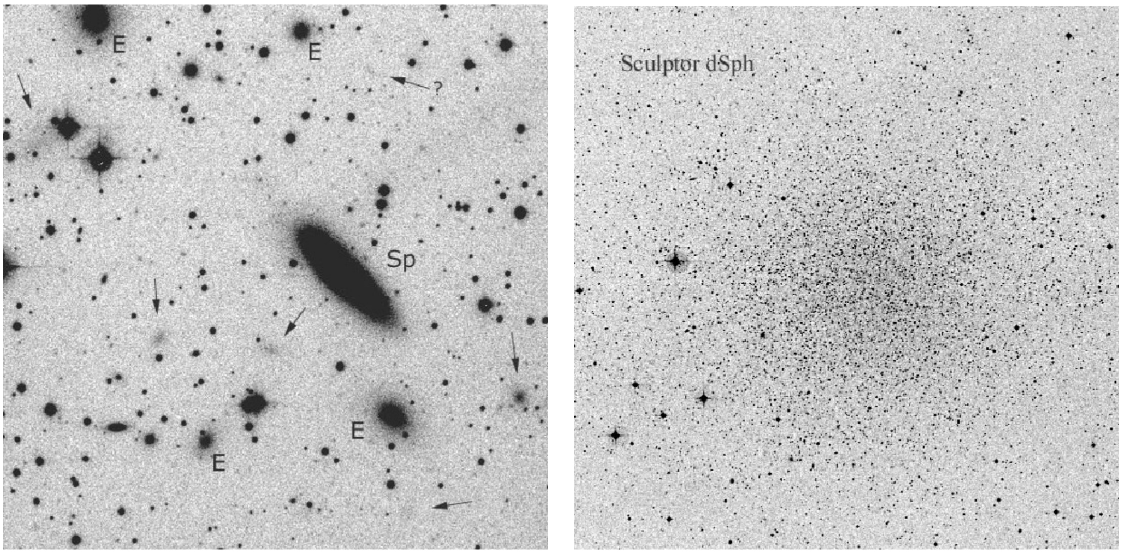
More recently, the discovery of the accelerating expansion of the Universe through observations of distant supernovae (Riess et al. 1998; Perlmutter et al. 1999) and accurate measurements of the Cosmic Microwave Background Radiation (CMB) by the WMAP satellite (Bennett et al. 2013) also indicated that another component is needed: dark energy.

Therefore, in the current cosmological ΛCDM model, using the new Planck results (Planck Collaboration et al. 2015), the energy content of the universe does not only consist of normal baryonic matter (5%) and radiation (0.008%), but also, to a much greater extent, of the mysterious components (cold) dark matter (26%) and dark energy (69%).

Problems arise since cold (non relativistic) dark matter models overpredict the number of small dwarf galaxies, the missing satellites problem (Klypin et al. 1999; Moore et al. 1999), and since the central density distributions of dark matter halos should be much more peaked than what is observed in galaxies by investigating their rotation curves, the cusp-core debate (Hui 2001). In addition Boylan-Kolchin et al. (2011) state that the predicted subhalos from ΛCDM simulations are too big to host the (small) Milky Way dwarf satellites, the too big to fail problem, although Vera-Ciro et al. (2013) show that this is dependent on the mass of the host halo. They find no too big to fail problem for a Milky Way mass around $8 \times 10^{11} M_\odot$.

Structures exist since the universe started in a hot and dense state in which tiny density fluctuations were already present. These fluctuations were for the first time found in the CMB radiation by the NASA Cosmic Background Explorer (COBE) in 1992. Warm dark matter (WDM) models are based on particles that initially had higher velocities than cold dark matter models. These particles became non-relativistic later. The good thing about this type of dark matter is that primordial density fluctuations on very small scales would get washed out because particles from over-dense regions would move towards under-dense regions. These models reduce the missing satellites problem and cusp-core debate and therefore seem to fit observations better. Peter (2012) has written down a nice short review about the possible dark matter candidates. We note that many of the ΛCDM problems could also have astrophysical solutions, since most predictions are based on dark matter only simulations, i.e. without baryonic physics included.

The dwarf spheroidal satellite galaxies (dSph's or dSph galaxies) of our Milky Way might be used to put constraints on the nature of dark matter, since these systems are believed to be highly dark matter dominated (Strigari et al. 2008; Walker et al. 2007; Wolf et al. 2010). Most work has assumed that dwarf spheroidal galaxies and their host halos are spherical, although we know from the light distribution that this is not true (Hayashi & Chiba 2015, and references therein). In fact, the natural shape of a dark matter halo is triaxial (Binney 1978) and this is confirmed in cosmological N-body simulations. In this work we test whether we can use the Schwarzschild's



(a) (b)

Figure 1.1: Dwarf spheroidal galaxies. Panel (a): Image of the nearby Centaurus galaxy cluster showing elliptical galaxies (E), a spiral galaxy (Sp) and several dwarf spheroidal galaxies denoted with arrows. Panel (b): The Sculptor dwarf spheroidal galaxy. The bright stars in the image are most likely foreground stars located in the Milky Way itself. Figure from Mateo (2000).

orbital superposition method to measure the mass content, shape and internal orbital structure of a system like the Sculptor Dwarf Spheroidal Galaxy. To this end we thus apply the Schwarzschild method in the axisymmetric regime to establish whether we are able to determine the properties of this potential, including a mass parameter and a flattening. We set up a mock galaxy made with a luminous and a dark component all together giving rise to an axisymmetric logarithmic potential.

1.1 Discovery of dwarf galaxies

In the early 1900's astronomers were unaware of the existence of other galaxies outside our own Milky Way (MW). Fuzzy patches in the sky were called 'nebulae' and were part of the MW. In fact the size of the Milky Way itself was not clear at all. Although Immanuel Kant, already in 1755 in his work 'Allgemeine Naturgeschichte und Theorie des Himmels' (Kant 1755), for the first time postulated that other nebulae could be large and distant disks of stars, similarly to the Milky Way, it took until 1920 before astronomers were truly debating about the 'Scale of the Universe' (Shapley & Curtis 1921). In this so-called 'Great Debate' Harlow Shapley and Heber Curtis discussed, among other things as well, whether the universe was composed of one big galaxy and that spiral nebulae were just nearby clouds or whether the universe was composed of many galaxies respectively. A few years later Hubble (1927) identified Cepheid variable stars in M31, the nearest major spiral nebulae, and showed that the distance was much greater than Shapley's proposed size of the Milky Way. The first proof of an extragalactic galaxy had been made.

The discovery of dwarf spheroidals took longer despite the fact that these systems orbit the Milky Way and therefore are relatively nearby. In 1938, the same Shapley announced the discovery of 'A Stellar System of a New Type' in the constellation Sculptor (Shapley 1938). Additional observations ruled out the possibility that it could be an extended cluster of galaxies. On a photographic plate, made in 1908 by S.I. Bailey during a site-testing expedition, the first confirmation of the reality of this object was already made. The same faint patch of light was seen at the position of the Sculptor system after a total exposure of five nights (23h and 16m) with a 1-inch telescope. (van Agt 1978)

The total light emitted by most spiral and elliptical galaxies over their respective area on the sky is comparable to the amount of light the sky itself emits in the same area. Most classical dwarf spheroidial galaxies have a surface brightness of around 1% of the night sky and are therefore much harder to detect. The Sculptor dwarf spheroidal was first found since it is one of the brightest nearby dSph galaxies and since it is located in a particular empty part of the sky. Many photographic plates have confirmed its discovery in the late 30s. (Mateo 2000)

Because dSph's are so difficult to detect, even today more and more dwarf spheroidal galaxies, including Ultra Faint Dwarfs (UFD's) having luminosities well below $10^5 L_\odot$ (Webster et al. 2014), are being discovered in the Local Group and beyond. For example Koposov et al. (2015); Bechtol et al. (2015) used data taken during the first year of the Dark Energy Survey (DES), and claim to have found nine (or eight) new dwarf galaxy candidates respectively in the region near the Large and Small Magellanic Cloud (LMC and SMC). Three objects appear to be new dwarf galaxies, while the others could be extended or disrupted globular clusters. dSph's are very low surface-brightness objects with half light radii of several hundred parsecs (McConnachie 2012) whereas luminous globular clusters mostly have half light radii of $\lesssim$ 10 pc (van den Bergh 2008). In addition, globular clusters do not contain significant amounts of dark matter, in contrast to dwarf spheroidals: the stars in dSph's would have been moving far to fast relative to one another to remain bound (Mateo 2000). Globular clusters typically contain 10^4-10^6 stars in a nearly spherical distribution and like dSphs they do not contain gas and dust (Binney & Tremaine 2008).

1.2 Dwarf galaxies and dark matter

Dwarf spheroidal galaxies are probably the most common type of galaxies in the universe. Since small galaxies are expected to have formed first in the ΛCDM model, it may be that dSph galaxies represent the basic units of galaxy formation. Dwarf spheroidal galaxies are very faint with absolute magnitudes ranging from Mv = -8 to Mv = -13.5 and have a velocity dispersion around 10 km/s (Mateo 1998; Walker et al. 2009b). Since they are abundant and may contain large amounts of dark matter, they might contribute greatly to the mass of the universe. The determination of the amount of dark matter is difficult, as, in order to estimate the mass content of a system, kinematics of gas and/or stars are needed. Since dwarf spheroidal galaxies generally do not contain much gas, we can only gain dynamical information from the velocities of its stars. Due to the large distance to Sculptor, it is expected that the Gaia satellite will only be able to measure proper motions of the horizontal branch (HB) or brightest red giant branch (RGB) stars, with errors similar or larger than the proper motion itself (Battaglia 2007; Jin et al. 2015). Therefore, only stars that can be observed spectroscopically can be used to get (line-of-sight) velocities. Breddels & Helmi (2013) have already used datasets (Battaglia et al. 2006; Walker et al. 2009a; Battaglia et al. 2008) consisting of roughly 2900 and 1700 probable members for Fornax and Sculptor respectively. As these numbers keep on increasing, the characterization of the line-of-sight velocity distributions (LOSVD) will keep on improving. Using the global velocity moments of the dSph's gives information about the mass content of the systems. In addition, the information about the underlying potential of the system and therefore the dark matter distribution can be obtained from studying the change in velocity moments throughout the system.

1.3 Dynamical modeling of dwarf galaxies

Making dynamical models of (dwarf) galaxies is easier when the systems are in dynamical equilibrium. Fortunately, most classical dSph's show no tidal streams, indicating that they are very likely to be in dynamical equilibrium. In this thesis we therefore assume that dynamical equilibrium holds. We shortly describe two methods of dynamical modeling that have been applied to such systems. In both methods the kinematics can be used to infer the mass distribution of the system, under the assumption that a dwarf galaxy may be considered as a collisionless system.

We follow Binney & Tremaine (2008) to show that a Sculptor like galaxy is indeed a collisionless system. Assuming a mass of $M = 10^8 M_\odot$ for the Sculptor dSph (Walker et al. 2007), the crossing time scale at a radius of $R = 1$ kpc will be:

$$t_{cross} \equiv R/v = 1/\sqrt{GM/R^3} = \sqrt{R^3/GM} \approx 47\,\mathrm{Myr} \tag{1.1}$$

For a luminosity $L \simeq 2 \cdot 10^6 L_\odot$ (Mateo 2000) and a typical stellar luminosity $L_\star \simeq L_\odot$ the number of stars will be $N \simeq 2 \cdot 10^6$. In crossing the galaxy once, the mean-square velocity change caused by stellar encounters is:

$$\triangle v^2 \simeq 8N \left(\frac{Gm}{Rv}\right)^2 \ln(\Lambda) \tag{1.2}$$

where $\ln(\Lambda) \simeq \ln\left(\frac{M}{m}\right)$ is the Coulomb logarithm and where $m \simeq M_\odot$ is a typical stellar mass. A typical speed v of a star is the circular velocity of a star at the edge of the galaxy:

$$v^2 \simeq \frac{GM}{R} \tag{1.3}$$

Therefore, the velocity $\boldsymbol{v}$ will change by an order of itself after n_{relax} crossings given by:

$$n_{relax} \equiv \frac{v^2}{\triangle v^2} \approx \frac{\left(\frac{M}{m}\right)^2}{8N\ln(\Lambda)} \approx 3.4 \cdot 10^7 \tag{1.4}$$

Then, the time in which a dSph star has changed its velocity by an order of itself can be estimated by:

$$t_{relax} \equiv n_{relax} \times t_{cross} = 1.6 \cdot 10^6 \,\mathrm{Gyr} \tag{1.5}$$

Since this so-called relaxation time scale is larger than the age of the universe ($t_{relax} > t_H$), its stars move under the influence of a gravitational field generated by a smooth mass distribution, rather than a distribution concentrated into nearly point-like stars.

The distribution function $f(\boldsymbol{x}, \boldsymbol{v}, t)$ of a system is defined such that $f(\boldsymbol{x}, \boldsymbol{v}, t)\,\mathrm{d}^3\boldsymbol{x}\,\mathrm{d}^3\boldsymbol{v}$ is the probability that at time t a randomly chosen star has phase-space coordinates in the given range. By definition the distribution function is normalized such that:

$$\int f(\boldsymbol{x}, \boldsymbol{v}, t)\,\mathrm{d}^3\boldsymbol{x}\,\mathrm{d}^3\boldsymbol{v} = 1 \tag{1.6}$$

The conservation of probability in phase-space is described by the Collisionless Boltzmann Equation (CBE) (Binney & Tremaine 2008):

$$\frac{\partial f}{\partial t} + \boldsymbol{v} \cdot \frac{\partial f}{\partial \boldsymbol{x}} - \frac{\partial \Phi}{\partial \boldsymbol{x}} \cdot \frac{\partial f}{\partial \boldsymbol{v}} = 0 \tag{1.7}$$

where $\Phi(\boldsymbol{x}, t)$ is the total gravitational potential of the system.

1.3.1 Jeans modeling

Since we assume that dwarf galaxies are in dynamical equilibrium, the distribution function is time-independent such that the first term of the CBE equals zero. In Jeans modeling one takes moments by multiplying the CBE by v_j^n and integrating over all velocities in order to obtain the Jeans equations. These moments should be compared to the observed (low-order) velocity moments. In general, assumptions need to be made while doing Jeans modeling.

For example in case of a spherical system, one could make an assumption on the velocity anisotropy

$$\beta(r) = 1 - \frac{\langle v_\theta^2 \rangle + \langle v_\phi^2 \rangle}{2\langle v_r^2 \rangle} \tag{1.8}$$

If one assumes that the density $\nu(r)$ is known and that $\beta(r)$ is constant with radius, then, if the line-of-sight velocities have been measured, we may be able to derive the mass distribution of the system (Binney & Tremaine 2008):

$$\langle v_r^2(r) \rangle = \frac{1}{r^{2\beta}\nu(r)} \int_r^\infty r'^{2\beta} \nu(r') \frac{\mathrm{d}\Phi}{\mathrm{d}r'} \mathrm{d}r' \tag{1.9}$$

We note that the intrinsic moment $\langle v_r^2(r) \rangle$ is not directly observable, but that it relates to the measured line-of-sight velocity moment $\langle v_{los}^2(R) \rangle$ via (Binney & Mamon 1982) :

$$\langle v_{los}^2(R) \rangle = \frac{2}{I(R)} \int_R^\infty \left(1 - \beta \frac{R^2}{r^2}\right) \langle v_r^2(r) \rangle \frac{r\mathrm{d}r}{\sqrt{r^2 - R^2}} \tag{1.10}$$

where $I(R)$ is the surface brightness and R the projected radius.

Using the spherical Jeans equations Walker et al. (2007) found a virial mass of $\sim (10^8 - 10^9) M_\odot$ for all dSph's. Modeling the Sculptor dSph, Battaglia et al. (2008) found a best-fit core radius of 0.5 kpc and an enclosed mass within 1.8 kpc of $(3.4 \pm 0.7) \times 10^8 M_\odot$, resulting in a mass-to-light ratio of $(158 \pm 33)(M/L_\odot)$.

For an axisymmetric system, Binney & Tremaine (2008) show that if one assumes that the density $\nu(R, z)$ and potential functional form $\Phi(R, z)$ are known and that if the distribution function is of the form $f(H, L_z)$ (such that the mixed moments vanish, $\langle v_R^2 \rangle = \langle v_z^2 \rangle$ and $\langle v_R \rangle = \langle v_z \rangle = 0$), that the Jeans equations can be solved. Here H and L_z denote the Hamiltonian and the angular momentum in the z-direction of the system. The disadvantages of Jeans modeling is that one should make such assumptions. In addition, not every solution to the Jeans equations has an associated distribution function that is physical and therefore positive everywhere.

Hayashi & Chiba (2015) have applied axisymmetric Jeans modeling to infer the axis ratio of the dark matter density distribution (Q) in several dSph's. For the Sculptor dwarf galaxy they find a very low axis ratio ($Q = 0.45 \pm 0.03$), whereas the observed projected flattening in the light ($q\prime$) is 0.68. They find a scale length of 0.6 kpc and an inclination of 87 degrees. Comparing both an oblate and prolate case, they find that the oblate case yields a much better fit than the prolate case.

1.3.2 Schwarzschild modeling

Schwarzschild models where first described by Schwarzschild (1979). In Schwarzschild modeling one assumes a specific gravitational potential. No additional assumptions have to be made for example about the form of the distribution function, which in fact is an output of the modeling. On the other hand the method requires a lot of computing power and therefore a smaller set of gravitational potentials can be explored than for Jeans modeling.

In Schwarzschild modeling orbits are used as building blocks of a system. Given a potential, a complete set of orbits are integrated numerically and for each orbit the predicted observables are stored in an orbit library. Varying the parameters of the potential or varying the potential form as a whole, will result in different libraries. Therefore, a lot of orbits need to be integrated and stored. The library from which a combination of weighted orbits matches the observations (light profile + kinematics) the best, corresponds to the best-fit potential and provides its corresponding distribution function, which will always be non-negative everywhere since the orbital weights are kept positive by construction. In figure 1.2 the principle of Schwarzschild modeling is visualized: a combination of modeled orbits are selected such that they match the observed properties of a galaxy.

1.4 This thesis

In this work we will test the Schwarzschild method on a mock galaxy with axisymmetric properties. The schwarzschild method has already been applied to elliptical galaxies with a central black hole in the axisymmetric regime (Cretton et al. 1999; van der Marel et al. 1998) and in the triaxial case (van den Bosch et al. 2008). However, in those models the dark matter was assumed to follow the light. Here we extend the work of Breddels et al. (2013) in their models of dSph's with the Schwarzschild technique beyond spherical symmetry, and consider an axisymmetric mass distribution both for the light and the dark components. To be able to test the reliability of our conclusions when using this modeling technique, we need to set up a mock galaxy model and simulate mock data (chapter 2). In chapter 3 we describe the Schwarzschild method, how we use it and how we performed basic tests to confirm that we implemented the method correctly (see section 3.6). Then, in chapter 4, we applied Schwarzschild modeling and show that we can recover the characteristic parameters of the mock galaxy potential. In chapter 5 we model our mock galaxy by an axisymmetric NFW potential form and show that we constrain the mass, scale radius and flattening for realistic datasets. We conclude in chapter 6.

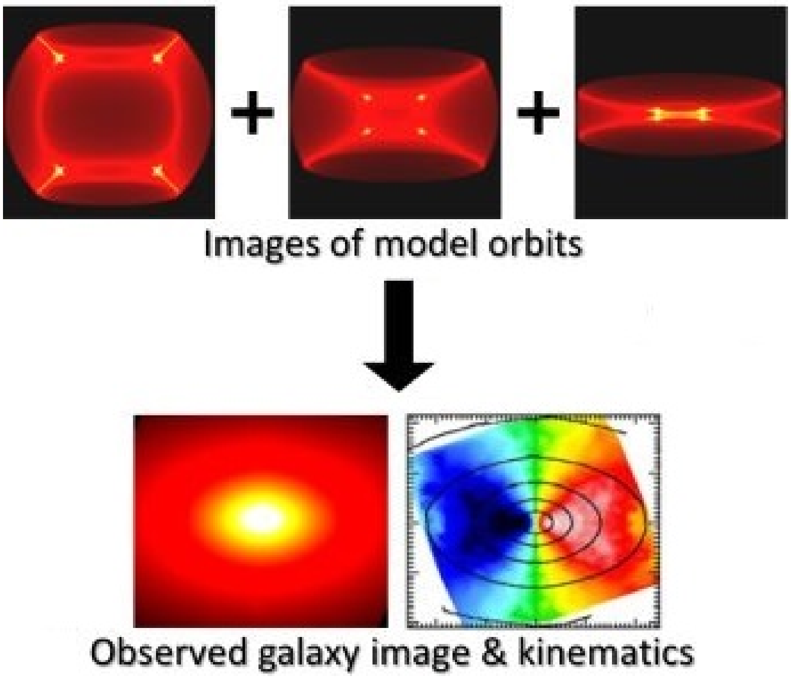


Figure 1.2: Illustrating the principle of Schwarzschild modeling. Individual orbits are combined to make the desired galaxy. Figure from Cappellari (2015).

Chapter 2

A mock galaxy

We have built a mock galaxy with characteristics similar to the Sculptor dSph and thus placed it at a distance of 80 kpc ($d = 0.08$ Mpc) and pretended to observe it with a square field of view (FOV) with a size of 7832"x7832", which then corresponds to 3x3 kpc, centered on the mock galaxy. Although we could have chosen any viewing angles, we have only explored an edge-on view in this work.

In this chapter, we first describe the properties of our mock galaxy. Then we explain how we constructed it and the various checks performed. We then generated mock observations, including realistic errors.

2.1 A composite axisymmetric mock dwarf galaxy

We chose to make an axisymmetric mock galaxy based on the composite system described by Evans (1993), since it has a simple distribution function. In this section we summarize the most important equations from that paper, which were used in our work.

Evans shows that analytically nice results can be derived for the line-of-sight velocity profile if the composite system has an axisymmetric logarithmic relative potential,

$$\Psi_{tot}(= -\Phi_{tot}) = -\frac{1}{2} v_0^2 \ln \left(R_c^2 + R^2 + \frac{z^2}{q^2} \right) \tag{2.1}$$

and a stellar component described by:

$$\rho_{lum} = \frac{\rho_0 R_c^p}{\left(R_c^2 + R^2 + \frac{z^2}{q^2} \right)^{p/2}} \tag{2.2}$$

Here the cylindrical coordinates (R, ϕ, z) are used. The parameter v_0 is a mass parameter, R_c is the core radius and parameter q is the axial ratio of the spheroidal equipotentials, which needs to satisfy $1/\sqrt{2} = 0.707 \leq q \leq 1.08$ to yield a distribution that is positive everywhere (Binney & Tremaine 2008; Evans 1993). With this choice, q is also the flattening of the stellar density component. In addition, the luminous density is characterized by the central density ρ_0 and slope parameter p. For $p > 3$ the total mass of the luminous component is finite and given by (Evans 1993):

$$M_{lum} = \frac{2^{p-3}}{p-2} \mathrm{B} \left[\frac{p-3}{2}, \frac{p-3}{2} \right] \pi \rho_0 q R_c^3 \tag{2.3}$$

where B is the beta function. The surface brightness profile is found by integrating the luminous density along the line of sight (Evans 1993):

$$I(x', y') = \frac{2^{p-2} \mathrm{B}[(p-1)/2, (p-1)/2]\, q \rho_0 R_c^p}{q' \left[x'^2 + R_c^2 + \dfrac{y'^2}{q'^2} \right]^{(p-1)/2}} \tag{2.4}$$

where ($x\prime$,$y\prime$) are the coordinates on the plane of the sky, $q\prime \equiv [\cos^2(i) + q^2 \sin^2(i)]^{0.5}$ and where i is the inclination towards the system.

The total density of the logarithmic potential is given by (Binney & Tremaine 2008):

$$\rho_{tot}(R,z) = \frac{v_0^2}{4\pi G q^2} \frac{(2q^2+1)R_c^2 + R^2 + \left(2 - \frac{1}{q^2}\right) z^2}{\left(R_c^2 + R^2 + \frac{z^2}{q^2}\right)^2} \tag{2.5}$$

We note that the flattening of the total density is not equal to q. In fact it is not constant as will be pointed out in section 5.1.

Both the dark mass and luminous mass have elementary distribution functions. The distribution function of the luminous mass is used to derive the kinematics of the stars:

$$F_{lum} = \mathrm{D} \exp(pE/v_0^2) \tag{2.6}$$

where $\mathrm{D} = \rho_0 R_c^p \left(\frac{p}{2\pi v_0^2}\right)^{3/2}$ is a constant and where $E = \Psi_{tot} - 0.5v_0^2$ is the binding energy. Since the luminous distribution function depends of the binding energy only, the velocity dispersion tensor is isotropic. Furthermore, because $F_{lum} = \mathrm{D}\exp[p\Psi/v_0^2]\exp[-pv^2/2v_0^2]$, integration over the line-of-sight and the tangential velocity components yields the result that the line-of-sight velocity profile is exactly Gaussian and has a velocity dispersion that is isotropic and constant everywhere:

$$\sigma = \frac{v_0}{\sqrt{p}} \tag{2.7}$$

In this work we made a mock galaxy with $v_0 = 20$ km/s, $R_c = 1$ kpc, $p = 3.5$. This means that the velocity dispersion is roughly 10.7 km/s (resulting in a second velocity moment of 114.3 km^2/s^2), a realistic number compared to the velocity dispersions of the classical dSph's. The central density parameter ρ_0 is irrelevant, since it has no influence on the total gravitional potential, which is fully determined by v_0, R_c and q: it is only an amplitude for the light profile, which will be normalized anyway (see section 3.5). Therefore, lowering ρ_0 is only equivalent to increasing the $\left(\frac{M}{L}\right)$-ratio of the system, which is given by (Evans 1993):

$$\begin{aligned} \upsilon(R,z) \equiv \left(\frac{M}{L}\right) &= \frac{\rho_{tot}}{\rho_{lum}} \\ &= \frac{v_0^2}{4\pi G \rho_0 q^2 R_c^p} \left[(2q^2+1)R_c^2 + R^2 + (2 - \frac{1}{q^2})z^2\right] \left(R_c^2 + R^2 + \frac{z^2}{q^2}\right)^{\frac{p-4}{2}} \end{aligned} \tag{2.8}$$

2.2 Constructing the mock galaxy

In this section we will decribe how we constructed our mock galaxy consisting of $N = 10^5$ stars by using the luminous density (equation 2.2) and distribution function (equation 2.22). We generate the positions and velocities of the mock stars separately.

2.2.1 Positions

Given the stellar density profile we can randomly draw positions of N stars to generate the spatial distribution of our mock galaxy. Since the probability functions are not independent in cylindrical coordinates (see equation 2.2 for the density), we need a conditional density function. This works as follows: suppose we want to generate a sample from a 2D joint density function $p(x,y)$. The first step is to draw a sample X whose x-coordinates follow the marginal density function $p(x) = \int_{-\infty}^{\infty} p(x,y)\,\mathrm{d}y$. Then, since $p(x,y) = p(y|x)p(x)$, we generate a sample Y according to the conditional density function $p(y|X)$.

In practise, we start by making the probability distribution function of the stellar density, $p(R,z,\phi)$ itself. It is defined such that

$$\int_0^{2\pi}\int_{-\infty}^{\infty}\int_0^{\infty} p(R,z,\phi)\,\mathrm{d}R\,\mathrm{d}z\,\mathrm{d}\phi \equiv 1 \tag{2.9}$$

and

$$p(R,z,\phi) = \frac{\rho_{lum}(R,z)R}{M_{lum}} \tag{2.10}$$

To compute the total luminous mass and its relation to the characteristic parameters of the density profile we proceed as follows

$$\begin{aligned}
M_{lum} &= \int_0^{2\pi}\int_{-\infty}^{\infty}\int_0^{\infty} \rho_{lum}(R,\phi,z)\,R\mathrm{d}R\,\mathrm{d}z\,\mathrm{d}\phi \\
&= \int_{-\infty}^{\infty}\int_0^{\infty} \rho_{lum}(R,z)\,2\pi R\mathrm{d}R\,\mathrm{d}z \\
&= \int_{-\infty}^{\infty}\int_0^{\infty} \frac{\rho_0 R_c^p}{\left(R_c^2+R^2+\frac{z^2}{q^2}\right)^{p/2}}\,2\pi R\mathrm{d}R\,\mathrm{d}z \\
&= \int_{-\infty}^{\infty}\int_0^{\infty} \pi\rho_0 R_c^p\left(R_c^2+R^2+\frac{z^2}{q^2}\right)^{-p/2}\,\mathrm{d}R^2\,\mathrm{d}z \\
&= \int_{-\infty}^{\infty} \frac{\pi\rho_0 R_c^p}{1-p/2}\left(R_c^2+R^2+\frac{z^2}{q^2}\right)^{1-p/2}\Bigg|_0^{\infty}\,\mathrm{d}z
\end{aligned} \tag{2.11}$$

for $p \neq 2$. Substituting $z \equiv qR_c\tan(\theta)$, such that $\mathrm{d}z = qR_c\sec^2(\theta)\mathrm{d}\theta$, we continue with:

$$\begin{aligned}
M_{lum} &= -\int_{-\infty}^{\infty} \frac{\pi\rho_0 R_c^p}{1-p/2}\left(R_c^2+\frac{z^2}{q^2}\right)^{1-p/2}\,\mathrm{d}z \\
&= -\left(\frac{\pi\rho_0 R_c^p}{1-p/2}\right)\int_{-\pi/2}^{\pi/2}\left[R_c^2\left(1+\tan^2(\theta)\right)\right]^{1-p/2} qR_c\sec^2(\theta)\mathrm{d}\theta \\
&= -\left(\frac{\pi\rho_0 R_c^3 q}{1-p/2}\right)\int_{-\pi/2}^{\pi/2}\left[\sec^2(\theta)\right]^{1-p/2}\sec^2(\theta)\mathrm{d}\theta \\
&= -\left(\frac{\pi\rho_0 R_c^3 q}{1-p/2}\right)\int_{-\pi/2}^{\pi/2}\sec^{4-p}(\theta)\,\mathrm{d}\theta
\end{aligned} \tag{2.12}$$

Then for $p = 3.5$, we get:

$$\begin{aligned}
M_{lum} &= \frac{4\pi\rho_0 R_c^3 q}{3}\int_{-\pi/2}^{\pi/2}\sqrt{\sec(\theta)}\,\mathrm{d}\theta \\
&= \frac{8\sqrt{2}\pi\rho_0 R_c^3 q}{3}\,K(\frac{1}{2}) \\
&\simeq 22.0\,\rho_0 R_c^3 q
\end{aligned} \tag{2.13}$$

The complete elliptical integral of the first kind, $K(x)$, is defined by evaluating the incomplete elliptical integral of the first kind, $F(y|x) = \int\limits_0^y \frac{1}{\sqrt{1-x\sin^2(t)}}\mathrm{d}t$, at $y = \pi/2$. Thus, $K(x) = F(\pi/2|x)$.

So now we can compute $p(R, z, \phi)$:

$$p(R,z,\phi) = \frac{\rho_{lum}(R,z)R}{M_{lum}} = \left(\frac{3\sqrt{R_c}}{8\sqrt{2}\pi q K(\frac{1}{2})}\right)\frac{R}{\left(R_c^2+R^2+\frac{z^2}{q^2}\right)^{\frac{7}{4}}} = p(R,z)p(\phi) \tag{2.14}$$

In our axisymmetric case the coordinate ϕ is independent of R and z. These coordinates $\boldsymbol{\phi}$ can therefore be generated separately. Since we have a uniform probability in ϕ, one can generate $\boldsymbol{\phi}$ very easily. Since ϕ can obtain values between 0 and 2π and since $\int\limits_0^{2\pi} p(\phi)\mathrm{d}\phi \equiv 1$, we directly see that $p(\phi) = 1/2\pi$. For every ϕ-coordinate we draw a random number u_1 from the continuous uniform distribution over the half-open interval between 0 and 1 and compute $\phi = 2\pi u_1$.

Now we continue with the coordinates z and R. We first need the probability function of R and z only:

$$p(R,z) = p(R,z,\phi)/p(\phi) = p(R,z,\phi)\,2\pi \tag{2.15}$$

In the following we decide to generate a sample $\boldsymbol{z}$ before generating a sample $\boldsymbol{R}$. Generating $\boldsymbol{R}$ first is possible as well but results in a harder integral that needs to be solved. We follow the sampling procedure as explained above. So we first compute $p(z)$:

$$\begin{aligned}
p(z) &= \int\limits_0^\infty p(R,z)\,\mathrm{d}R \\
&= \int\limits_0^\infty \left(\frac{3\sqrt{R_c}}{4\sqrt{2}qK(\frac{1}{2})}\right)\frac{R}{\left(R_c^2+R^2+\frac{z^2}{q^2}\right)^{\frac{7}{4}}}\,\mathrm{d}R \\
&= \int\limits_0^\infty \left(\frac{3\sqrt{R_c}}{8\sqrt{2}qK(\frac{1}{2})}\right)\frac{1}{\left(R_c^2+R^2+\frac{z^2}{q^2}\right)^{\frac{7}{4}}}\,\mathrm{d}R^2 \\
&= \left(\frac{3\sqrt{R_c}}{8\sqrt{2}qK(\frac{1}{2})}\right)\left(-\frac{4}{3}\right)\left(R_c^2+R^2+\frac{z^2}{q^2}\right)^{-\frac{3}{4}}\Bigg|_0^\infty \\
&= \left(\frac{\sqrt{R_c}}{2\sqrt{2}qK(\frac{1}{2})}\right)\left(R_c^2+\frac{z^2}{q^2}\right)^{-\frac{3}{4}}
\end{aligned} \tag{2.16}$$

As for u_1, we draw a random number u_2 for every z-coordinate and, while substituting $z \equiv qR_c\tan(\theta)$ again, we solve:

$$
\begin{aligned}
\int_0^{u_2} 1 du_2\prime &= \int_{-\infty}^{z} p(z')\mathrm{d}z' \\
u_2 &= \left(\frac{\sqrt{R_c}}{2\sqrt{2}qK(\frac{1}{2})}\right) \int_{-\infty}^{z} \left(R_c^2 + \frac{z'^2}{q^2}\right)^{-\frac{3}{4}} \mathrm{d}z' \\
u_2 &= \left(\frac{\sqrt{R_c}}{2\sqrt{2}qK(\frac{1}{2})}\right) \int_{-\pi/2}^{\theta} \left[R_c^2\left(1+\tan^2(\theta')\right)\right]^{-\frac{3}{4}} qR_c \sec^2(\theta')\mathrm{d}\theta' \\
u_2 &= \left(\frac{1}{2\sqrt{2}K(\frac{1}{2})}\right) \int_{-\pi/2}^{\theta} \left[sec^2(\theta')\right]^{-\frac{3}{4}} \sec^2(\theta')\mathrm{d}\theta' \\
u_2 &= \left(\frac{1}{2\sqrt{2}K(\frac{1}{2})}\right) \int_{-\pi/2}^{\theta} \sqrt{sec(\theta')}\,\mathrm{d}\theta'
\end{aligned}
\tag{2.17}
$$

Using Mathematica we find:

$$
u_2 = \left(\frac{1}{2\sqrt{2}K(\frac{1}{2})}\right)\left[2F(\frac{\theta}{2}, 2) + \sqrt{2}K(\frac{1}{2})\right] \tag{2.18}
$$

At this point we could have inverted the equation to compute θ given u_2, however we computed θ numerically by using a solver with bounds $(-\pi/2 \leq \theta \leq \pi/2)$.
Next, we compute the probability function of drawing coordinate R given coordinate z, $p(R|z)$. This is quite simple since:

$$
\begin{aligned}
p(R|z) &= p(R,z)/p(z) \\
&= \frac{3}{2}R\left(R_c^2 + R^2 + \frac{z^2}{q^2}\right)^{-\frac{7}{4}}\left(R_c^2 + \frac{z^2}{q^2}\right)^{\frac{3}{4}}
\end{aligned}
\tag{2.19}
$$

Again we draw a random number u_3 for every R-coordinate and we solve:

$$
\begin{aligned}
\int_0^{u_3} 1\mathrm{d}u_3' &= \int_0^{R} p(R'|z)\,\mathrm{d}R' \\
u_3 &= \int_0^{R} \frac{3}{2}R'\left(R_c^2 + R'^2 + \frac{z^2}{q^2}\right)^{-\frac{7}{4}}\left(R_c^2 + \frac{z^2}{q^2}\right)^{\frac{3}{4}} \mathrm{d}R' \\
u_3 &= \int_0^{R^2} \frac{3}{4}\left(R_c^2 + R'^2 + \frac{z^2}{q^2}\right)^{-\frac{7}{4}}\left(R_c^2 + \frac{z^2}{q^2}\right)^{\frac{3}{4}} \mathrm{d}R'^2 \\
u_3 &= \frac{3}{4}\left(-\frac{4}{3}\right)\left(R_c^2 + R'^2 + \frac{z^2}{q^2}\right)^{-\frac{3}{4}}\left(R_c^2 + \frac{z^2}{q^2}\right)^{\frac{3}{4}}\Bigg|_0^{R^2} \\
u_3 &= -\left(R_c^2 + \frac{z^2}{q^2}\right)^{\frac{3}{4}}\left[\left(R_c^2 + R^2 + \frac{z^2}{q^2}\right)^{-\frac{3}{4}} - \left(R_c^2 + \frac{z^2}{q^2}\right)^{-\frac{3}{4}}\right] \\
u_3 &= \left[1 - \left(\frac{R_c^2 + R^2 + \frac{z^2}{q^2}}{R_c^2 + \frac{z^2}{q^2}}\right)^{-\frac{3}{4}}\right]
\end{aligned}
\tag{2.20}
$$

Inverting the equation gives:

$$R = \sqrt{\left(R_c^2 + \frac{z^2}{q^2}\right)\left([1-u_3]^{-\frac{4}{3}} - 1\right)} \tag{2.21}$$

So, given random numbers u_1,u_2 and u_3 a star's position can be obtained by the equations stated above.

2.2.2 Velocities

Besides the positions, stars have a certain velocity in a certain direction. These velocities are computed through the distribution function

$$F_{lum} = \mathrm{D}\exp[p\Psi/v_0^2]\exp[-pv^2/2v_0^2] \tag{2.22}$$

where $v^2 = v_x^2 + v_y^2 + v_z^2$ or $v^2 = v_R^2 + v_z^2 + v_\phi^2$. Since we see that we could write $F_{lum}(\boldsymbol{x}, \boldsymbol{v}) = F_{lum}(\boldsymbol{x})F_{lum}(\boldsymbol{v}) = F_{lum}(\boldsymbol{v}|\boldsymbol{x})F_{lum}(\boldsymbol{x})$ we conclude that $F_{lum}(\boldsymbol{v}|\boldsymbol{x}) = F_{lum}(\boldsymbol{v}) \propto \exp[-pv^2/2v_0^2]$, which is independent of position and Gaussian in all three velocity components. Therefore in order to generate the velocities we could simply draw N velocities from a Gaussian distribution with velocity dispersion $v_0/\sqrt{p}$ in each direction.

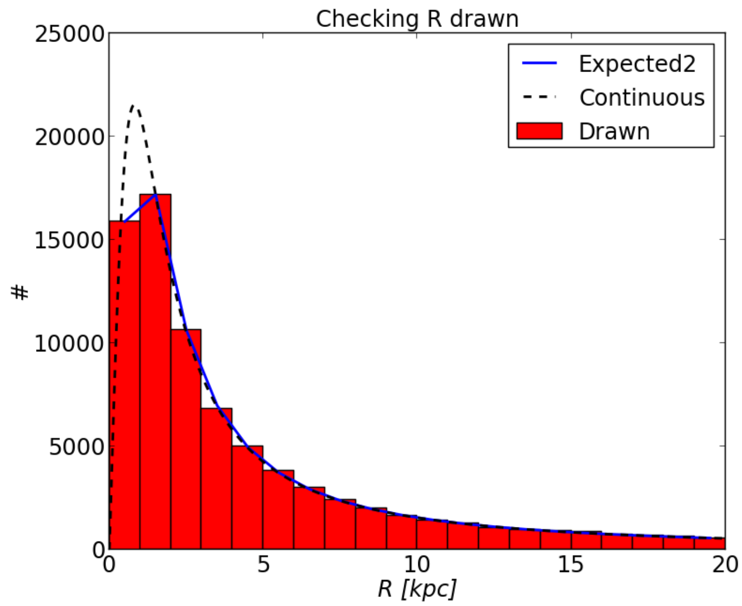
2.3 Checking the properties of our mock galaxy

We checked our procedure by drawing a large number of stars, such that the global properties should not be affected by the random number generator. We used $N = 10^5$ stars to check that this is indeed not the case.

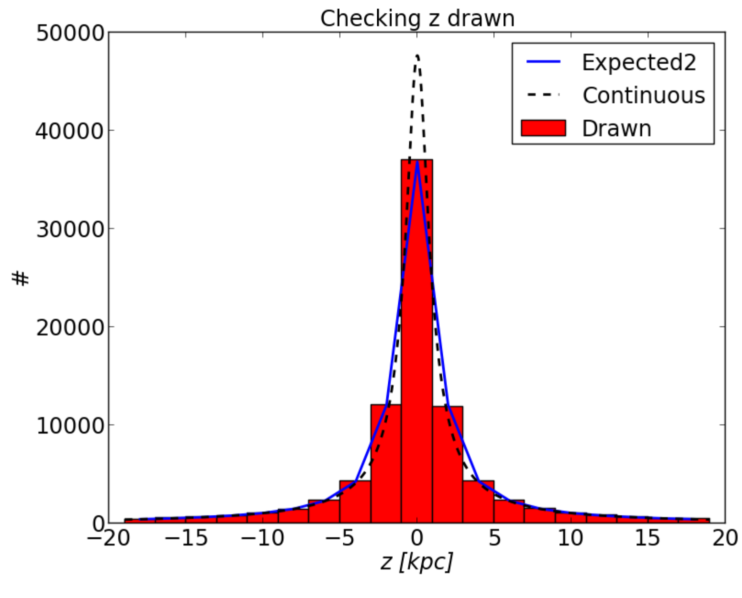
Figure 2.1 shows 1D histograms of the R- and z-coordinates. We compare them to the corresponding theoretical probability curves $p(R)$ (equation 2.23) and $p(z)$ (equation 2.16). To compute $p(R)$, we use the substitution $z \equiv q\sqrt{R_c^2 + R^2}\tan(\theta)$ such that $\mathrm{d}z = q\sqrt{R_c^2 + R^2}\sec^2(\theta)\mathrm{d}\theta$. Then:

$$\begin{aligned}
p(R) &= \int_{-\infty}^{\infty} p(R,z)\,\mathrm{d}z \\
&= \int_{-\infty}^{\infty} \left(\frac{3\sqrt{R_c}}{4\sqrt{2}qK(\frac{1}{2})}\right) \frac{R}{\left(R_c^2 + R^2 + \frac{z^2}{q^2}\right)^{1\frac{3}{4}}}\,\mathrm{d}R \\
&= \int_{-\pi/2}^{\pi/2} \left(\frac{3\sqrt{R_c}R}{4\sqrt{2}qK(\frac{1}{2})}\right) \left[\left(R_c^2 + R^2\right)\left(1 + \tan^2(\theta)\right)\right]^{-1\frac{3}{4}} q\sqrt{R_c^2 + R^2}\sec^2(\theta)\mathrm{d}\theta \\
&= \left(\frac{3\sqrt{R_c}R}{4\sqrt{2}K(\frac{1}{2})}\right)\left(R_c^2 + R^2\right)^{-\frac{5}{4}} \int_{-\pi/2}^{\pi/2} \left(\sec^2(\theta)\right)^{-\frac{7}{4}} \sec^2(\theta)\mathrm{d}\theta \\
&= \left(\frac{3\sqrt{R_c}R}{4\sqrt{2}K(\frac{1}{2})}\right)\left(R_c^2 + R^2\right)^{-\frac{5}{4}} \int_{-\pi/2}^{\pi/2} \sec^{-\frac{3}{2}}(\theta)\,\mathrm{d}\theta \\
&= \left(\frac{3\sqrt{R_c}R}{4\sqrt{2}K(\frac{1}{2})}\right)\left(R_c^2 + R^2\right)^{-\frac{5}{4}} \left[\frac{2\sqrt{2}K(\frac{1}{2})}{3}\right] \\
&= \frac{0.5R\sqrt{R_c}}{(R_c^2 + R^2)^{\frac{5}{4}}}
\end{aligned} \tag{2.23}$$

where the latter integral was solved by using Mathematica.

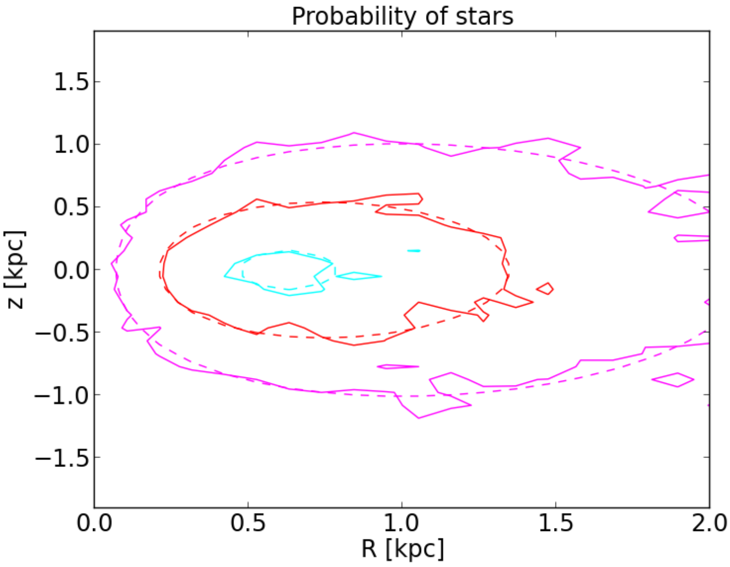


(a)

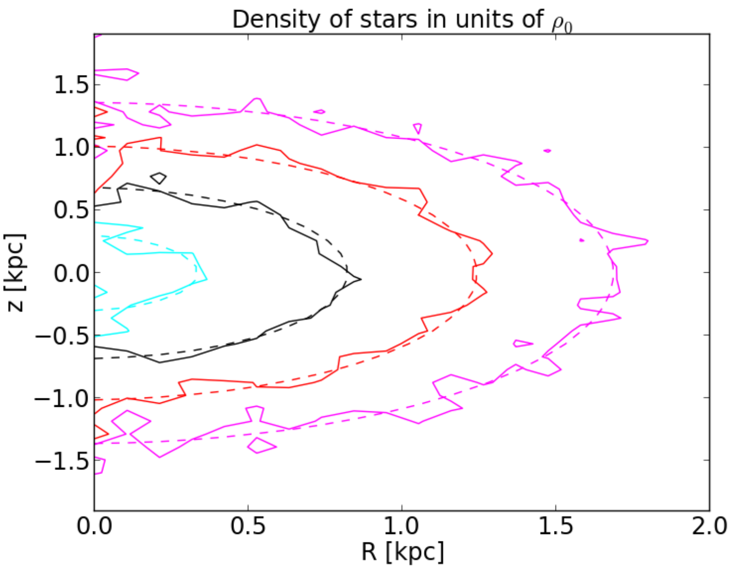


(b)

Figure 2.1: Number counts of stars drawn from the distribution described in section 2.2.1 for 10^5 stars in total. (a): The histogram shows the drawn number of stars in 20 radial bins up to a distance of 20 kpc ($\triangle R = 1$ kpc). The dashed line shows the theoretical function (see equations 2.16 and 2.23). The expected number of stars in each radial bin is shown by the full line and equals the mean number of the dashed line in each radial bin. (b): Similar to (a), but now for the z-coordinate. Here 19 bins are plotted ($\triangle z = 2$ kpc), to get a bin centered around $z = 0$.



(a)



(b)

Figure 2.2: Theoretical and drawn contours for (a) the probability density function $p(R, Z)$ and (b) density $\rho(R, Z)$. Full contours show the drawn probability of stars. Dashed contours show the theoretical probability distribution.

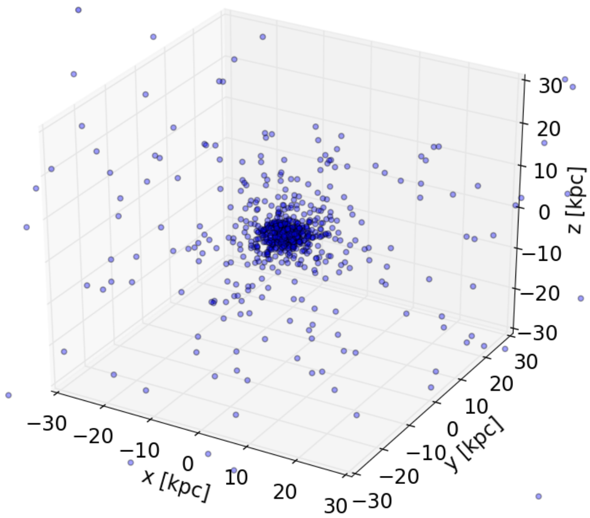


Figure 2.3: A 3-dimensional visualization of our mock galaxy. All axes have units of kpc. We note that a significant fraction of stars have a large distance to the center of the mock galaxy.

In figure 2.2 we confirm that a 2D histogram of our mock stars shows an axial ratio equal to q when plotting the stars in an edge-on view. We do this by first checking whether our dataset follows the $p(R, z)$ distribution. Then we compute $\rho(R, z)$ by multiplying $p(R, z)$ by $M/(2\pi R)$.

To conclude this section, we add a 3d-dimensional visualization of our mock galaxy in figure 2.3. We emphasize that a high fraction of stars have a high distance to the center of the mock galaxy. In fact, depending on the viewing angles towards the system, in between 28.7% (face-on view) and 31.5% (edge-on view) of all drawn stars end up in our field of view (see section 2.1).

2.4 Constructing mock data with realistic errors

Nowadays we have accurate data from roughly 2000 stars of for example the Sculptor dwarf galaxy (see section 1.2). Therefore we need to downsample the number of stars of our mock galaxy to make a fair comparison of what we could find with our Schwarzschild method when applying it to real data. In addition, we always have uncertainties in our velocity measurements. In the nearby dwarf spheroidal galaxies these are of order $dv = 2$ km/s per star (Mateo et al. 1991; Breddels et al. 2013). So, if we want to use our mock galaxy as if we had been observing those stars, we should convolve the line of sight velocities with a Gaussian with standard deviation equal to 2 km/s, under the assumption that the measurement errors are normally distributed and independent. In fact we do not use the kinematics of single stars, but we use the moments by combining the velocities of all available stars in a certain bin on the sky (in the following 'kinematic bin') of our field of view. In this section we describe how we can estimate these the best, including their corresponding errors.

2.4.1 Real raw moments estimators

We follow the method by Breddels et al. (2013) and make small modifications. We define v_i as the real line of sight velocity of star i and ϵ_i as the measurement error on that star. Therefore $v_i + \epsilon_i$ is the observed velocity of star i. We note that the expectation values for the errors, which are drawn from a Gaussian distribution with $\sigma = 2$ km/s, are given by: $E\left[\langle\epsilon_i^n\rangle\right] = E\left[\epsilon_i^n\right] = 0$ for odd n and $s_n \equiv E\left[\langle\epsilon_i^n\rangle\right] = E\left[\epsilon_i^n\right] = (n-1)!!\sigma^n$ for even n. Since we want to know the true value of the moments, i.e. without measurement errors, we compute the moment estimators, $\hat{\mu_n}$, in every kinematic bin. We note that we calculate raw moments, which are not taken about the

mean velocities. In the following we just mention 'moments' to denote 'raw moments'.
For the first moment, the expectation value for the observed moment equals:

$$
\begin{aligned}
E[m_1] &= E\left[\frac{1}{N}\sum_{i=1}^{N}(v_i + \epsilon_i)\right] \\
&= E\left[\frac{1}{N}\sum_{i=1}^{N} v_i\right] + E\left[\frac{1}{N}\sum_{i=1}^{N} \epsilon_i\right] \\
&= E\left[\langle v_i \rangle\right] + \cancelto{0}{E\left[\langle \epsilon_i \rangle\right]} \\
&= \mu_1
\end{aligned}
\tag{2.24}
$$

We note that in our notation the $\mu_n = E\left[\langle v_i^n \rangle\right] = E\left[v_i^n\right]$. Thus, the observed first moment, m_1, equals the real moment, μ_1. In general, we do not know the underlying function form of the velocity profile. Only when assuming a symmetric velocity profile the odd real moments vanish. Therefore, the real moment estimator is given by:

$$
\hat{\mu_1} = \frac{1}{N}\sum_{i=1}^{N}(v_i + \epsilon_i) \tag{2.25}
$$

For the second moment we have:

$$
\begin{aligned}
E[m_2] &= E\left[\frac{1}{N}\sum_{i=1}^{N}(v_i + \epsilon_i)^2\right] \\
&= E\left[\frac{1}{N}\sum_{i=1}^{N}\left(v_i^2 + 2v_i\epsilon_i + \epsilon_i^2\right)\right] \\
&= E\left[\langle v_i^2 \rangle\right] + 2\cancelto{0}{E\left[\langle v_i\epsilon_i \rangle\right]} + E\left[\langle \epsilon_i^2 \rangle\right] \\
&= \mu_2 + s_2
\end{aligned}
\tag{2.26}
$$

where $s_2 = \sigma_i^2 = dv^2$. Thus, the best estimate for the real second moment is given by:

$$
\hat{\mu_2} = \frac{1}{N}\sum_{i=1}^{N}(v_i + \epsilon_i)^2 - s_2 \tag{2.27}
$$

We repeat the procedure up to the 8^{th} moment. Here we show the best estimates for these real moments, assuming $\hat{\mu_n} \simeq \mu_n$ on the right-hand side of the equations. For completeness the fifth and seventh moments are shown as well, although they are not needed in our analysis.

$$
\hat{\mu_3} = \frac{1}{N}\sum_{i=1}^{N}(v_i + \epsilon_i)^3 - 3\mu_1 s_2 \tag{2.28}
$$

$$
\hat{\mu_4} = \frac{1}{N}\sum_{i=1}^{N}(v_i + \epsilon_i)^4 - 6\mu_2 s_2 - 3s_2^2 \tag{2.29}
$$

$$
\hat{\mu_5} = \frac{1}{N}\sum_{i=1}^{N}(v_i + \epsilon_i)^5 - 10\mu_3 s_2 - 15\mu_1 s_2^2 \tag{2.30}
$$

$$
\hat{\mu_6} = \frac{1}{N}\sum_{i=1}^{N}(v_i + \epsilon_i)^6 - 15\mu_4 s_2 - 45\mu_2 s_2^2 - 15s_2^3 \tag{2.31}
$$

$$
\hat{\mu_7} = \frac{1}{N}\sum_{i=1}^{N}(v_i + \epsilon_i)^7 - 21\mu_5 s_2 - 105\mu_3 s_2^2 - 105\mu_1 s_2^3 \tag{2.32}
$$

$$\hat{\mu_8} = \frac{1}{N}\sum_{i=1}^{N}(v_i+\epsilon_i)^8 - 28\mu_6 s_2 - 210\mu_4 s_2^2 - 420\mu_2 s_2^3 - 105 s_2^4 \tag{2.33}$$

In figures 2.4, 2.5, 2.6 and 2.7 we show 2 examples of generating our mock observables. In figures 2.4 and 2.6 we show how we started for the case of observing 2000 stars in a face-on view and 10^4 stars in an edge-on view respectively. We visualize our field of view, in which we randomly select 2000 (or 10^4) stars, we check the Gaussianity of the velocity profile and add velocity measurement errors. Then we divide the field of view in 3x3 kinematic bins. In each kinematic bin the kinematic moments are calculated. For every kinematic bin separately we show the kinematic information in figures 2.5 and 2.7. We show the true (T) and observed (O) velocity profile of our stars. It is clearly visible that in the case of 10^4 stars, all profiles are much closer to a Gaussian distribution than in the case of 2000 stars. The second and fourth raw moment estimators for all 9 kinematic bins are visualized in the first column of the bottom panel. The larger the number of stars the smaller is the range of the obtained kinematic moments: for example, for 2000 stars the second moment ranges in between 90 and 125 $\mathrm{km}^2/\mathrm{s}^2$, whereas for 10^4 stars it ranges in between 104 and 120 $\mathrm{km}^2/\mathrm{s}^2$. The errors and S/N-ratios, shown in the second and last column are described in sections 2.4.2 and 2.4.3.

2.4.2 Realistic error of the real raw moments estimators

To compute the error on these moments, we compute the square root of the variance of the moments; $Var(\mu_N) \approx Var(m_N) = E[{m_N}^2] - (E[m_N])^2$ (Breddels et al. 2013). We will work out the variance for the first two moments and show the results for the other moments.

$$\begin{aligned}
Var(m_1) &= Var\left[\frac{1}{N}\sum_{i=1}^{N}(v_i+\epsilon_i)\right] \\
&= E\left\{\left[\frac{1}{N}\sum_{i=1}^{N}(v_i+\epsilon_i)\right]^2\right\} - \left[E\left\{\frac{1}{N}\sum_{i=1}^{N}(v_i+\epsilon_i)\right\}\right]^2 \\
&= E\left\{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(v_i+\epsilon_i)(v_j+\epsilon_j)\right\} - \mu_1^2 \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[E(v_iv_j) + \cancelto{0}{E(v_i\epsilon_j)} + \cancelto{0}{E(\epsilon_iv_j)} + E(\epsilon_i\epsilon_j)\right] - \mu_1^2 \\
&= \frac{1}{N^2}N(N-1)\left[E(v_i)E(v_j) + \cancelto{0}{E(\epsilon_i)E(\epsilon_j)}\right] \\
&+ \frac{1}{N^2}N\left[E(v_i^2) + E(\epsilon_i^2)\right] - \mu_1^2 \\
&= \frac{(N-1)\mu_1^2}{N} + \frac{\mu_2+s_2}{N} - \mu_1^2 \\
&= \frac{\mu_2+s_2-\mu_1^2}{N}
\end{aligned} \tag{2.34}$$

where we note that $\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{N} 1 = N$ for $i=j$ and $\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{N} 1 = N(N-1)$ for $i \neq j$.

Again, assuming $\mu_n \simeq \hat{\mu_n}$, we can also write:

$$\begin{aligned}
Var(m_1) &= \frac{\mu_2 + s_2 - \mu_1^2}{N} \\
&\simeq \frac{\left[\frac{1}{N}\sum\limits_{i=1}^{N}(v_i+\epsilon_i)^2 - s_2\right] + s_2 - \left[\frac{1}{N}\sum\limits_{i=1}^{N}(v_i+\epsilon_i)\right]^2}{N} \\
&= \frac{1}{N}\left\{\frac{1}{N}\sum_{i=1}^{N}(v_i+\epsilon_i)^2 - \left[\frac{1}{N}\sum_{i=1}^{N}(v_i+\epsilon_i)\right]^2\right\}
\end{aligned} \tag{2.35}$$

For the variance of the second raw moment estimator we find:

$$\begin{aligned}
Var(m_2) &= Var\left[\frac{1}{N}\sum_{i=1}^{N}(v_i+\epsilon_i)^2\right] \\
&= E\left\{\left[\frac{1}{N}\sum_{i=1}^{N}(v_i+\epsilon_i)^2\right]^2\right\} - \left[E\left\{\frac{1}{N}\sum_{i=1}^{N}(v_i+\epsilon_i)^2\right\}\right]^2 \\
&= E\left\{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(v_i+\epsilon_i)^2(v_j+\epsilon_j)^2\right\} - [\mu_2+s_2]^2 \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E\big[v_i^2v_j^2 + \cancelto{0}{2v_i^2v_j\epsilon_j} + v_i^2\epsilon_j^2 + \cancelto{0}{2v_i\epsilon_i v_j^2} + 4v_i\epsilon_i v_j\epsilon_j + \cancelto{0}{2v_i\epsilon_i\epsilon_j^2} + v_j^2\epsilon_i^2 \\
&+ \cancelto{0}{2\epsilon_i^2 v_j\epsilon_j} + \epsilon_i^2\epsilon_j^2\big] - [\mu_2+s_2]^2 \\
&= \frac{N-1}{N}\big[E(v_i^2)E(v_j^2) + E(v_i^2)E(\epsilon_j^2) + \cancelto{0}{4E(v_i\epsilon_i v_j\epsilon_j)} + E(v_j^2)E(\epsilon_i^2) + E(\epsilon_i^2)E(\epsilon_j^2)\big] \\
&+ \frac{1}{N}\big[E(v_i^4) + 6E(v_i^2)E(\epsilon_i^2) + E(\epsilon_i^4)\big] - [\mu_2+s_2]^2 \\
&= \left[\cancel{1} - \frac{1}{N}\right]\left[\mu_2^2 + 2\mu_2 s_2 + s_2^2\right] + \frac{1}{N}\left[\mu_4 + 6\mu_2 s_2 + s_4\right] - \cancel{\left[\mu_2^2 + 2\mu_2 s_2 + s_2\right]} \\
&= \frac{1}{N}\left[\mu_4 - \mu_2^2 + 4\mu_2 s_2 + 2s_2^2\right]
\end{aligned} \tag{2.36}$$

As before, assuming $\mu_n \simeq \hat{\mu_n}$, we get:

$$\begin{aligned}
Var(m_2) &= \frac{1}{N}\left[\mu_4 - \mu_2^2 + 4\mu_2 s_2 + 2s_2^2\right] \\
&\simeq \frac{1}{N}\left\{\left[\frac{1}{N}\sum_{i=1}^{N}(v_i+\epsilon_i)^4 - 3s_s^2 - 6\mu_2 s_2\right] - \mu_2^2 + 4\mu_2 s_2 + 2s_2^2\right\} \\
&= \frac{1}{N}\left\{\frac{1}{N}\sum_{i=1}^{N}(v_i+\epsilon_i)^4 - \left[\frac{1}{N}\sum_{i=1}^{N}(v_i+\epsilon_i)^2 - s_2\right]^2 - s_2^2 - 2s_2\left[\frac{1}{N}\sum_{i=1}^{N}(v_i+\epsilon_i)^2 - s_2\right]\right\} \\
&= \frac{1}{N}\left\{\frac{1}{N}\sum_{i=1}^{N}(v_i+\epsilon_i)^4 - \left[\frac{1}{N}\sum_{i=1}^{N}(v_i+\epsilon_i)^2\right]^2\right\}
\end{aligned} \tag{2.37}$$

Similarly the errors of the third and fourth moment estimators can be computed. They are given by:

$$Var(m_3) = \frac{1}{N}\left[\mu_6 + 15\mu_4 s_2 + 45\mu_2 s_2^2 + 15 s_2^3 - \mu_3^2 - 6\mu_3\mu_1 s_2 - 9\mu_1^2 s_2^2\right] \tag{2.38}$$

$$Var(m_3) = \frac{1}{N}\left\{\frac{1}{N}\sum_{i=1}^{N}(v_i+\epsilon_i)^6 - \left[\frac{1}{N}\sum_{i=1}^{N}(v_i+\epsilon_i)^3\right]^2\right\} \tag{2.39}$$

$$Var(m_4) = \frac{1}{N}\left[\mu_8 + 28\mu_6 s_2 - \mu_4^2 - 12\mu_4\mu_2 s_2 + 204\mu_4 s_2^2 - 36\mu_2^2 s_2^2 + 384\mu_2 s_2^3 + 96 s_2^4\right] \tag{2.40}$$

$$Var(m_4) = \frac{1}{N}\left\{\frac{1}{N}\sum_{i=1}^{N}(v_i+\epsilon_i)^8 - \left[\frac{1}{N}\sum_{i=1}^{N}(v_i+\epsilon_i)^4\right]^2\right\} \tag{2.41}$$

Clearly, the error decreases as the number of stars in a kinematic sky bin increases. In the second columns of figures 2.5b and 2.7b we show the errors obtained like explained in this section. When using $N = 10^5$ stars inside the field of view and 9x9 kinematic bins, the error of the second moment is roughly 3-9 km^2/s^2 and for $N = 10^4$ and 3x3 kinematic bins the error is similar, roughly 4-7 km^2/s^2. The errors on the higher order fourth moment are relatively larger.

We verified the errors by generating M samples, each containing N stars. For every sample we computed the raw moment estimators. Then we computed the standard deviation over those M samples for the first 4 moments. We confirmed this Monte Carlo Method gives errors of the same magnitude as our analytic calculations.

2.4.3 Requiring a high signal to noise ratio

In our analysis we choose to have a high signal to noise (S/N) ratio, here simply defined as the value of the moments divided by their errors, for our moment estimators. This reflects into a minimum number of stars per kinematic bin. In the third columns of figures 2.5b and 2.7b we show the S/N-ratios of our mock datasets. Using either $N = 10^5$ stars/9x9 kinematic bins or $N = 10^4$ stars/3x3 kinematic bins inside the field of view result both in a high S/N-ratio ($\gtrsim 20$). We note that the relative errors on the fourth moments are larger and thus result in lower S/N-ratios. Decreasing the number of stars to 2000 results in S/N-ratios ranging from 3 to 6 for the fourth moment. Taking fewer kinematic bins to decrease the errors does not make sense anymore. As the errors become significant, we must be careful when interpreting results. Therefore we did not model yet datasets containing 2000 stars. To summarize, we used 9x9 kinematic bins ($N_{kin} = 81$) when we simulate an observation containing 10^5 stars and we used 3x3 kinematic bins ($N_{kin} = 9$) when observing 10^4 stars in order to retain a high signal to noise ratio.

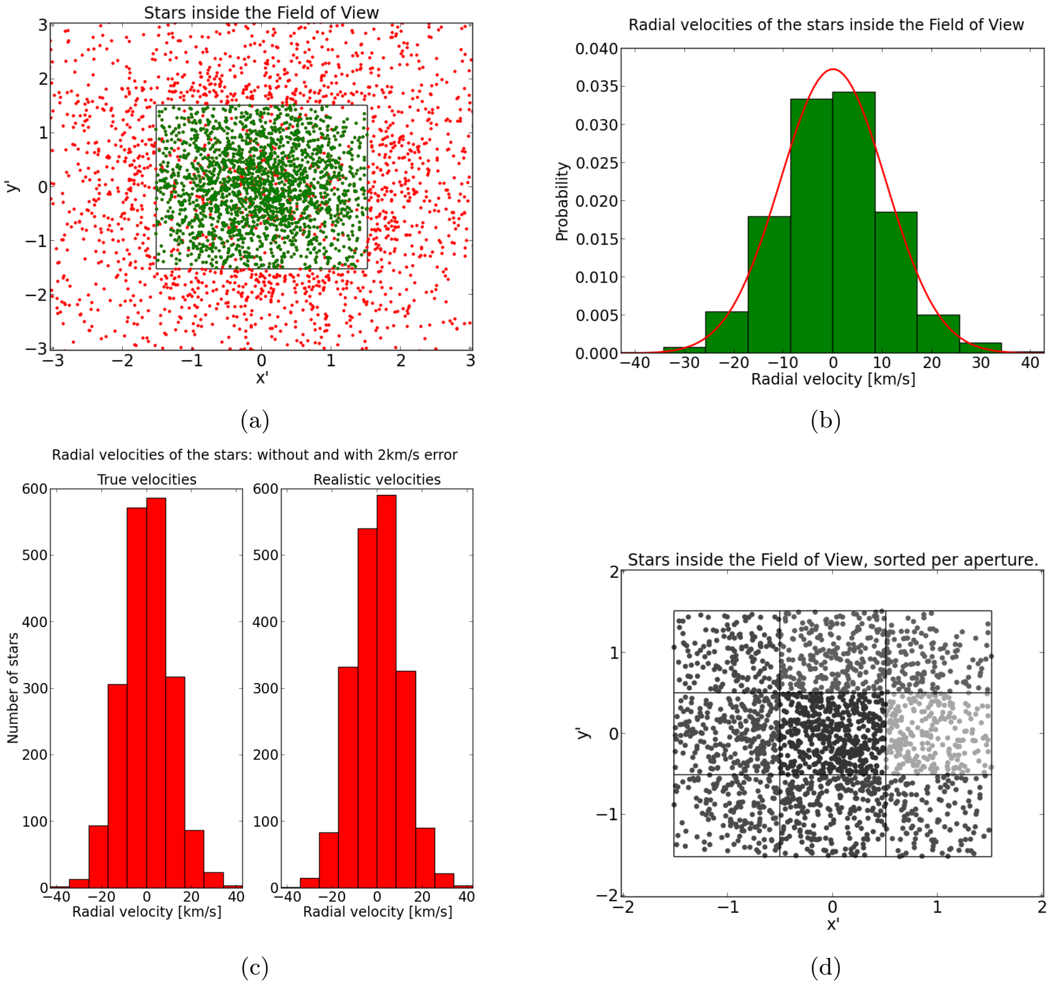
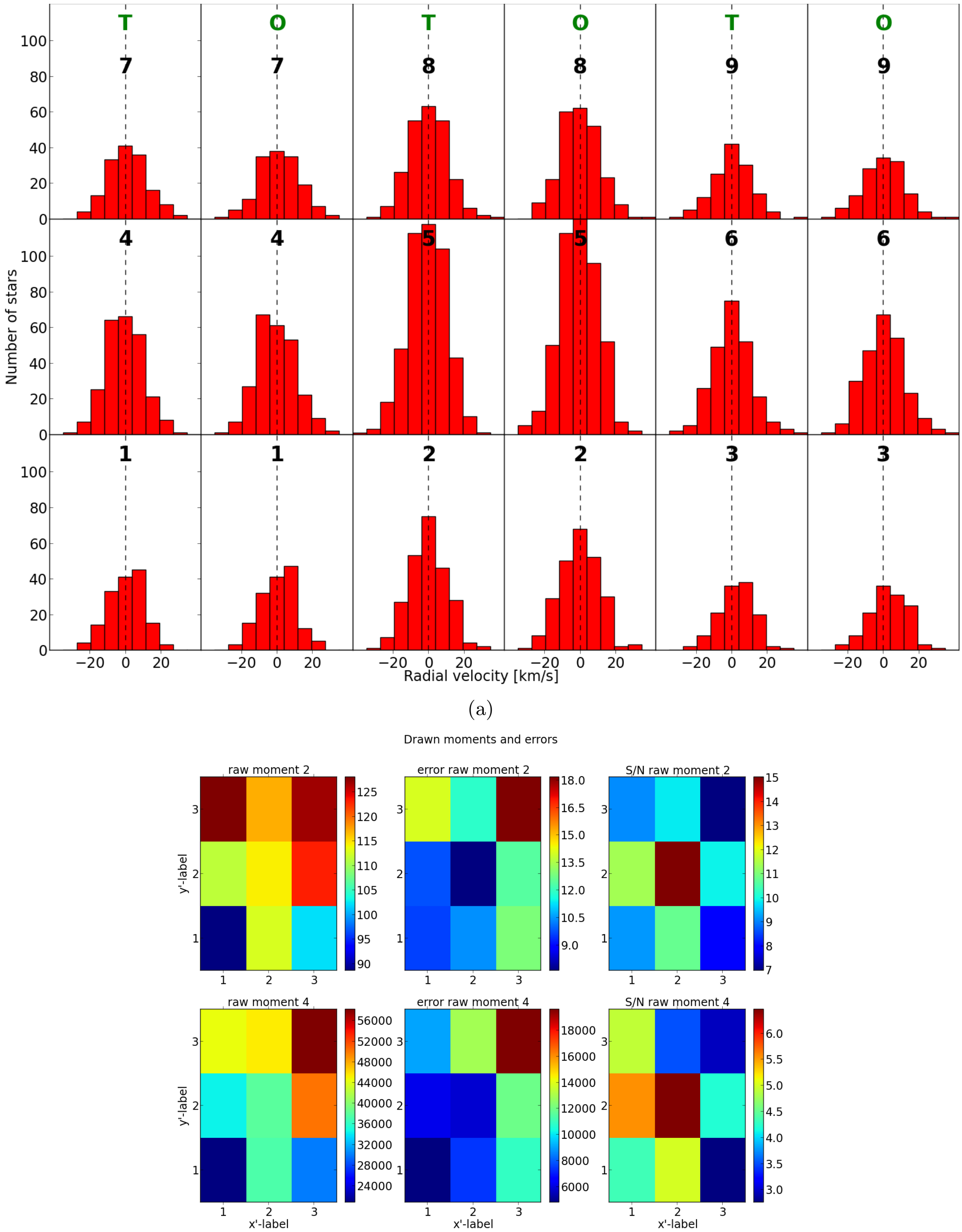


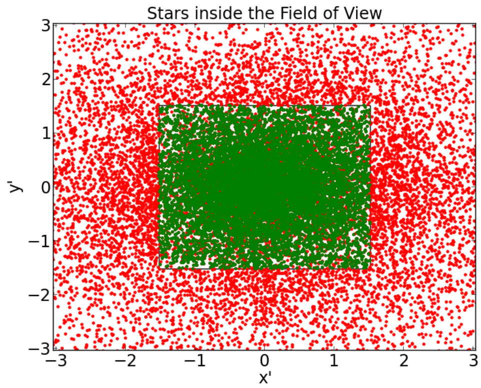
Figure 2.4: Selecting and assigning $N = 2000$ stars in total, using 3x3 kinematic bins and assuming a face-on view: First a sufficient number of stars (red+green in panel (a)) are drawn from the distribution function. In our field of view we randomly select N stars (green) and we verify that the velocities are Gaussian distributed (panel (b)). To all N stars we add a velocity drawn from a Gaussian distribution with standard deviation $dv = 2$ km/s to simulate measurement errors. We pretend that these new velocities (right histogram of panel (c)) are observed. In panel (d) we assigned to proper kinematic sky bin to all N stars. The corresponding moments and errors are visualized in figure 2.5.



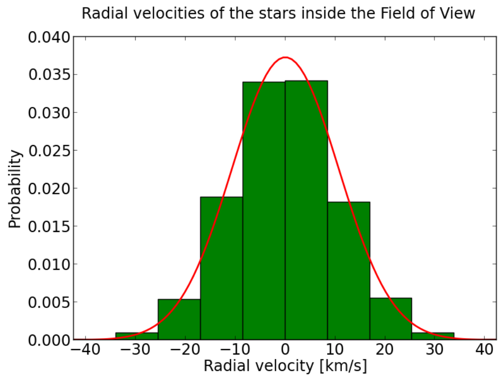
(a)

(b)

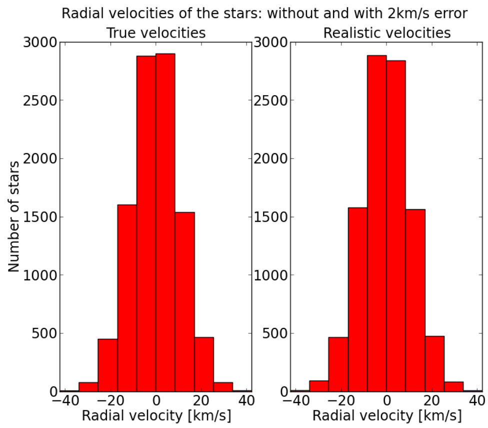
Figure 2.5: Obtaining the kinematic moments for 2000 stars in a face-on view. In panel (a) we show the velocity histograms of the stars for each kinematic bin (see figure 2.4d), before (true velocities: 'T') and after (observed velocities: 'O') adding the measurement errors. For each of the observed velocity histograms we compute the raw moments, as described in section 2.4.1. The second and fourth moments are shown in the left column of panel (b). The error on these moments are estimated by the method described in section 2.4.2. The corresponding errors are shown in the middle column of panel (b). The S/N-ratios, here simply defined as the value of the moments divided by their errors, are shown in the rightmost column of panel (b).



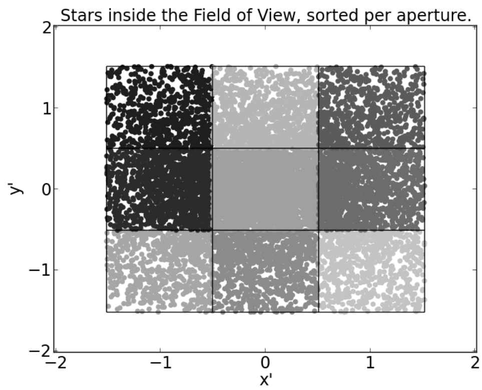
(a)



(b)

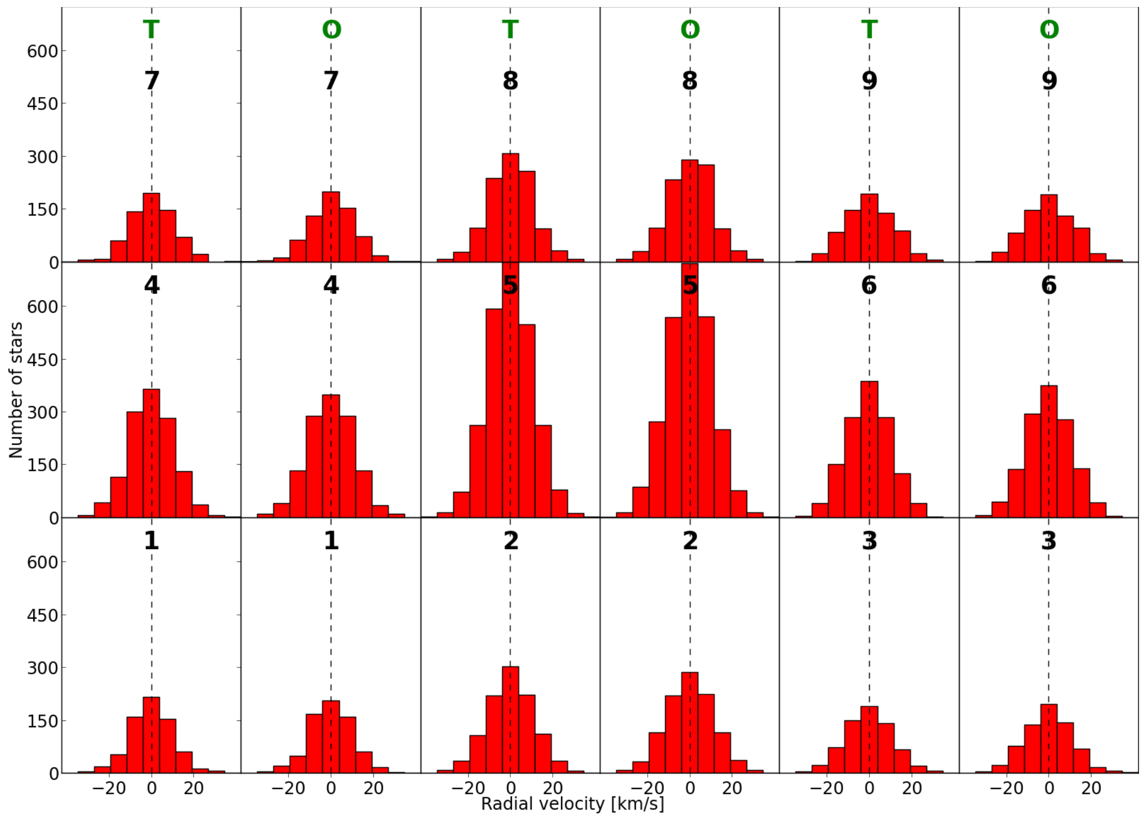


(c)

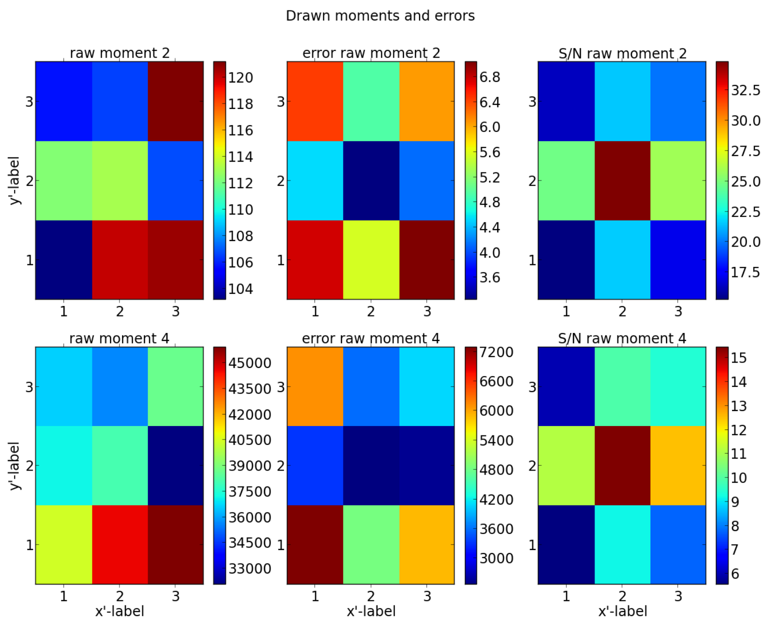


(d)

Figure 2.6: Selecting and assigning $N = 10^4$ stars in total, using 3x3 kinematic bins and assuming an edge-on view. See figure 2.4 for more information.



(a)



(b)

Figure 2.7: Obtaining the kinematic moments for 10^4 stars in an edge-on view. See figures 2.5 and 2.6d for more information. Increasing the number of stars by a factor M does increase the signal-to-noise ratio by a factor $\sqrt{M}$.

Chapter 3

Methodology

In this chapter we will explain in more detail how our Schwarzschild method works. For each potential, as complete as possible sets of orbits are integrated and stored in orbit libraries. Then, a fitting routine assigns orbital weights to the orbits, in such a way that the observational data is fitted the best. In this way, the potential that fits the observables the best will reveal its characteristic parameters. We first describe how the potential and accelerations are computed in order to do orbital integration. Then we describe how the code produces as complete as possible sets of orbits by generating proper initial conditions. Then we explain what orbital properties are stored in the libraries and how this is done. Finally we describe how we can use these libraries together with our 'observed' mock galaxy to find the best-fit model. More information can be found in van den Bosch et al. (2008).

3.1 Potential, mass and acceleration

We use the same code as van den Bosch et al. (2008). In their work they applied their model to the kinematically-decoupled-core galaxy NGC4365. This is a giant E3 elliptical galaxy that shows minor axis rotation. Their models are triaxial and their modeled potentials are based on a mass distribution following the light and a central black hole. They do not implement an additional potential form for a dark matter halo, but the code was written such that this can be implemented very easily. In our mock galaxy the logarithmic potential describes both the luminous and dark matter mass distribution (see equation 2.1). In this section we will, as an example, describe the accelerations from this underlying potential, although any other potential form can be modeled as well (for an example, see chapter 5).

Given the analytical form of the potential, we can compute the accelerations everywhere in all three directions, x, y and z. These accelerations are needed to integrate individual orbits. We use cartesian coordinates since the Schwarzschild code is designed to use them and since there is no need to modify this. We do not include the contribution of a black hole. The accelerations are determined by the gradient of the potential:

$$\boldsymbol{a} = -\bigtriangledown \Phi = -\left(\frac{\mathrm{d}}{\mathrm{d}R}, \frac{\mathrm{d}}{\mathrm{d}z}, \frac{1}{R}\frac{\mathrm{d}}{\mathrm{d}\phi}\right)\Phi \tag{3.1}$$

Of course, in our axisymmetric potential there is no acceleration in the ϕ-direction ($a_\phi = 0$). Its acceleration in the R-direction is:

$$\begin{aligned} a_R(R,z) &= -\frac{\mathrm{d}}{\mathrm{d}R}\Phi(R,z) = -\frac{\mathrm{d}}{\mathrm{d}R}\left[\frac{1}{2}v_0^2 \ln\left(R_c^2 + R^2 + \frac{z^2}{q^2}\right)\right] \\ &= -v_0^2 \frac{R}{R_c^2 + R^2 + \frac{z^2}{q^2}} \end{aligned} \tag{3.2}$$

Similarly, the vertical acceleration is given by:

$$
\begin{aligned}
a_z(R,z) &= -\frac{\mathrm{d}}{\mathrm{d}z}\Phi(R,z) = -\frac{\mathrm{d}}{\mathrm{d}z}\left[\frac{1}{2}v_0^2 \ln\left(R_c^2 + R^2 + \frac{z^2}{q^2}\right)\right] \\
&= -v_0^2 \frac{z/q^2}{R_c^2 + R^2 + \frac{z^2}{q^2}}
\end{aligned}
\tag{3.3}
$$

To convert the acceleration in cylindrical radius to accelerations in cartesian coordinates, we use the simple relations $x = R\cos(\phi)$ and $y = R\sin(\phi)$ and decompose the radial acceleration vector into vectors along the x- and y-direction. Thus:

$$
a_x = a_R \cos(\phi) = a_R \frac{x}{R} = -v_0^2 \frac{x}{R_c^2 + R^2 + \frac{z^2}{q^2}} \tag{3.4}
$$

$$
a_y = a_R \sin(\phi) = a_R \frac{y}{R} = -v_0^2 \frac{y}{R_c^2 + R^2 + \frac{z^2}{q^2}} \tag{3.5}
$$

In stead of providing an analytic expression of the potential, one can also use the light profile of the system to model the potential of the system. To convert a surface brightness profile into a triaxial density, a multi-Gaussian expansion or MGE (Cappellari 2002) can be used and assumptions need to be made about the mass-to-light ratio of the system. An example of such a MGE is given in appendix A.

3.2 Initial conditions

In Schwarzschild modeling orbits are stored in a so-called orbit library. In order to produce an orbit library we need to compute many different orbits, which are specified by initial conditions. In this section, we describe how initial conditions are sampled, after we give a short overview of the different orbital families that might exist in a galaxy.

3.2.1 Types of orbits

In order to be able to identify a best fitting model (in terms of determining the intrinsic properties of the observed galaxy), the libraries need to consist of all orbital types. To make a complete as possible set of orbits one needs to know what orbital types one can expect.
In general, orbits in axisymmetric potentials conserve two or three integrals motion. If an orbit admits two integrals of motion, E and L_z, it would ultimately fill the entire area within the zero velocity curve (ZVC). The ZVC is defined as the curve in the meridional (R,z)-plane where the velocity of the orbit in this plane is zero ($v_R = v_z = 0$). The ZVC is unique for every combination of energy and angular momentum. The ZVC of an orbit equals its equipotential ($E = \Psi$) in the limit $L_z \to 0$ and reduces to a point when $L_z \to L_{z,circ}$, the circular orbit (see figure 3.1). If an orbit is restricted to a sub-area within the ZVC (see figure 3.2), then this means that the orbit admits a third integral of motion, which is not a classical one, i.e. we do not know yet how to express it analytically. Since the angular momentum in the z-direction is conserved in an axisymmetric potential, orbits with nonzero L_z never cross the center. These Z-tube orbits (they are often referred to as loop orbits) circulate in a fixed sense about the center of the potential, while oscillating in radius. The circular orbit that orbits the z-axis is the parent of the Z-tubes: a circular orbit closes on itself after one revolution and does not oscillate in radius. Given its energy it has the highest possible angular momentum, $L_{z,circ}$.

As we use a triaxial Scwarzschild code, the initial conditions generator also samples box orbits, to be able to reproduce triaxial systems (see section 3.2.4). Box orbits have zero time-averaged angular momentum. In this orbital family a star eventually passes close to every point inside a rectangular box and therefore can cross the center.

In addition to tube orbits and box orbits, a subset of phase-space will also be occupied by resonant or irregular orbits, although this will not be a large fraction.

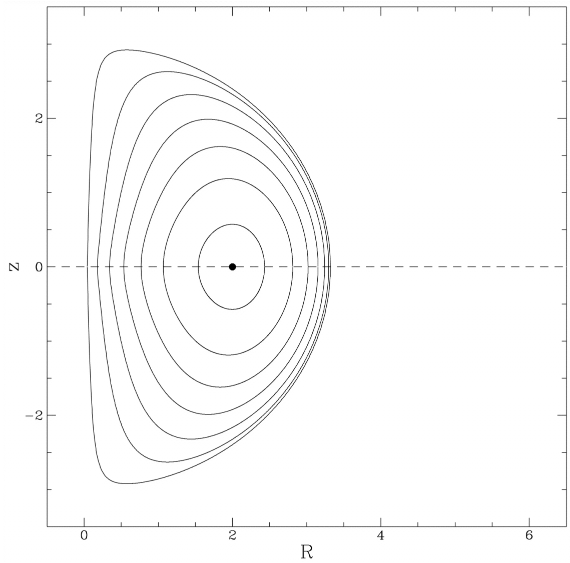


Figure 3.1: The ZVC curves for seven values of L_z at an energy corresponding to a circular orbit indicated by the dot. The higher the angular momentum, the smaller the region covered by the ZVC. The figure is taken from Cretton et al. (1999) and is based on their test model.

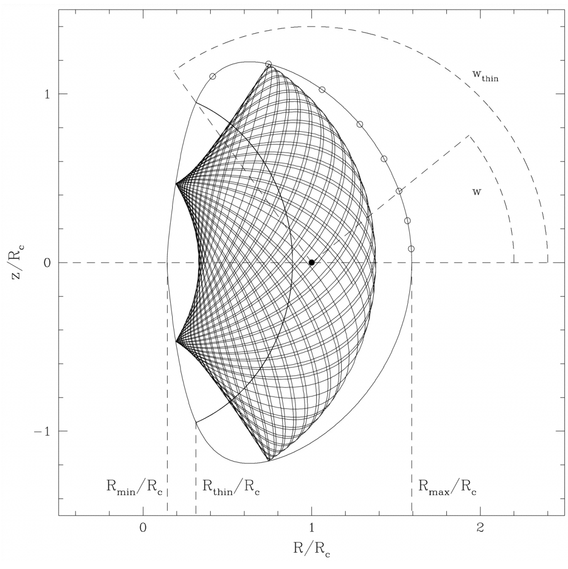


Figure 3.2: A regular orbit, admitting a third integral of motion, the thin tube orbit and the ZVC curve around them in the meridional plane. The energy of the orbits correspond to the energy of the circular orbit indicated by the dot. The definition of the angles w_{thin} and w are indicated. The figure is taken from Cretton et al. (1999) and is based on their test model.

3.2.2 Initial conditions from observable space

Most often two options are used to generate initial conditions. The first one is to sample the space of the integrals of motion E, L_z and I_3. Every orbit with nonzero L_z touches its ZVC (Ollongren 1962). Given a set of (E, L_z) an angle w or a turning radius R_{zvc} is used to parametrize I_3. This angle (or radius respectively) was defined by the point at which the orbit touches the ZVC. To be more precise, for every (E, L_z) there is exactly one orbit, the so-called thin-tube orbit, that touches the ZVC at only one angle w_{thin}. All other regular orbits touch the ZVC for at least two values of w, one smaller and one larger than w_{thin} (see figure 3.2). Orbits that do not have a third integral of motion touch the ZVC everywhere. See Levison & Richstone (1985); van der Marel et al. (1998); Cretton et al. (1999) for a more detailed description.

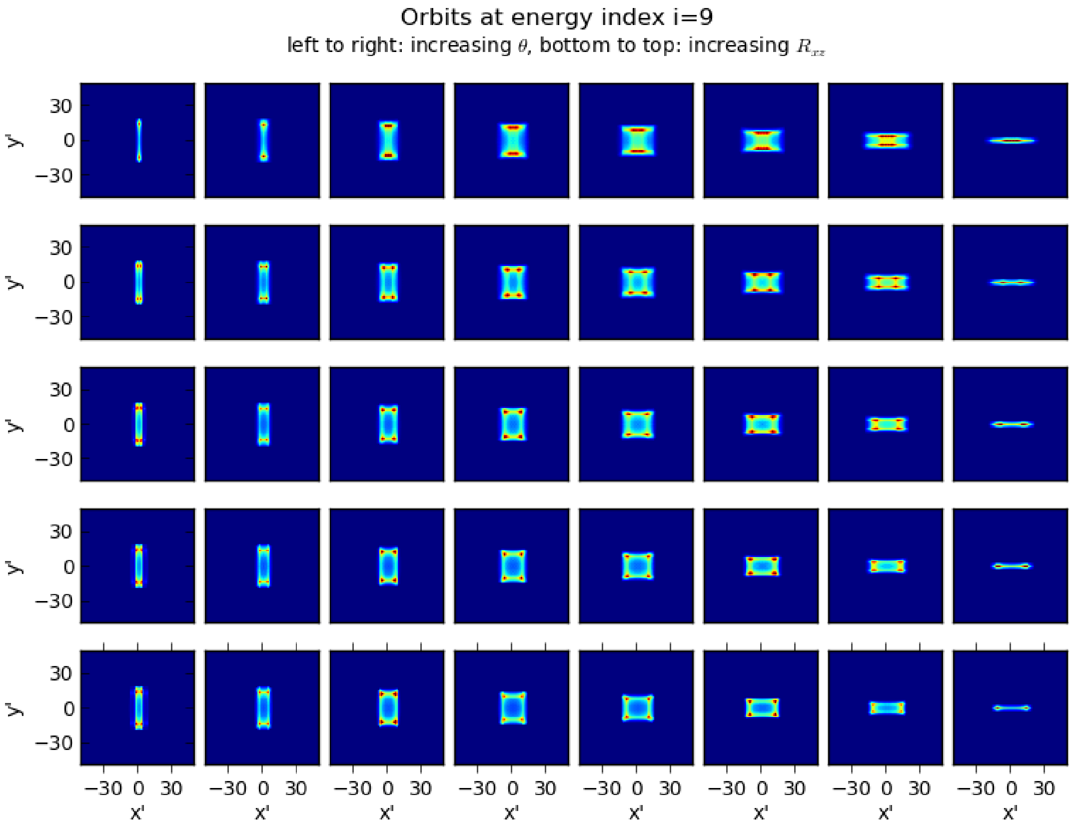


Figure 3.3: The surface brightness profiles of the individual orbits shown in all 99x99 light bins of our field of view (3x3 kpc), seen edge-on, having an energy index of 9 in the library of q80v20 from attempt 3 in section 4.3 which consist of 20 different energies and $N_{I_2} = 8$ and $N_{I_3} = 5$. The labels on the projected $x\prime$- and $y\prime$-axis show the separation from the central light bin in units of bin numbers. We note that each orbit consist of $N^3_{dither} = 5^3$ suborbits.

The second option is to sample observable space as uniformly as possible (Cappellari et al. 2006). In this work we generate initial conditions with this option. The code is written such that two sets of initial conditions (or start-spaces) are being generated, the (x,z)-start space and the stationary start space, following the method described by Schwarzschild (1993). We choose a number of energies N_{ener} for our models and sample the orbital energies through a logarithmic grid in radius. So, we specify a minimum, r_{min}, and maximum radius, r_{max} and space all N_{ener} radii logarithmically between r_{min} and r_{max}. The energies are defined by the potential at $(x, y, z) = (r_i, 0, 0)$ and are identical for both start spaces.

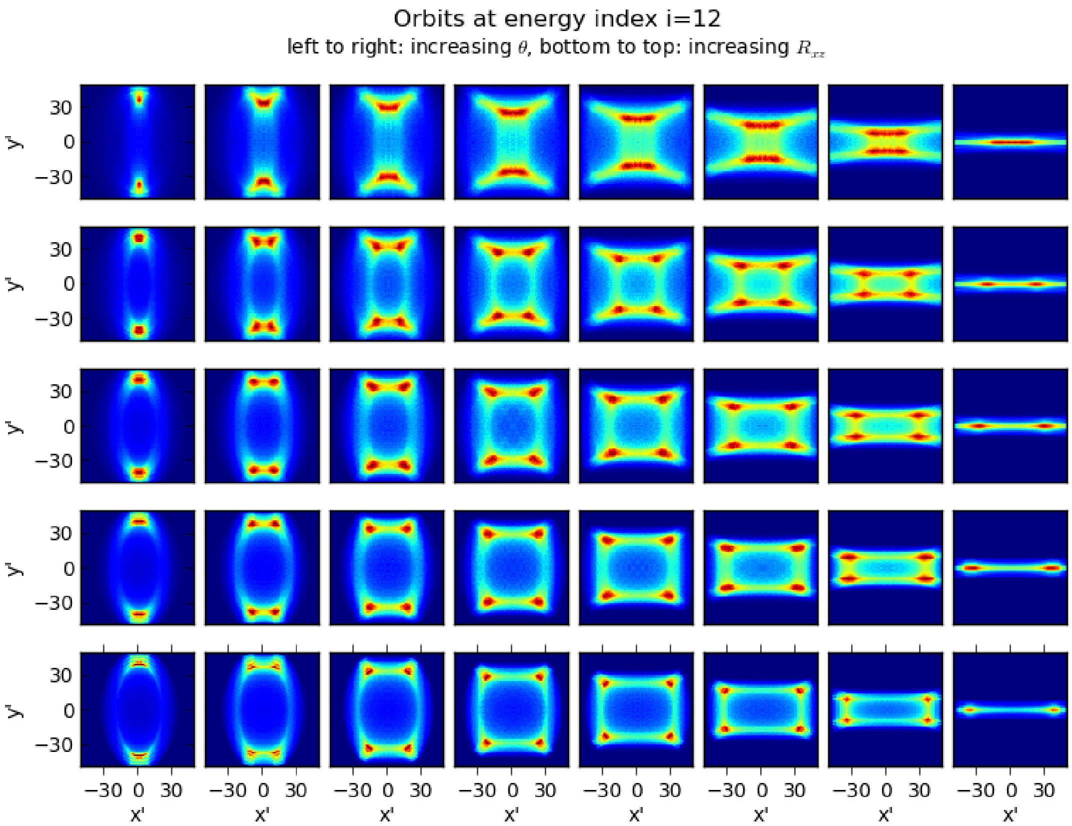


Figure 3.4: As in figure 3.3, but now having an energy index of 12. Higher energy orbits extend up to larger radii.

3.2.3 (x,z)-start space

Most orbits cross the (x,z)-plane perpendicularly twice at $z > 0$ (see section 3.2.2). To reduce duplications of orbits in our library we therefore first find the location of the thin orbit curves. At a fixed angle $\theta_0 = \arctan(x_0/z_0)$ stars are launched at different values of radii in the (x,z)-plane ($z = R_{xz,0}\cos(\theta_0)$, where $R_{xz,0} = \sqrt{x_0^2 + z_0^2}$) until the width of the orbit is minimal. This is repeated for different angles θ_0 and for all energies. In an axisymmetric potential there is only one thin tube curve per energy. We sample starting positions in a linear open polar grid in between the thin orbit curve and the equipotential. We choose N_{I_2} angles θ_0 and N_{I_3} radii. Therefore, this 'orbit library' will consist of $N_{orb} = N_{ener} \times N_{I_2} \times N_{I_3}$ orbits. The initial y_0-coordinate and initial velocities in the x- and z-direction are set to zero. The initial velocity in the y-direction is simply determined by: $\Psi(x_0, 0, z_0) - E = 0.5v_{y,0}^2$. We only sample initial conditions with $v_{y,0} > 0$ since the trajectories of the orbits with the opposite velocity direction will be identical. We include such counter-rotating orbits at a later stage, by reversing the signs of the velocity vector correctly (see section 3.5). Thus, effectively the single orbit library will be used twice. The set of initial conditions generated as explained above is called the '(x,z)-start space'.

In figure 3.3 we show the surface brightness profiles seen edge-on of strongly bound orbits for the q80v20 model with parameters $q = 0.8$ and $v_0 = 20$ km/s for the logarithmic potential, as described in section 4.3.3. This figure shows the different orbits for 8 different values for θ_0 and 5 different values for I_3 at energy index 9 (out of 20). The colors range from red to green to blue, where red colors denote a relative high fraction of time the orbit spent in that particular light bin and blue colors a low fraction of time. The left column corresponds to orbits with low values of θ_0 and arise from integrating initial conditions sampled near the minor axis of the system. θ_0 increases as we move to different columns towards the right. Thus, in last column, the orbits are sampled near the major axis of the system. In addition the bottom row of orbits are sampled near the location of the thin orbit ($R_{xz,0} \simeq R_{xz,thin}$). Successive rows are sampled at higher $R_{xz,0}$-radii, up

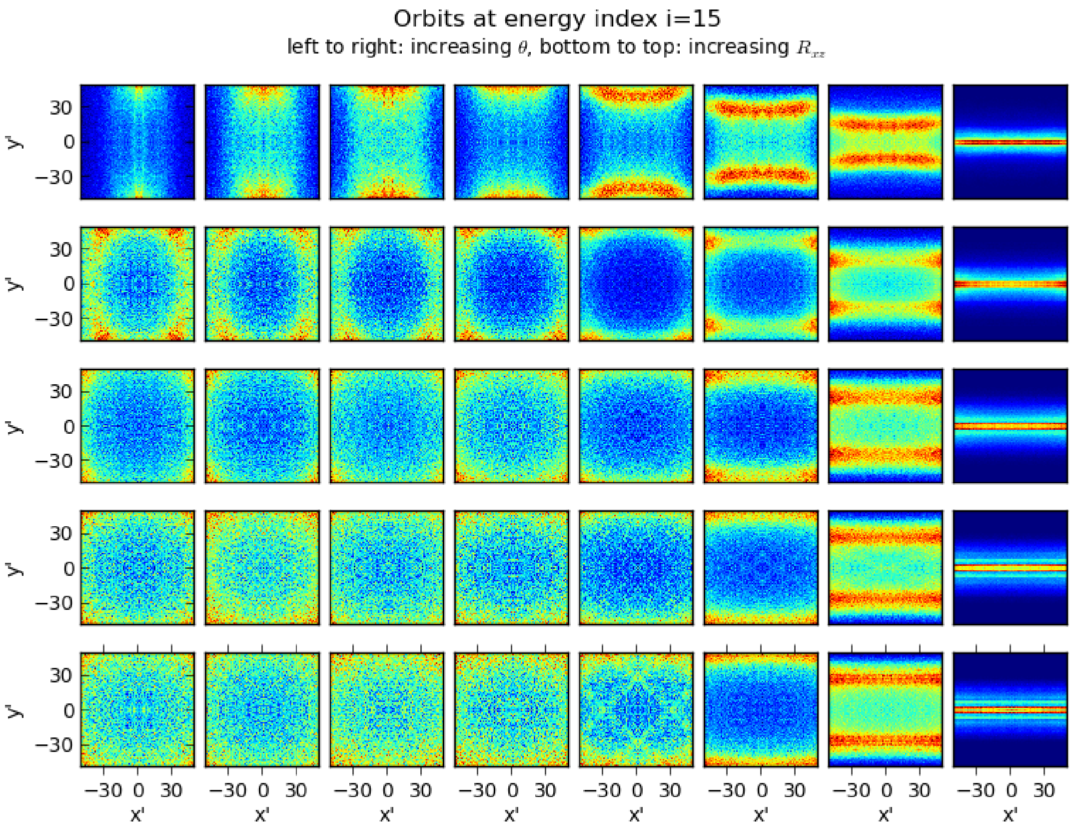


Figure 3.5: As in figure 3.3, but now having an energy index of 15. Higher energy orbits extend up to larger radii.

to an orbit near the $R_{xz,0}$-radius of the equipotential satisfying $E = \Psi$. With these definitions, the orbit that approximates a circular orbit the best is seen in the right bottom panel. In reality every orbit seen in the figure is the sum of $5^3 = 125$ neighbouring suborbits: this is called ‘dithering’ and will be explained in 3.3. In figures 3.4 and 3.5 similar sets of surface brightness profiles of less bound orbits are shown (energy indices 12 and 15 respectively). Less bound orbits extend up to larger radii.

3.2.4 Stationary start space

Since the Schwarzschild code is designed to be able to reproduce triaxial systems, which consist of a significant number of box orbits, the models need to include box orbits as well. The (x,z)-start space only has few box orbits (van den Bosch et al. 2008). To include additional box orbits, another ‘stationary start space’ is being made. Since box orbits always reach a point in which they have no velocity component anymore, i.e. they reach the equipotential $E = \Psi$ (Schwarzschild 1979), the initial conditions are chosen to lie on successive equipotential surfaces. For every energy, a 2D linear grid of spherical angles θ and ϕ determines the initial position r_0 on the equipotential. We specify N_{I_2} initial angles θ_0 and N_{I_3} initial angles ϕ_0. The ‘stationary start space’ contains (resonant) box orbits, but none of the tube orbits, such that it is a useful space in addition to the (x,z)-start space. Similar as in the ‘orbit library’, the ‘box orbit library’ consist of $N_{ener} \times N_{I_2} \times N_{I_3}$ orbits. Together with the 2 (mirrored) orbit libraries from the (x,z)-start space, the fitting routine will assign weights to all orbits from these three libraries.

In figure 3.6 we show the surface brightness profiles seen edge-on of the strongly bound box orbits for the q80v20 model with parameters $q = 0.8$ and $v_0 = 20$ km/s for the logarithmic potential, as described in section 4.3.3. This figure shows the different box orbits for 8 and 5 different initial values for the spherical angles θ_0 and ϕ_0 respectivly at energy index 9 (out of 20). As before the colors range from red to green to blue, where red colors denote a relative high fraction of time the

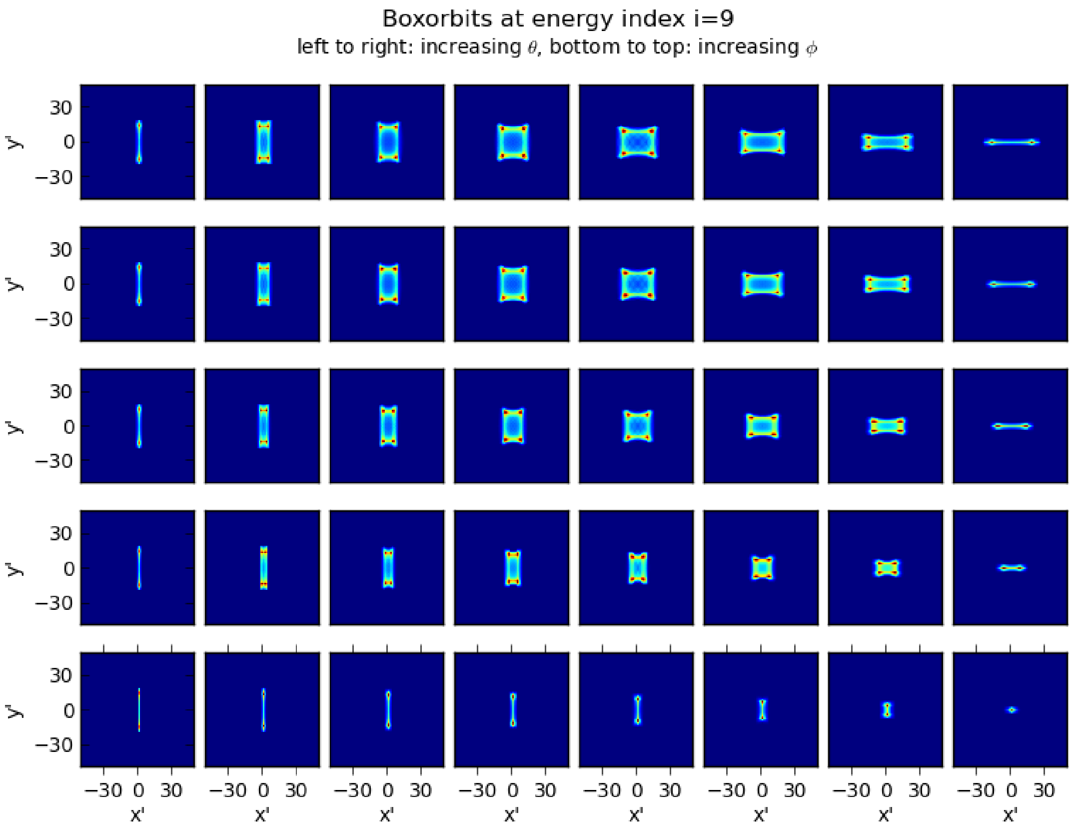


Figure 3.6: The surface brightness profiles of the individual box orbits in all 99x99 light bins of our field of view (3x3 kpc), seen edge-on, having an energy index of 9 in the library of q80v20 from attempt 3 in section 4.3 which consist of 20 different energies and $N_{I_2} = 8$ and $N_{I_3} = 5$. The labels on the projected $x\prime$- and $y\prime$- axis show the separation from the central light bin in units of bin numbers. We note that each orbit consist of $N^3_{dither} = 5^3$ suborbits.

box orbit spent in that particular light bin and blue colors a low fraction of time. From left to right the angle θ_0 is increased. From bottom to top angle ϕ_0 is increased. Since here dithering is used as well, every box orbit seen in the figure is the sum of $5^3 = 125$ neighbouring suborbits (see section 3.3). In figures 3.7 and 3.8 similar sets of surface brightness profiles of less bound box orbits are shown (energy indices 12 and 15 respectively). Less bound box orbits extend up to larger radii.

3.3 Integrator and orbital dithering

The intial conditions of both start spaces are integrated with a Runge Kutta integrator. To speed up orbital integration, the accelerations in x, y and z (see equations 3.4, 3.5 and 3.3 in the case of the logarithmic potential) are stored in a three-dimensional polar grid. During integration the accelerations are computed by the method of trilinear interpolation from this grid. The code ensures that the minimum relative accuracy of the interpolation grid is better than 10^{-4}. Each orbit is integrated for a duration of roughly 200 orbital time scales. We require that the energy of each orbit is always conserved better than 0.1%. The integrator uses dense output, which makes it possible to use more integration steps when the star changes direction quickly. We make sure that the stored properties (see section 3.4) of all orbits are based on equal time intervals, even if temporarily more steps are used.

van den Bosch et al. (2008) find that their models do not significantly improve when the total number of orbital weights ($3 \times N_{orb}$, see section 3.2.4) that need to be fitted is larger than 2000. When trying to recover the mock galaxy parameters, we used $N_{ener} = 32$, $N_{I_2} = 16$ and $N_{I_3} = 16$ to get a total of 3x32x16x16=24576 orbits. We found that choosing 3x20x8x5=2400 orbits was not

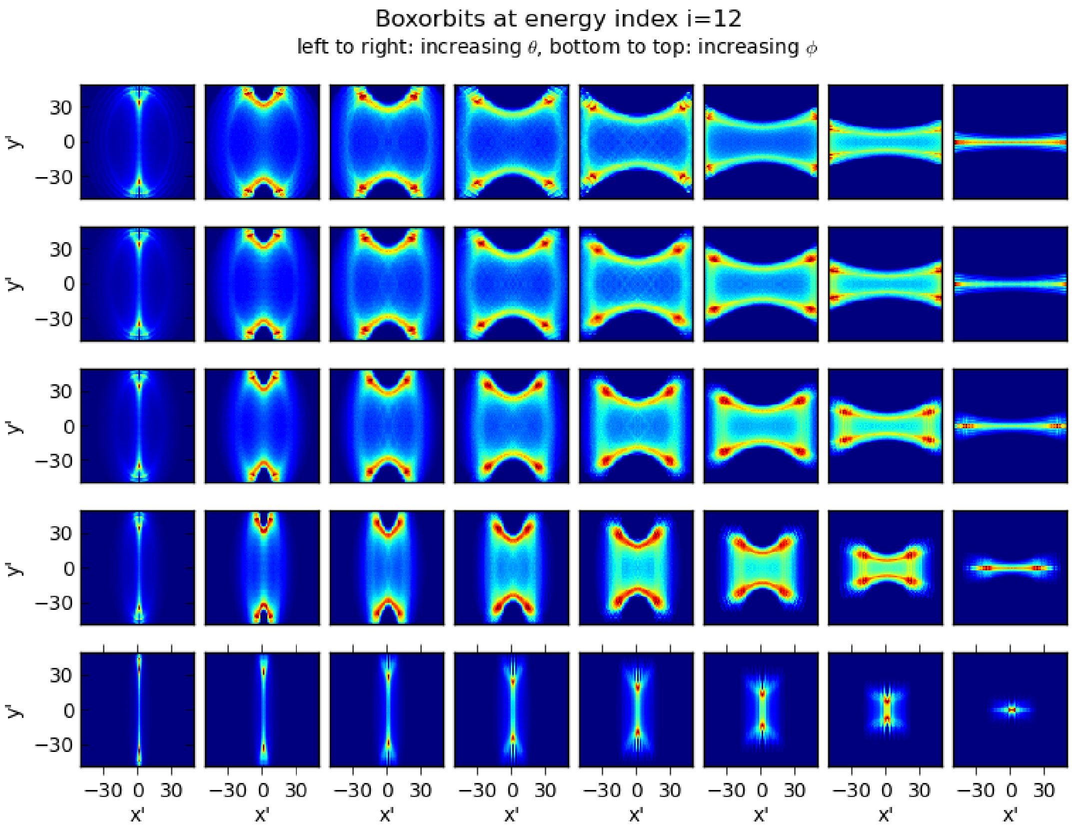


Figure 3.7: As in figure 3.6, but now having an energy index of 12. Higher energy box orbits extend up to larger radii.

good enough.

In stead of integrating N_{orb} orbits in each start space, we could increase our sampling to enlarge the accuracy of the model. In addition, to smooth the building blocks of the galaxy we are using the method of 'dithering'. This means that we split every orbit into N_{dither}^3 suborbits: each coordinate is replaced by N_{dither} adjacent coordinates. Choosing an odd number for N_{dither} ensures that the original orbit will be the central suborbit of the bundle. The observables of all suborbits will be summed and stored as being the observables of the (bundled) orbit. In all our results we used $N_{dither} = 5$. Every orbit is thus made from a bundle of $5^3 = 125$ neighbouring suborbits. For our choice of parameters, this means that we integrated 32x16x16x125 = 1024000 suborbits.

3.4 Storage grids and symmetries

3.4.1 Velocity histogram

For every orbit two histograms are being stored. The first histogram consist of a velocity axis and an axis containing sky bins. The velocity bins are specified by the user and are determined by three parameters: v_{width}, v_{cen} and N_v. We used $v_{width} = 80$ km/s, $v_{cen} = 0$ km/s and $N_v = 41$, such that 41 velocity bins are linearly spaced between -40 km/s and +40 km/s. We choose a velocity range of 80 km/s such that we cover a 4σ (in our mock galaxy model $\sigma = 10.7$ km/s) interval in velocity. We choose an odd number of velocity bins N_v, such that the central velocity bin is given by v_{cen}. A higher number of velocity bins could increase the accuracy of the velocity profile, but could also unnecessarily enlarge the disk space needed for storage. The sky bins are identical to the kinematic bins that were used to obtain the kinematic moments of our mock galaxy.

The program stores what fraction of time an orbit was positioned at a specific sky bin while having a velocity that corresponds to the velocity range of the specific velocity bin. In every

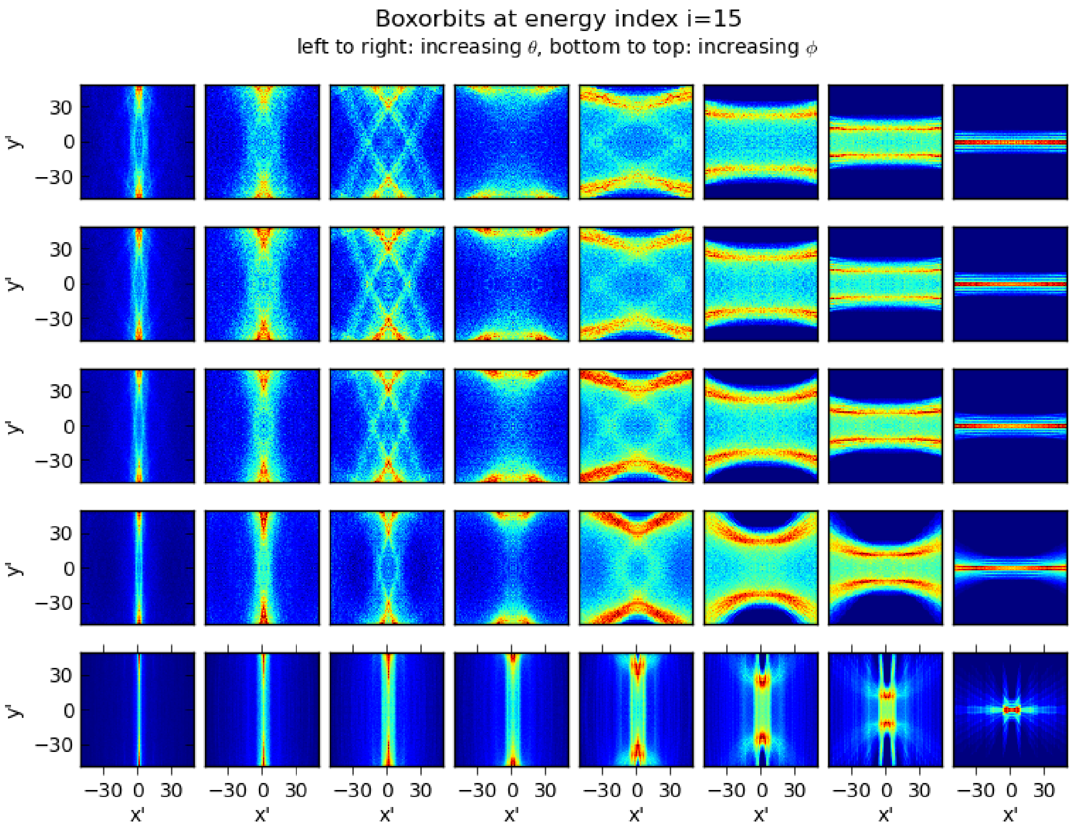


Figure 3.8: As in figure 3.6, but now having an energy index of 15. Higher energy box orbits extend up to larger radii.

time step an orbit has velocities exceeding the velocity range of the histogram, a count will be added in either the first or last velocity bin, depending on the sign of the velocity. Therefore the first velocity bin theoretically includes orbits that have velocities up to $-\infty$ km/s, and the last velocity bin includes orbits having velocities up to $+\infty$ km/s. Summing over the velocity axis of the histogram will give the fraction of time that an orbit has spent in a particular sky bin. For low energy (strongly bound) orbits that completely fall into the field of view, an extra sum over all sky bins will results in a fraction of time equal to unity. Higher energy (less bound) orbits, that partially spend their time outside the FOV, will have a smaller sum, while the velocity histograms of orbits that never enter the field of view will remain empty. We note that this is just a way of normalizing the histograms.

We verified that choosing a velocity width of 80 km/s and 41 velocity bins, does indeed give the correct properties of the velocity profile. We note that, in our case, in every kinematic bin the theoretical velocity profile is Gaussian ($\equiv G(v)$). However, the velocity histograms have finite width and are discrete. First of all, we can not use the outer velocity bins, since these contain velocities with a large range (up to $\pm\infty$ or up to $\pm v_{esc}$, the escape velocity). Secondly, suppose velocity bin k allows the contribution of an orbit that satisfies $v_{cen,k} - \widetilde{\triangle v} < v < v_{cen,k} + \widetilde{\triangle v}$, where $\widetilde{\triangle v}$ is the half velocity width of a velocity bin and where $v_{cen,k}$ is the central velocity in velocity bin k. However, when computing the kinematics from the histograms the only thing we know is what fraction of time the orbit spent in velocity bin k. Therefore the real velocity v is being replaced by $v_{cen,k}$. To check how large these effects are, we evaluate what the determined raw moments would be if the real line profiles are indeed Gaussian:

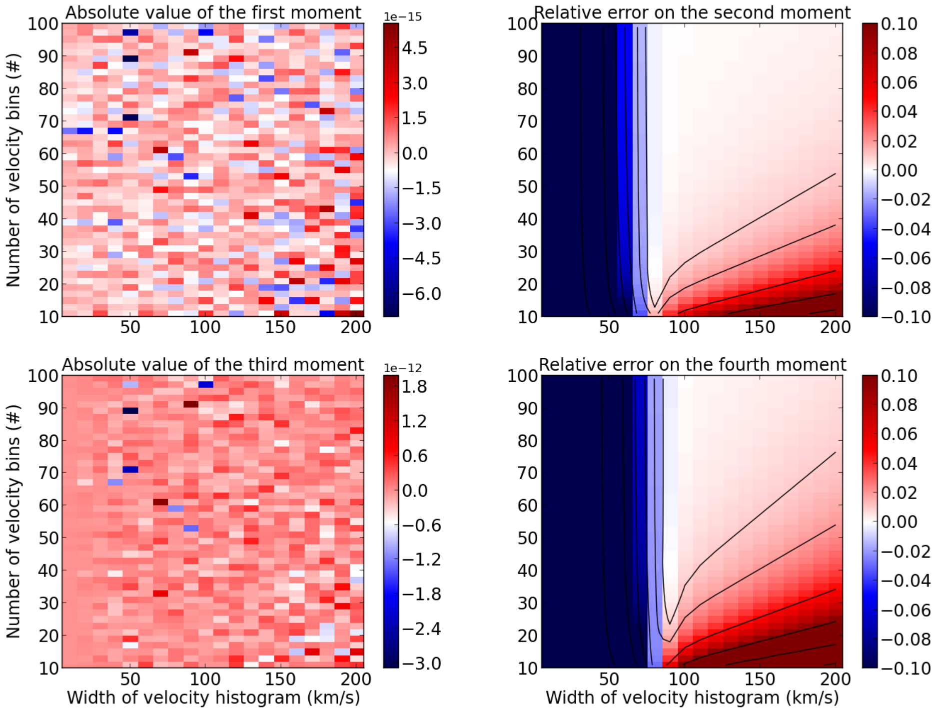


Figure 3.9: The errors in the velocity histogram on all four raw moments when choosing a certain combination of v_{width}(km/s) and N_v. The odd moments are not affected by the choice of parameters. The even moments are. For the odd moments we plot the absolute error of the library with respect to the theoretical value. For the even moments we plot the relative error. Negative (blue) values show that the library will underestimate the selected moment. Obviously, the best results are obtained when choosing both a high velocity width and a high number of velocity bins. A higher number of velocity bins will increase the storage space. Good results can already be obtained when using v_{width} = 80-90 km/s and $N_v > 20$.

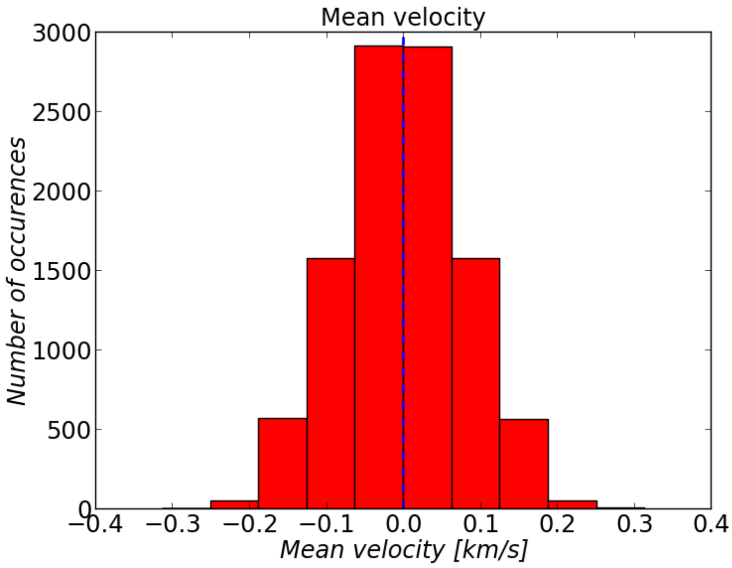


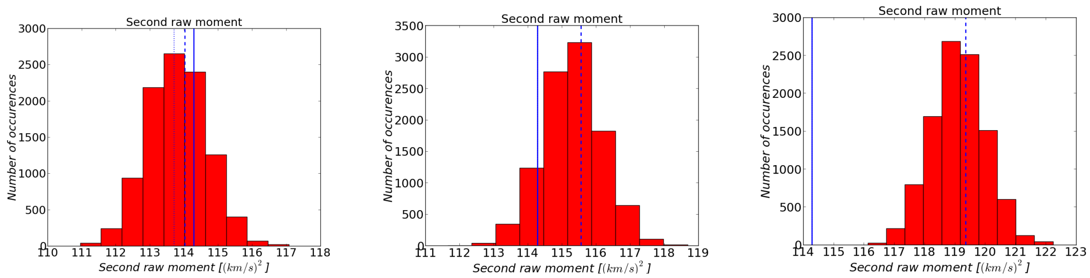
Figure 3.10: The first raw moments can always be obtained very accurate, whatever combination of number of velocity bins and velocity width you take. See figure 3.9 for more information.

$$m_n = \langle v^n \rangle = \frac{\displaystyle\sum_{k-2}^{N_v-1} \int_{v_{cen,k}-\widetilde{\triangle v}}^{v_{cen,k}+\widetilde{\triangle v}} G(v)\, v_{cen,k}^n \, \mathrm{d}v}{\displaystyle\sum_{k-2}^{N_v-1} \int_{v_{cen,k}-\widetilde{\triangle v}}^{v_{cen,k}+\widetilde{\triangle v}} G(v) \, \mathrm{d}v} \tag{3.6}$$

where we sum over all, except the outer, velocity bins k. In figure 3.9 the relative errors caused by this velocity binning only is shown. The first and third moment are not affected by the discreteness of the histogram, as confirmed in figure 3.10 where the recovered mean velocities are well distributed around the theoretical zero mean velocity. The second and fourth moment however show deviations when using $v_{width} < 70$ km/s or when a small number of velocity bins is chosen. We see that the best results can be obtained using a velocity width between 80 and 100 km/s and for at least 20 velocity bins. Due to this histogram binning the relative error on the second and fourth moment are of order 0.1% and 2% respectively for our choice of parameters. If we, for example, used a $v_{width} = 40$ km/s, the estimated second moment would have been 78.3 km^2/s^2 in stead of 114.3 km^2/s^2. In figures 3.10 and 3.11 we show the recovered first and second moment after using all our 100000 mock stars as initial conditions to make an ideal library (see section 3.6.1). In this way all orbits of the library are equally important. However, recovering the second moments (figure 3.11), by adding up all orbital contributions, show systematic offset if bad choices are made for the parameters of this velocity histogram. We confirmed that the first moment is not affected for the same choices.

3.4.2 Surface brightness histogram

A second histogram is introduced to fit the surface brightness profile of our mock galaxy. Since we will only use a relatively small number of kinematic sky bins, to keep a high S/N-ratio, we can not use the same few kinematic bins to constrain the flattening of the mock galaxy. To fit the surface brightness, we therefore use 'light bins'. Since we use an additional histogram for the light only, we do not need to store the line profiles of the orbits in this surface brightness histogram. In stead, we just store the fractional time an orbit spent in each of the N_{light} light bins.



(a) (b) (c)

Figure 3.11: Demonstrating the effect of choosing a wrong combination of v_{width}(km/s) and N_v when trying to model the raw second moment. Panel (a): Good results, v_{width} = 80 km/s and N_v = 41. The histogram shows the recovered raw second moments for all bins on the sky plane (here 101x101=10201) when making an ideal library in which the stars of the mock galaxies are taken as initial conditions (see section 3.6.1). The full line shows $\sigma^2 = 114.3$ km^2/s^2. The dashed line includes the effect on σ^2 when using a discrete histogram and fixed velocity width (see text). Panel (b) shows the same as panel (a), but now the velocity width of the histogram has been doubled. In panel (c) the velocity width is doubled again.

3.4.3 Octant grid

Both histograms are used in the fitting routine to find the weights of the orbits (section 3.5). Once the orbital weights are determined, the distribution function of the system is known and predictions can be made about the internal properties of the system. In order to do this, an octant grid is being made during the orbital integration. This spherical grid consists of 40 radial bins, 5 angular bins for θ and 4 angular bins for ϕ. The radial bins are logarithmically sampled between r_{min} and r_{max}, similar to what was done to determine the orbital energies that were used to make the libraries. In the octant grid, however, the first radial bin also inserts orbits that come arbitrarily close to the center, while the last radial bin extends to infinity. Both spherical angles are sampled linearly between 0 and $\pi/2$, such that an octant of space is being covered. At every equally spaced time interval we determine in what 3D-cell of the octant grid the orbit is located. If the orbit is currently not in the positive octant satisfying $x > 0$, $y > 0$ and $z > 0$, then by symmetry arguments, we compute what its coordinates and velocities would be if it was in this octant (see section 3.4.4). Then, the x-,y- and z-positions, the v_x-, v_y- and v_z-velocities and all second (mixed) moments are added to this cell. After the orbital integration is completed the mean of all quantities and the fractional time an orbit has spent in a certain cell is computed. This is done for every orbit separately. The octant grid can be used to predict the intrinsic properties of the mock galaxy as soon as the determination of the best-fit orbital weights is completed.

3.4.4 Symmetries

Here we summarize the most important information from the similar section in van den Bosch et al. (2008). All orbits in a separable potential are eightfold symmetric, whereas resonant and irregular orbits from other potentials might not be. In order to be able to use them, the orbits are made to satisfy the symmetries by applying a folding scheme. An asymmetric orbit has up to seven mirror images when reflecting it in the principal planes. Every point can be mirrored in x-, y- and z-coordinate. The resulting mirrored orbits would not have entered the library without this approach, since the initial conditions are sampled from one octant only. Thus, in the Schwarzschild code all eight mirrors are added to obtain an orbit that has three planes of symmetry.

In order to obtain the correct kinematic observables, we must also change the sign of the velocities. However, changing it in the same way as the positional coordinates would result in no net angular momentum for the resulting image. This is correct for box orbits, but can not be the end of the story for tube orbits. Tube orbits must preserve the sign of one component of the angular momentum. For example, the sign of $L_z = xv_y - yv_x$ must be preserved for Z-tubes. Therefore, when mirroring (x,y,z) into, let's say, ($-x$,y,z), the velocities need to be changed from

(v_x,v_y,v_z) towards (v_x,-v_y,v_z). The full symmetry relations can be found in van den Bosch et al. (2008).

3.5 Nonnegative least square fitting

After completing the orbit libraries, we can proceed with the fitting routine. We require that the sum of the orbital weights w_i equals unity, i.e.

$$\sum_{i=1}^{N_{orb}} w_i = 1 \tag{3.7}$$

By construction, this requirement is equivalent to fitting the fractional light of the system in our field of view. In addition, both the measured kinematics inside our kinematic bins and the surface brightness inside all light bins separately must be fitted by finding the best combination of orbits from the libraries. In contrast to van den Bosch et al. (2008) we do not use the surface brightness as a *constraint* only, since we model a composite system in which the luminous mass has a different flattening than the dark mass (see equations 2.2 and 2.5). We calculate the theoretical surface brightness (see equation 2.4) in each of our light bins and normalize this by the total luminous mass (see equation 2.13).

Since our FOV does not cover the whole of our extended mock galaxy, we only observe roughly 28-32% of the luminous mass, depending on the viewing angles (see section 2.3). The model returned by the fitting routine should fit the fractional light, compared to the system's total light, in each of the light bins. Thus:

$$m_j = \sum_{i=1}^{N_{orb}} w_i \, m_{ij} \tag{3.8}$$

where we sum over all orbits i, m_{ij} is the fraction of time orbit i spent in light bin j and where m_j is the 'observed' theoretical and fractional surface brightness in light bin j.

Simultaneously the kinematics are fitted. In every kinematic bin j we compute the mass-weighted raw moments $\langle v_j^n \rangle$:

$$m_j \langle v_j^n \rangle = \sum_{i=1}^{N_{orb}} w_i \, m_{ij} \langle v_{ij}^n \rangle \tag{3.9}$$

where again we sum over all orbits i. This time m_{ij} is the fraction of time orbit i spent in kinematic bin j and m_j is the 'observed' theoretical fractional surface brightness in kinematic bin j. $\langle v_{ij}^n \rangle$ is the raw n^{th} moment of orbit i in kinematic bin j:

$$\langle v_{ij}^n \rangle = \frac{\displaystyle\sum_{k=2}^{N_v-1} h_{ij}(v_{cen,k}) \, v_{cen,k}^n \, \triangle v}{\displaystyle\sum_{k=2}^{N_v-1} h_{ij}(v_{cen,k}) \, \triangle v} \tag{3.10}$$

where again we sum over all, except the outer, velocity bins k for every orbit i in kinematic bin j. $\triangle v$ is the full size of a velocity bin and $h_{ij}(v_{cen,k})$ is the fraction of time that orbit i spent in kinematic bin j with velocity v in the range $[v_{cen,k} - \widetilde{\triangle v} \ , \ v_{cen,k} + \widetilde{\triangle v}]$.

We use a non negative least square solver to ensure that all orbital weights are positive. The fit is based on minimizing $\chi^2_{\rm tot}$:

$$\chi^2_{\rm tot} = \sum_{m=1}^{N_{obs}} \left[\frac{\text{Model[m]} - \text{Data[m]}}{\text{Error[m]}} \right]^2 \tag{3.11}$$

where m runs over all N_{obs} observables. The number of observables is given by:

$$N_{obs} = 1 + N_{light} + 4N_{kin} \tag{3.12}$$

which includes the contribution of the total light inside our FOV (1), the surface brightness inside all light bins (99x99), and 4 times (four moments) the number of kinematic bins (either 9x9 or 3x3).

We note that we can investigate the contribution of all terms to the total $\chi^2_{\rm tot}$ by writing:

$$\begin{aligned}\chi^2_{\rm tot} &= \chi^2_{\rm light} + \chi^2_{\rm kin} + \chi^2_{\rm FOV\ light} \\ &= \chi^2_{\rm light} + \chi^2_{\rm mom1} + \chi^2_{\rm mom2} + \chi^2_{\rm mom3} + \chi^2_{\rm mom4} + \chi^2_{\rm FOV\ light}\end{aligned} \tag{3.13}$$

The fitter minimizes: $|\boldsymbol{Ax} - \boldsymbol{b}|^2$, where $\boldsymbol{A}$ is a (m x n)-matrix in which all m=N_{obs} properties of all n=N_{orb} orbits are stored. Vector $\boldsymbol{x}$ is a (n x 1)-matrix, that will contain the best fit orbital weights. Vector $\boldsymbol{b}$ is the (m x 1)-matrix containing all corresponding observables.

3.5.1 Observables

The estimators of the kinematic moments are described in section 2.4.1. We estimate the corresponding errors of these moments by assuming a Gaussian error of 2 km/s on the line-of-sight velocity measurements of the stars, as described in section 2.4.2. The errors on the light in each of the light bins are set to 2% as in van den Bosch et al. (2008).

3.5.2 Regularization

The solution of our minimization problem may be a distribution for the orbital weights that is rapidly varying, essentially a sum of delta-functions. The orbital distribution function can be smoothened by adding extra terms to the χ^2-fitting algorithm, such that:

$$\widetilde{\chi^2_{\rm tot}} = \chi^2_{\rm tot} + \chi^2_{\rm reg} \tag{3.14}$$

This is called regularization. Regularization should not change the solution of recovered best-fit parameters of the galaxy. Adding regularization does cause the confidence intervals of the parameters to become smaller, since it decreases the freedom of the models (van den Bosch et al. 2008). In this work we do *not* add regularization terms in our fitting routine.

3.6 Testing the integrator and fitting routine

In chapter 4 we will check whether our method is able to recover the true parameters of our mock galaxy given the true potential functional form. Then, in chapter 5 we search for the best-fit parameters for more commonly used potentials, because, in reality, the gravitational potential functional form in a system like the Sculptor dSph is unknown. Before we can actually use the method to recover characteristic parameters of the modeled potentials, we perform tests to check whether the integrator and fitting routine work fine when using an ideal library.

3.6.1 An ideal orbit library

We tested the Schwarzschild code by integrating the positions and velocities of all 10^5 stars in our mock galaxy as initial conditions for the orbits. In this way we construct an ideal library, because these are drawn from the true distribution function of the system. This implies that all orbits should be represented in our model with equal weights. As we used $N_{dither} = 5$, $10^5/5^3 = 800$ orbits are stored in this library.

The first test we performed was to give by hand all orbits of this ideal library equal weights. We found that this recovers the properties of our mock galaxy, such as the light distribution and the kinematic moments, showing that the potential was correctly added to the code and that the orbits were well integrated. In figure 3.12 we show the values for the kinematic moments recovered for all 101x101 sky bins, that were used for both the kinematics and the light in making the ideal library, on the sky. The median of the first three moments agree with the expected values, while the fourth moment is predicted too low. We explain this by the fact the velocity histograms that were used to compute all moments, are binned and can cause deviations. In section 3.4.1 we show

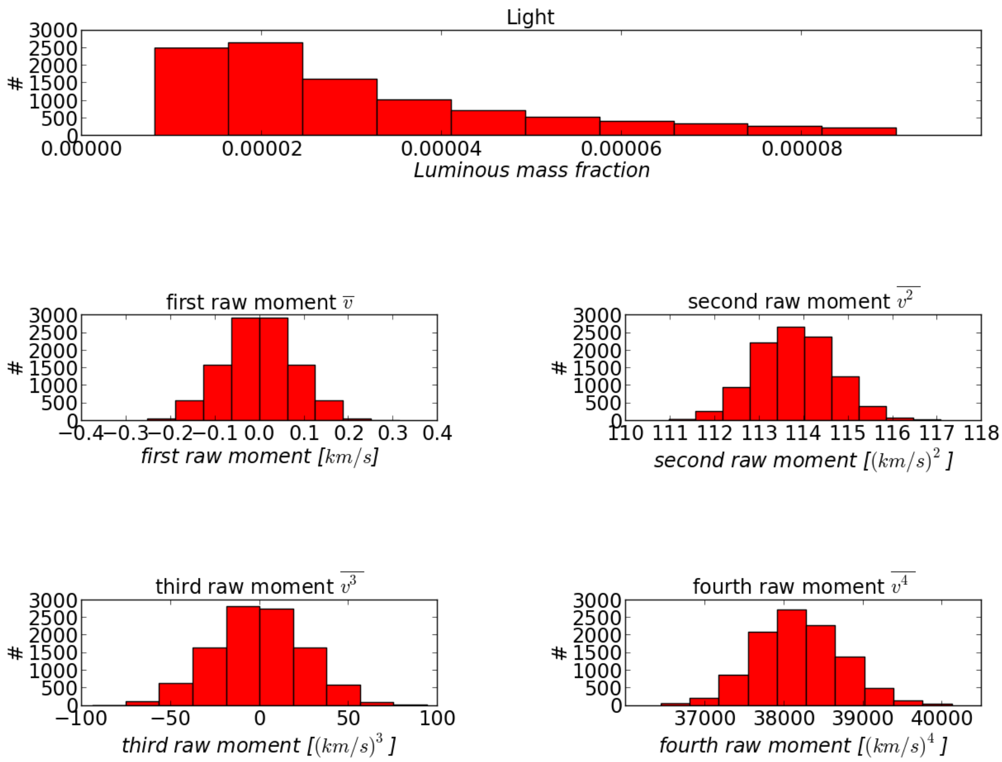


Figure 3.12: Histograms showing how well the observables are recovered when all orbits in our ideal orbit library are given equal weights. The odd moments show no deviation from zero. The second moment roughly has its mean at $\sigma^2 = 114.3$ (km/s)^2. The fourth moment, which should be $3\sigma^4 = 39184$ (km/s)^4, is underestimated by roughly 2%, which matches the predictions made from figure 3.9.

that the fourth moment would be underestimated by roughly 2% by our choice of parameter v_{width} and N_v.

The second test we performed was to use the fitting routine to find the orbital weights. As input to the fitting routine we used the surface brightness profile $I(x\prime, y\prime)$ given by equation 2.4 and the theoretical properties for the kinematics, as described in section 2.1. Panel (a) of figure 3.13 shows orbits were roughly given equal weights as a best fit, implying that the correct distribution function was recovered. In panel (b) we show the isodensity contours for both the model (full contours) and our mock galaxy (dashed countours following equation 2.4). The recovered variances from the line-of-sight profiles in all bins of our field of view are shown in panel (c). In panel (d) we show how the octant grid (see section 3.4.3) can be used to investigate the internal second moment in direction parallel to the major axis of the galaxy. The median shows that the octant grid does recover this second moment quite well, although it shows a wide range of values.

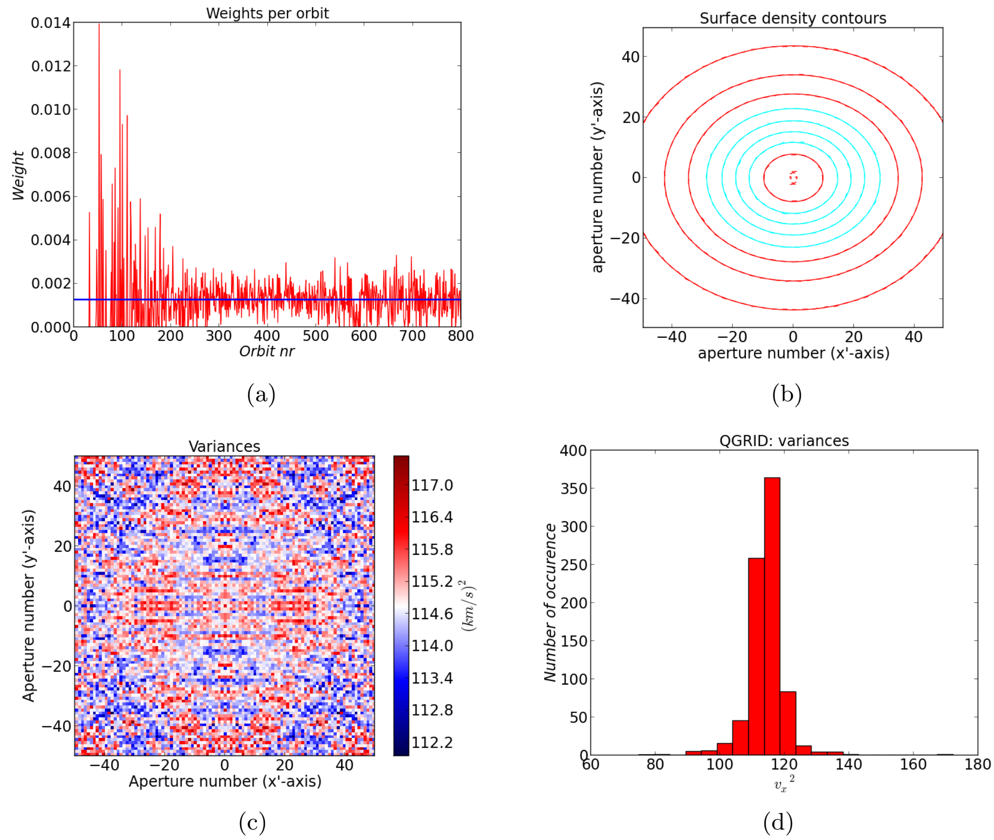


Figure 3.13: Some of recovered properties of our mock galaxy after fitting our ideal library to our mock data. All orbital weights are roughly equal to $1/800 = 0.00125$ (panel (a)). In panel (b) we show that the same isodensity contours for both the model and our mock galaxy overlap. In panel (c) and (d) we see that both the variance for the radial velocities in all 101x101 sky bins and for all intrinsic 3d-elements have values around σ^2.

Chapter 4

Recovering the mock galaxy parameters and properties

In section 3.6 we used an ideal orbit library, based on the true distribution function of the mock galaxy, and showed that our Schwarzschild model does recover its properties (light + kinematics). In reality we do not know which is the correct potential and want to find it with our Schwarzschild method. In order to test whether our model can recover the characteristic parameters q and v_0 of the true potential, we make a grid of models in which we vary the values of these parameters. This allows us to check whether the correct underlying model gives the best fit. This is indeed the case, but we note that one needs to be careful when making choices about how the models should be built. We therefore start this chapter by showing how we recovered the mock galaxy parameters. Later we elaborate on the wrong choices we made for the various models, such that, even if the correct underlying potential was chosen for the model, the results were not satisfactory.

4.1 High resolution models

We make high resolution models after setting $N_{ener} = 32$, $N_{I_2} = 16$ and $N_{I_3} = 16$. For the dithering we used $N_{dither} = 5$, such that both the orbit and the box orbit library consist of 32x16x16x5^3 = 1024000 suborbits and thus of 32x16x16=8192 independent orbits. The fitting routine makes a mirrored copy of the orbits and is therefore fitting 3x8192=24576 weights. This number is much larger than 2000 above which van den Bosch et al. (2008) found no significant improvement in their fits. We note that using roughly one-tenth as many orbits does not result in good models in our case (see section 4.3). The lowest energy orbits are set by $r_{min} = 10^{1.81} = 64.5$ arcsec along the major axis, which corresponds to 25 pc at a distance of 80 kpc. The highest energy orbits are set by $r_{max} = 10^{5.11} = 1.29\text{x}10^5$ arcsec, which corresponds to 50 kpc.

We decided to fix the viewing angles to an edge-on view and the core radius to $R_c = 1$ kpc, the true value. In our grid of input characteristic parameters used to build the orbit libraries, we therefore vary the flattening parameter q and mass parameter v_0. Since our mock galaxy has been made for $q = 0.8$ and $v_0 = 20$ km/s we sample the grid around these values from 0.72 to 0.96 in flattening and from 11 km/s to 29 km/s for v_0. We chose $\triangle q = 0.04$ and $\triangle v_0 = 3$ km/s. We name the models by the values of their parameters: qXXvYY in which XX = q100 $\equiv 100q$ and YY = v_0 in km/s. We make mock observables following the description in section 2.4. Since we do not want to have errors on the moments which are comparable to the moments themselves, we chose to sample 9x9 bins on the sky when pretending to observe 10^5 stars and 3x3 bins when pretending to observe 10^4 stars (see section 2.4.3). Using the high resolution models we found that, in order to extract information from the light profile, taking 45x45 light bins with 2% error is too few. Therefore, we set the number of light bins to 99x99. We kept the error of 2% in each light bin.

For each of the models using a 2.2Ghz AMD Opteron(tm) Processor 6174, generating the initial conditions for both the orbits and the box orbits takes 2-3 hours. Making one orbit library (out of two) takes 3-4 days. The libraries each take roughly 2.5-3.0 GB disk space. The fitting routine needed 1-3 days to compute the best fit orbital weights, while using 99x99 light bins, either 9x9 or 3x3 kinematic bins for using 10^5 or 10^4 stars respectively and while not adding regularization.

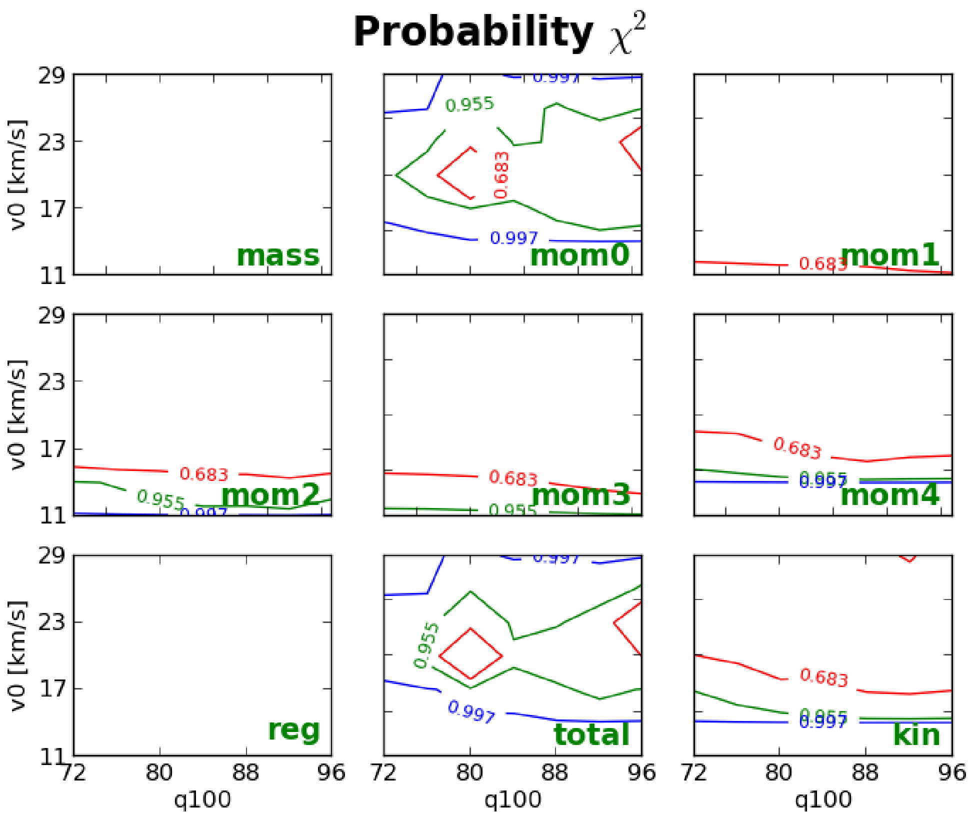


Figure 4.1: Contours of constant χ^2 computed after fitting our mock data consisting of 10^5 stars inside our field of view containing 9x9 kinematic bins, in a known edge-on view. Each of the nine subplots show the corresponding 1-,2- and 3-sigma probability contours around the best fit model on the grid. From top left to bottom right: fitting total mass, light, first raw moment, second raw moment, third raw moment, fourth raw moment, regularization terms, all constraints (total), all raw moments (kinematics). We note that all terms are not fitted independently from eachother: χ^2_{tot} is being minimized. The term concerning the light ($\chi^2_{mom0} \equiv \chi^2_{light}$) is dominating the shape of the contours in the total fit (χ^2_{tot}).

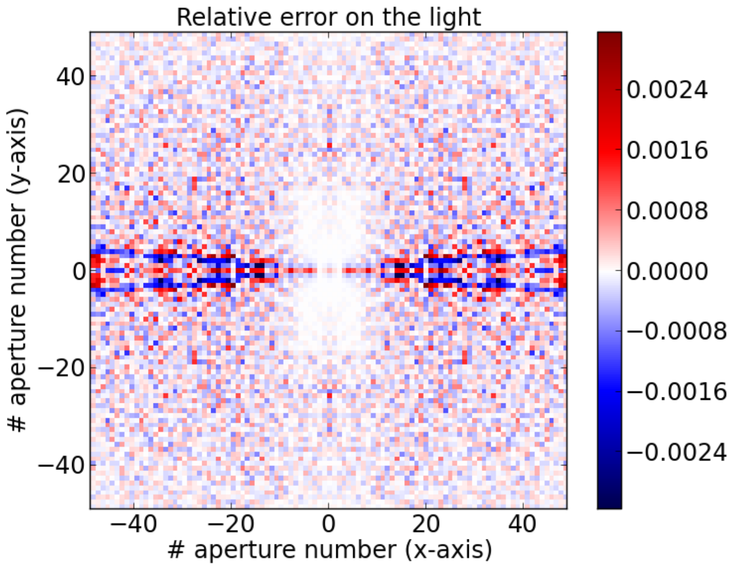
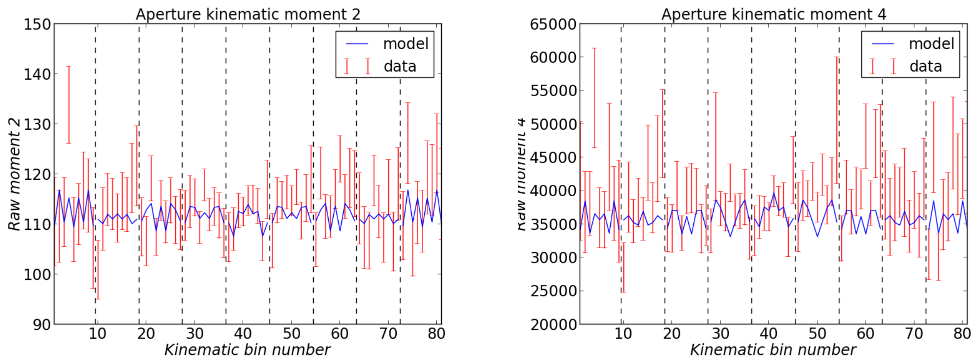


Figure 4.2: The light distribution after fitting the high resolution q80v20 library to our mock data consisting of 10^5 stars in our field of view, assuming an edge-on view. The light is fitted in 99x99 bins, the kinematics in 9x9 bins. We show the relative error on the surface brightness profile. The relative errors on the light are lower than in those of the q72v11 and q96v29 models seen in figures 4.4 and 4.6.



(a) (b)

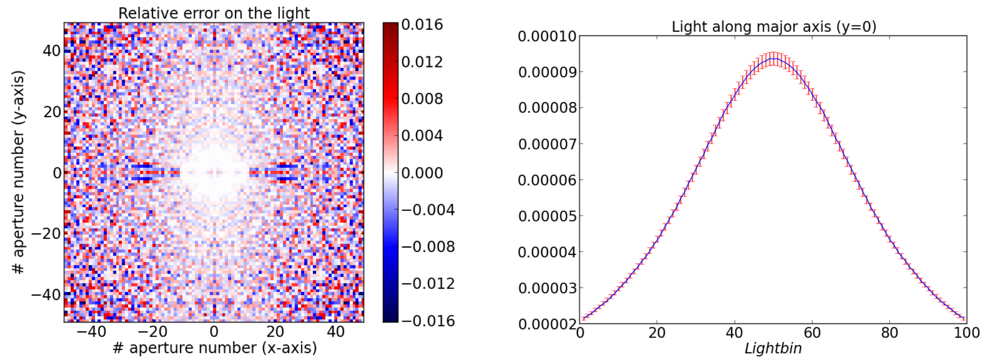
Figure 4.3: As in figure 4.2, but now showing the fitted second (a) and fourth (b) velocity moment.

The fitting routine needs 1.9 GB memory usage.

Figure 4.1 shows the results after 'observing' 10^5 stars edge-on, while using 9x9 kinematic bins on the sky. Colored contours of constant χ^2, corresponding to 1σ-, 2σ- and 3σ-confidence intervals (red, green and blue respectively), are shown in each of the panels. If shown, light grey contours show either $\triangle\chi^2 = 1.0$ or $\triangle\chi^2 = 0.1$, where:

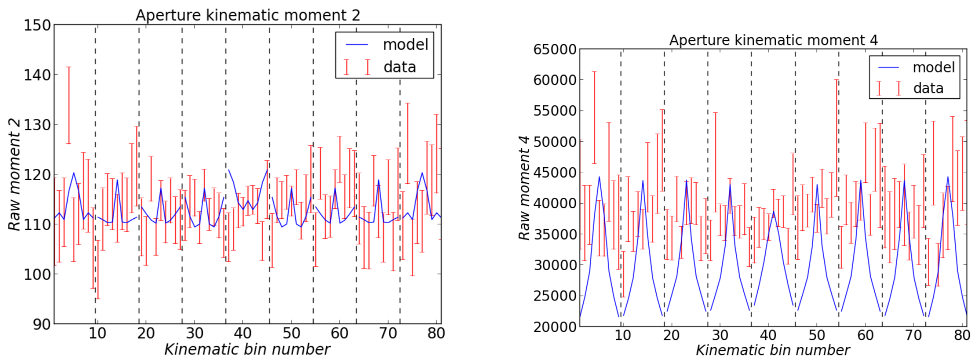
$$\triangle\chi^2 \equiv \triangle\chi^2_{ij,k} = \chi^2_{ij,k} - \min(\chi^2_k) \tag{4.1}$$

where $\chi^2_{ij,k}$ corresponds to the value of χ^2 for model $q = i$ and $v_0 = j$ and where $\min(\chi^2_k)$ denotes the χ^2-value for the best fit model. The subscript k emphasizes that the χ^2-values are computed for the observables in panel k only. For two degrees of freedom the 1σ-, 2σ- and 3σ-confidence intervals are defined by $\triangle\chi^2$ equal to 2.30, 6.17 and 11.8 respectively (Press et al. 1992). The panels show, from top left to bottom right, the decomposition of χ^2_{tot} in the χ^2-terms corresponding to the total light (mass), surface brightness (mom0), all kinematics moments separately (mom1 to mom4), regularization (reg), total fit (total) and all kinematic moments together (kin), see equation



(a) (b)

Figure 4.4: As in figure 4.2 in panel (a), but now showing the light distribution after fitting the high resolution q72v11 library to our mock data. In panel (b) we show the light fit along the major axis of the system.

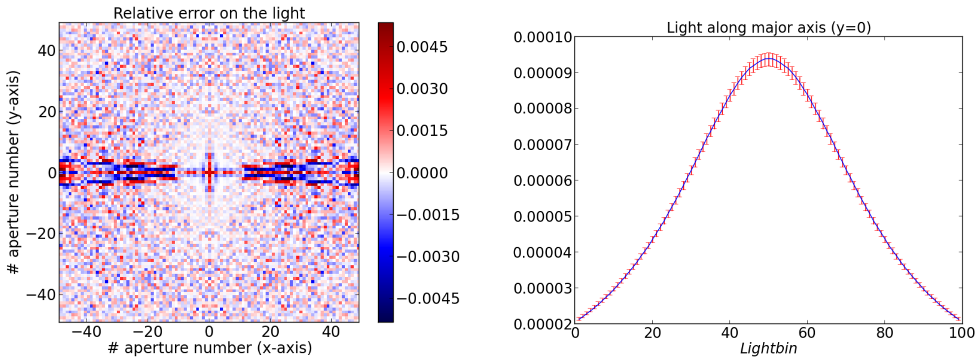


(a) (b)

Figure 4.5: As in figure 4.4, but now showing the fitted second (a) and fourth (b) velocity moment.

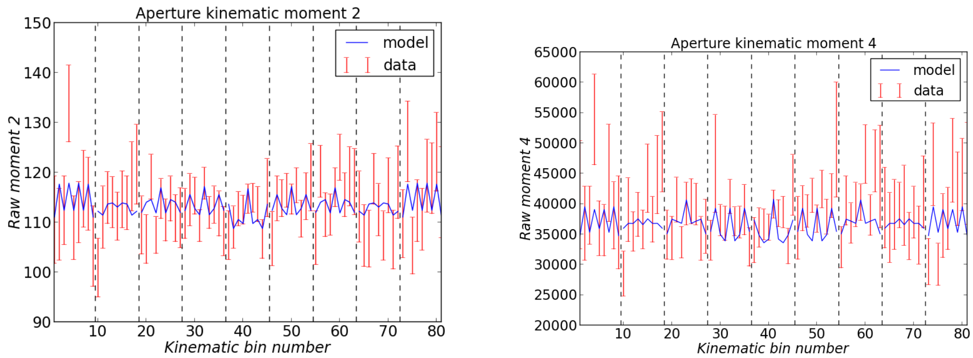
3.13 and 3.14. Therefore, it can be seen that the term χ^2_{light} (surface brightness) is the dominant term in χ^2_{tot}. The total light in our field of view (top left panel) is not significantly fit better for any of the models and that is why no contours are visible in this panel. In addition, only for low mass models, the kinematic terms are significantly worse. We note that the fitter minimizes χ^2_{tot} and that, because it is not minimizing all χ^2-terms separately, it is not necessarily true that the best fit model arose because of fitting the light only (as one might guess from comparing the contours). When we downsample the number of stars this becomes clear, as the contours span a larger area although the to be fitted surface brightness profile is identical (see section 4.2 and figure 4.8). Since we did not use regularization, that panel (bottom left) remains empty. From the bottom middle panel, showing the total fit, we conclude that we can recover the true parameters of our mock galaxy, although the flattening parameter is not constrained very well within the 2σ-boundary. For datasets with 10^5 stars, seen edge-on, we find $q = 0.8^{+0.04}_{-0.04}$ and $v_0 = 20^{+3}_{-3}$ km/s, where the upper and lower bounds are set by a 1σ-confidence interval, although we can also not rule out a model with $q = 0.96$ and $v_0 = 23$ km/s. For the best-fit model q80v20 we find $\chi^2_{red} \equiv \chi^2_{tot}/N_{total} = 0.0209$, where χ^2_{tot} and N_{total} are given by equation 3.11 and 3.12. We note that this reduced-χ^2 is much smaller than unity, because of the fact that the light profile is fitted very well and because the number of light bins dominates the term N_{total} for our choices.

In figure 4.3 we show how the second and fourth moment are fitted in the best fit and true q80v20 model. The modeled moments are shown by the full blue lines and the observed moments with errorbars are shown by the red data points. The 9x9 kinematic bins are aligned into this 1-dimensional plot by simply running over all 81 kinematic bins from the bottom left to the top right kinematic bin in our field of view. Therefore, kinematic bin numbers 36 to 45 show the fit



(a) (b)

Figure 4.6: As in figure 4.2 in panel (a), but now showing the light distribution after fitting the high resolution q96v29 library to our mock data. In panel (b) we show the light fit along the major axis of the system.



(a) (b)

Figure 4.7: As in figure 4.6, but now showing the fitted second (a) and fourth (b) velocity moment.

along the major axis of the system. Note that the dashed lines visualize the boundary of our field of view and that blue lines are thus not connected at these boundaries. As the model is symmetric in the even moments the modeled moments in each row of 9 kinematic bins on the sky are symmetric around its central kinematic bin (for example kinematic bins 1 and 9). With the same argument, all fits, except fits of the central $z = 0$ row, appear twice. The modeled even moments in kinematic bin numbers 1-9 are therefore identical to them in kinematic bin numbers 73-81. In figures 4.5 and 4.7 we show the same results, but for the q72v11 and q96v29 model respectively. A bad kinematic fit in the fourth moment is clearly visible for the q72v11 model.

In figures 4.2, 4.4 and 4.6 we show how the light is fitted for the q80v20, q72v11 and q96v29 model. In panels (a) the relative error on the surface brightness of the model with respect to the 'observed' surface brightness (see section 3.5) in each light bin is shown. Note the small relative errors, but that the q80v20 model has the smallest range in relative errors. In panels (b) the modeled surface brightness profile (blue full line) together with the 'observed' data with 2% error bars along the major axis are plotted, showing that all models fit the light quite well. Thus, although the surface brightness profile is fitted well in all models, there is still a significant difference between the models when combining all N_{light} light bins.

4.2 Downsampling and folding data

Assuming an edge-on view and using 10^4 stars in the field of view and 3x3 kinematic bins, we see in figure 4.8 that the uncertainties in the parameter ranges become large. The best fit is still found

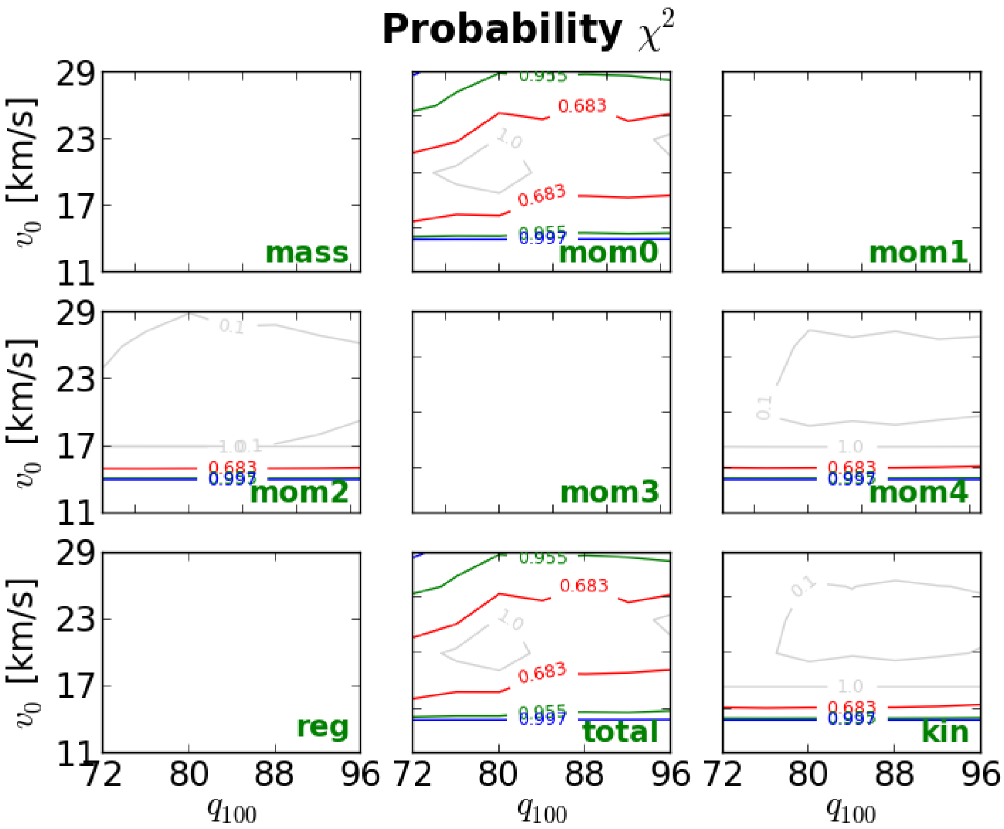


Figure 4.8: Similar to figure 4.1, but now using 10^4 stars and 3x3 kinematic bins. Both flattening parameter q and mass parameter v_0 are not constrained very well anymore.

at the q80v20 model, as seen by the grey contours that satisfy $\triangle\chi^2 = 1.0$ with respect to the best fit model, although the 1σ- and 2σ-contours show that many other models can not be ruled out as they do not appear to produce significantly worse fits.

To confine the best-fit contours to a smaller region, we started to fold the data in the case of 10^4 stars. Since we have an axisymmetric system, we could fold our data into the kinematic bins corresponding to the positive semi-major axis, the positive semi-minor axis and the remaining kinematic bins of the positive quadrant. Assuming that the system is not rotating, which is true in our case, we can simply move the stars towards the analogous kinematic bins without changing the velocity signs of the stars. Using 3x3 kinematic bins, this means that the kinematic bins in the corners of our field of view will be combined, as well as the two outer kinematic bins along the major and minor axis. The central kinematic bin is not combined with any other kinematic bin and we end up with 4 kinematic bins in which we now included all 10^4 stars. Statistically, the raw moment estimators will be closer to the real raw moments. The errors on the moments become smaller as the number of stars per kinematic bin has increased, but on the other hand the number of kinematic bins to be fitted decreases. We compute the raw moments and their errors in these 4 bins, but as our program still expects an input having 3x3 kinematic bins, we copy the kinematics back to the corresponding other 5 kinematic bins. Since these terms must not contribute anymore to the value of χ^2, we have increased the errors in their kinematic moments by a factor of 10^5. Because of symmetry in the model on the field of view, changing the orbit weights will affect the fits in all corresponding kinematic bins in the same way. Therefore, the kinematic bins having the non-increased errors will always be the dominant term in χ^2 compared to them having the large errors. In figure 4.9 we show the resulting χ^2 contours. Using this approach we find no significant difference with respect to the approach in which we do not fold the data (see figure 4.8).

In stead of folding the data when using 3x3 bins, we also folded the data using 9x9 bins. The mean number of stars in each bin roughly increase in by a factor of $9^2/5^2 = 3.24$ (with respect

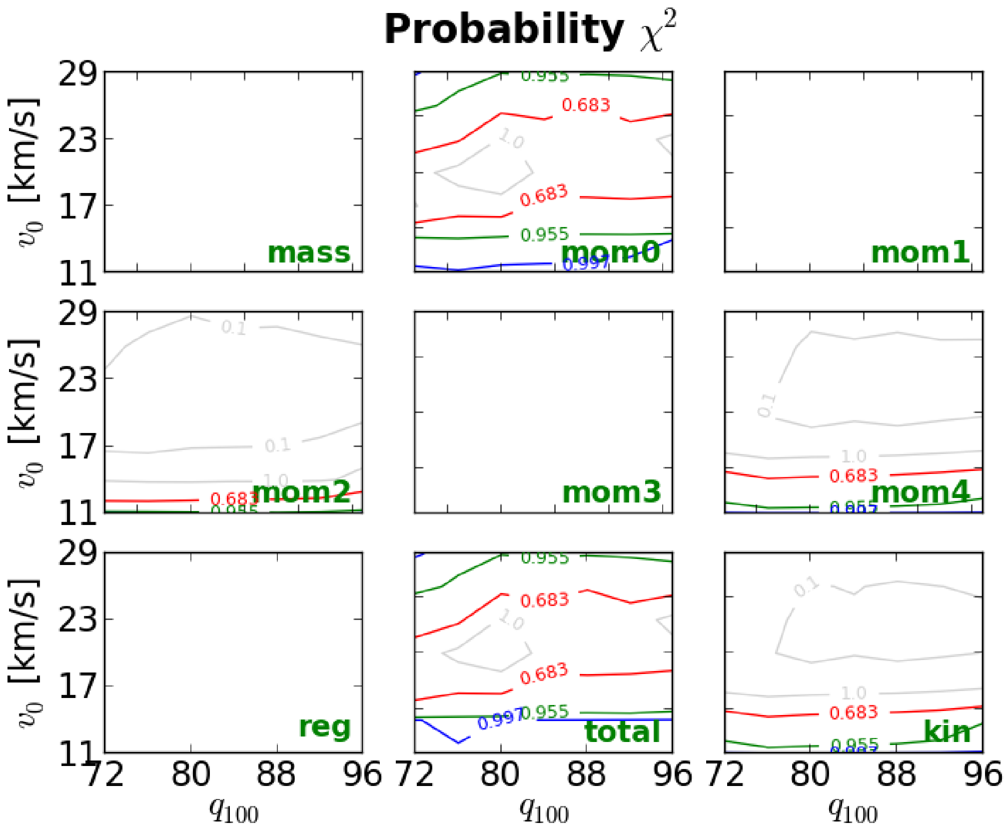


Figure 4.9: Similar to figure 4.8, but now using the approach of folding the data from 3x3 into 2x2 kinematic bins. The probability contours are similar to the case without folding (see figure 4.8).

to not folding the data) to $\frac{10^4}{25} = 400$. This number is lower than 4 since the number of stars in the central bin remains the same and the number of stars in the kinematic bins along the positive semi-major axis and the positive semi-minor axis are roughly doubled. The errors on the velocity moments will be still higher than in the case of using 10^5 stars, in which the mean number of stars equals $\frac{10^5}{81} = 1235$, but the difference is not that large anymore (factor of ~ 3). We argue that using more and therefore smaller kinematic bins might put stronger constraints on the best-fit model. The price being paid are the larger errors compared to folding with 3x3 kinematic bins. Folding the data in this way, we find that the regions spanned by the 1σ-, 2σ- and 3σ-contours do get smaller (see figure 4.10) and that the correct characteristic parameters of the potential can be recovered within a 1σ-confidence interval when observing 10^4 stars in an edge-on view. We find that $q = 0.80^{+0.04}_{-0.04}$ and $v_0 = 20^{+3}_{-3}$ km/s, within the 1σ-confidence interval, although we can also not rule out a model with $q = 0.96$ and $v_0 = 23$ km/s and that $\chi^2_{red} = 0.00221$ for the best-fit model q80v20. As expected, the constraints that can be put on these parameters are still less strong than in the case of 10^5 stars (see figure 4.1), as the 2σ-confidence interval now span over a larger area.

4.3 Rejecting bad models

Provided that the models fit the light profile in a sufficient number of bins and that the libraries itself consist of a sufficient number of orbits, our models fit the data very well and recover the mock galaxy parameters. We will now elaborate on the choices that we have made in order to converge to those good models and why not meeting these requirements will not result in reliable models.

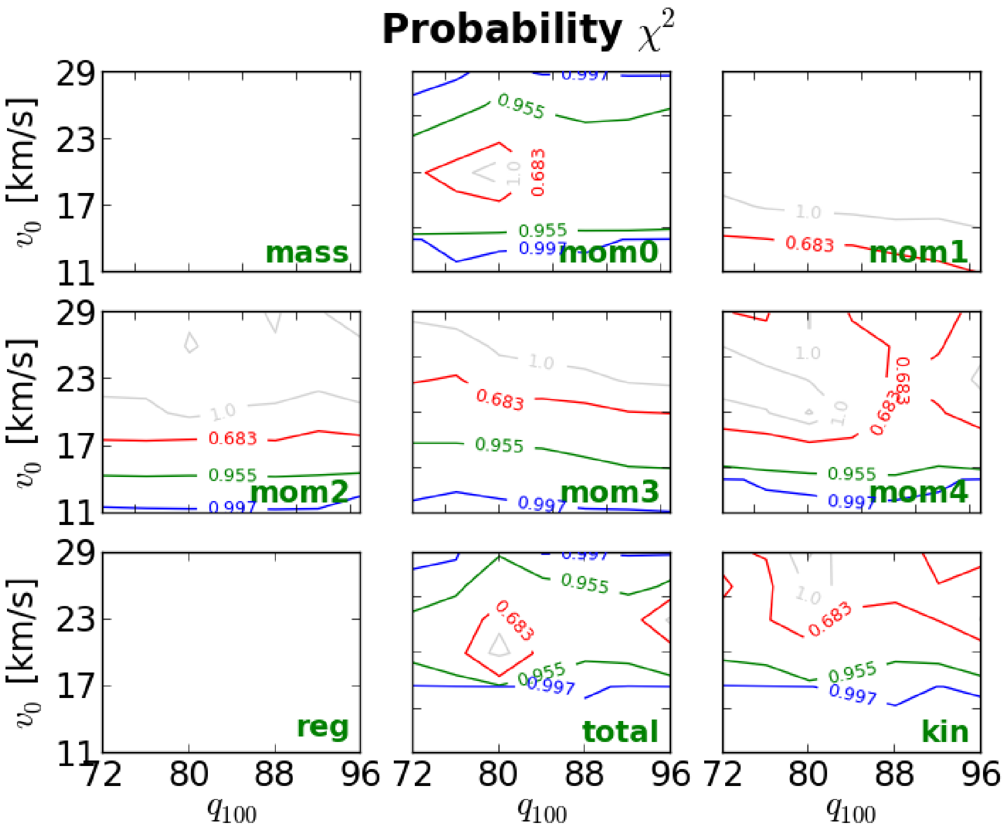


Figure 4.10: Similar to figure 4.9, but now using the approach of folding the data from 9x9 into 5x5 kinematic bins using 10^4 stars.

4.3.1 Attempt 1: Too small a number of light bins (9x9)

Our first guess was to integrate 10^5 suborbits and 10^5 box suborbits and to use $N_{ener} = 20$, $N_{I_2} = 8$ and $N_{I_3} = 5$ in order to get 800 independent orbits for both types of orbits (in all attempts we used $N_{dither} = 5$). The fitting routine was therefore fitting 3x800=2400 weights. This number was comparable to the number of orbits that van den Bosch et al. (2008) used. The range of energies, defined by r_{min} and r_{max}, and the error in each light bin were identical to those used in the high resolution models. The main difference is that we were using the same number of light bins as kinematic bins (9x9). The kinematic observables were computed in the same way as before. In this case, the light was fitted extremely well for almost all models, except the models with very low mass parameter v_0. As shown in figure 4.11, we found that using this configuration a minimum mass can be constrained assuming an edge-on view towards the system. But, as we only use a few number of light bins, it is not possible to put strong constraints on the flattening of the system. We were not able to determine the true flattening ($q = 0.8$) of the system.

In figures 4.12, 4.13 and 4.14 we will show some of the recovered properties of the mock galaxy for 3 different models: q72v20, q80v20 and q80v30. The q72v20 model is significantly worse than the q80v20 and q80v30 models. The isodensity contours of the light along the major axis for both model and data do not overlap (they do in the other models) and the recovered second velocity moments have much broader range of values compared to the other two models.

4.3.2 Attempt 2: Too large a number of kinematic bins (31x31)

As a second attempt we tried to model 31x31 bins for both the light and the kinematics. As the errors on the moments will be significantly larger as the mean number of stars per kinematic bin drops by a factor $\frac{31^2}{9^2} \simeq 12$, we decided to fit the moments obtained from inserting equal weights after making an ideal orbit library (see section 3.6.1) with 31x31 kinematic bins, in order to rule

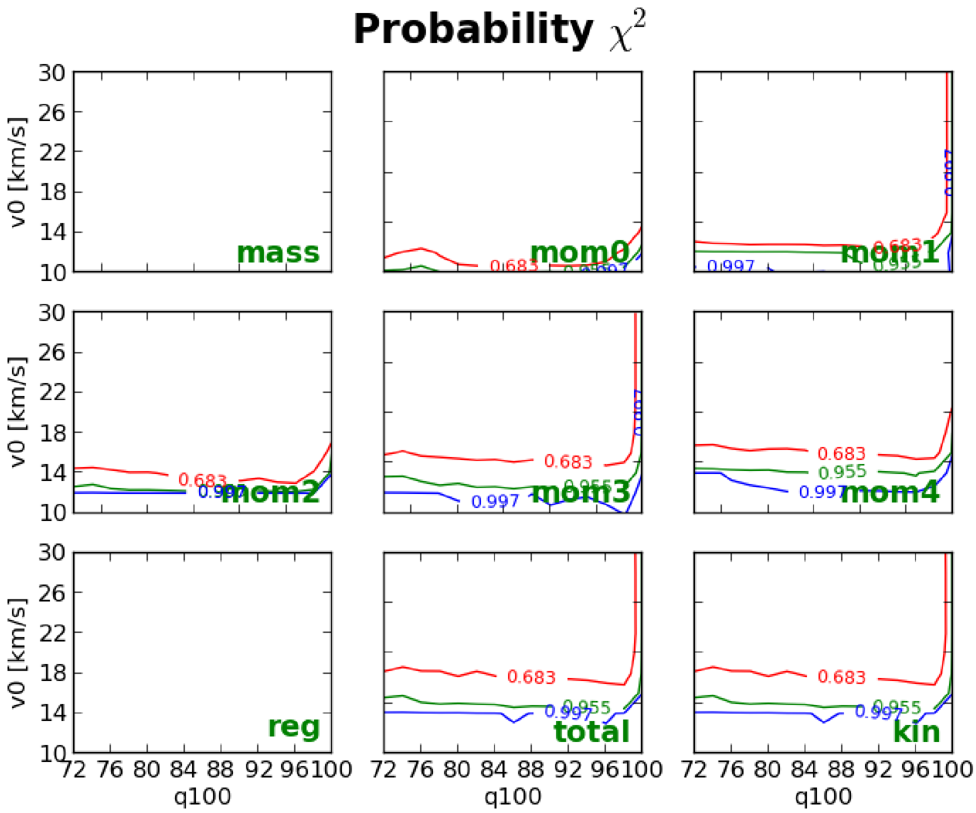
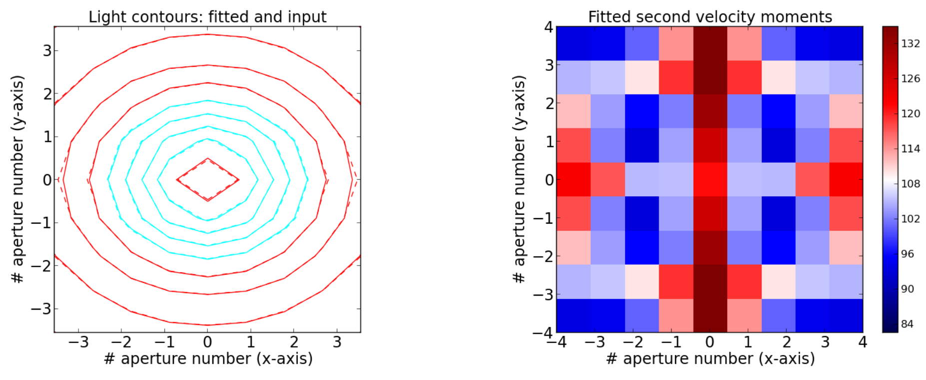
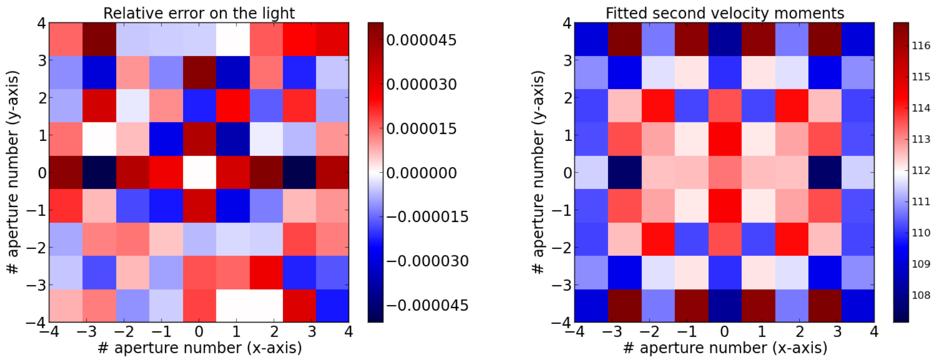


Figure 4.11: Figure showing contours of constant χ^2 for our models in attempt 1, using 9x9 light and kinematic bins and generating 10^5 stars in our field of view, assuming an edge-on view. The grid shows all models between $q = 0.72$ and $q = 0.98$ ($\triangle q = 0.02$) and between $v_0 = 10$ km/s and $v_0 = 30$ km/s ($\triangle v_0 = 2.0$ km/s). Only models with mass parameters $v_0 < 12$ km/s can be ruled out.



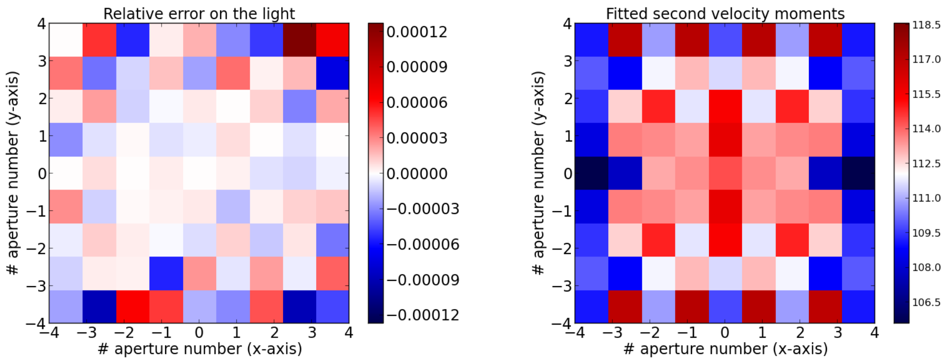
(a) (b)

Figure 4.12: Some of the recovered galaxy properties after fitting the q72v10 library of attempt 1 to our mock data consisting of 10^5 stars in our field of view (edge-on view). In panel (a) we show the light contours from both model (dashed) and 'observation' (full). The contours do not fully overlap along the major axis of the system. In panel (b) we show the recovered second moment in each kinematic bin. Along the minor axis, the second moments are much too high, whereas their values are much too low diagonally going outwards. This is an example of a bad model.



(a) (b)

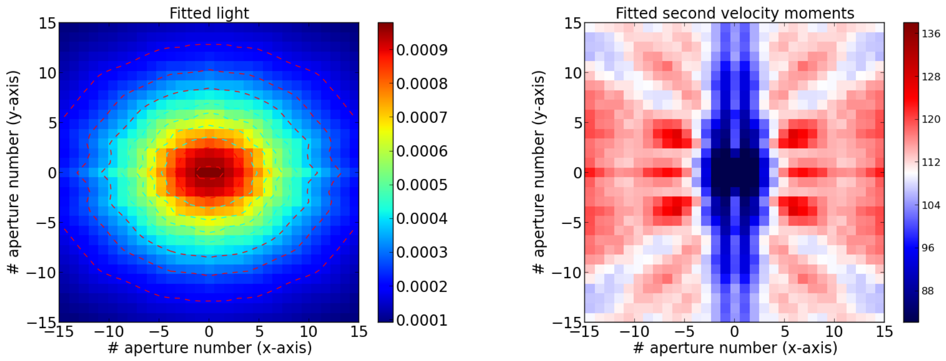
Figure 4.13: Some of the recovered galaxy properties after fitting the 'correct' q80v20 library of attempt 1 to our mock data consisting of 10^5 stars in our field of view (edge-on view). In panel (a) we now show the relative error on the light profile ([model-data]/data). Contours of both model and data do not give any information, as they fully overlap. In panel (b) we show the recovered second moment in each kinematic bin, which is clearly much better than in the case of q72v10.



(a) (b)

Figure 4.14: As in figure 4.13, but now for model q80v30. χ^2-analysis shows that this model is fitting the data even better, although not significantly better, than the true model q80v20: the light profile is fitted slightly worse, whereas the kinematic are fitted better.

out that such bad, but realistic, kinematics were going to worsen our results. With this approach we ensured that possible binning effects are diminished (see section 3.4.1). For this attempt we made a coarse grid and used the same number of orbits as in attempt 1 (see section 4.3.1). The flattening was varied from 0.6 to 1.0 in steps of 0.1, while mass parameter v_0 was varied from 10 to 30 km/s in steps of 5 km/s. This resulted in a best-fit model at a flattening of 0.6, but we note that physical models are only recovered as long as $q \geq 0.707$ (see section 2.1). Comparing the models with high enough flattening did give the correct best-fit model. However, a closer look at the recovered properties of this q80v20 model showed us that the light was not fitted well along the major axis of the system. In addition, the second raw moments in all kinematic bins showed a large range in values. Both claims are supported by figure 4.15. We checked whether the initial conditions were not complete in energy and angular momentum space, but we did not find a shortcoming. We decided to change our models another time by decreasing the number of kinematic bins while increasing the number of light bins. We decreased the number of kinematic bins, such that the model gets more freedom in fitting the velocity moments, and since it seemed that the kinematics were fitted the worst (as χ^2_{kin} was the dominant term in χ^2_{tot}).



(a) (b)

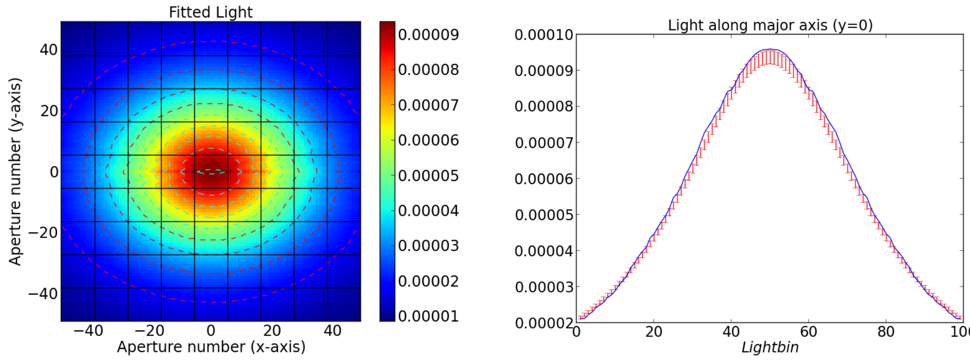
Figure 4.15: The recovered surface brightness profile and second velocity moments after fitting the 'correct' q80v20 library of attempt 2 to our ideal dataset with an edge-on view. In panel (a) we show the fitted light profile, where you can clearly see an abnormal pattern along the major axis. In panel (b) we show the recovered second moment in each kinematic bin, which shows a large range of values, implying that the fit is not very well.

4.3.3 Attempt 3: Too few orbits while using 9x9/3x3 kinematic bins and 99x99 light bins

In this section we show the shortcomings when using the same number of kinematic and light bins as in our good high resolution models, but when using the same number of orbits as in attempt 1 and 2 (800, see section 4.3.1). We show that generating orbit libraries that do not consist of a large enough number of orbits will not give satisfactory results.

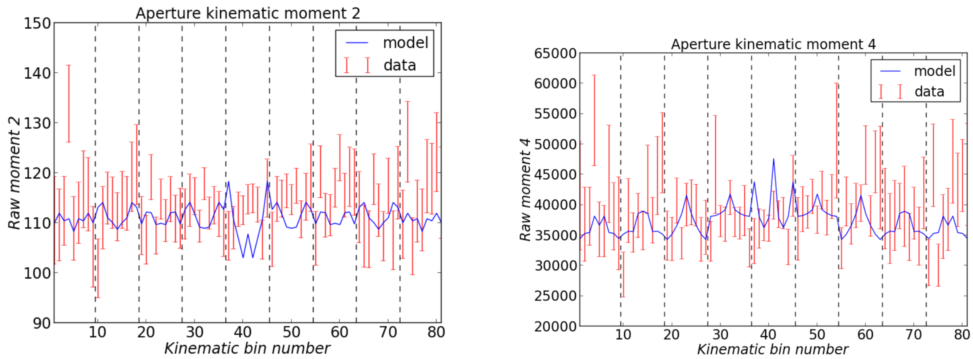
Since it is possible that the libraries contain too few different types of orbits in order to fit the kinematics in all 31x31 kinematic bins well, while keeping the light fitted well, we started to use a different number of bins for both the light and the kinematics in our third attempt. The number of kinematic bins was decreased to 9x9 again (as in attempt 1), while the number of light bins was increased to 99x99 to be able to get more information from the light profile. We used the realistic mock kinematics again (as in attempt 1) and the same number of orbits as in our first two attempts (800, see 4.3.1). We again did not recover the correct parameters of our potential. A possible explanation is that still the light profile was not fitted well along the major axis of the system, even if the true q80v20 model (figure 4.16) or the best-fit q96v20 model (figure 4.18) was chosen. In figures 4.17 and 4.19 we show the corresponding recovered even velocity moments. In both cases the second moments in the central kinematic bins along the major axis (bins 40-42) are rather underestimated. A systematic trend is visible in the recovered fourth moment, in the sense that kinematic bins in the columns around the minor axis (the central kinematic bins in each row of kinematic bins) have significantly larger values for its fourth moments than in the rest of our field of view, although the model does fit the data within the errors. In our good high resolution models, which are already described in section 4.1, we have therefore increased the number of orbits roughly by a factor of 10 in order to rule out that, in case we still do not fit the light profile well, we had used too few orbits in our previous attempts. As we have shown in that section, the data *is* fitted well when using more orbits.

To summarize section 4.3, bad choices for models are (1) using too few a number of light bins, (2) using too many kinematic bins and (3) using too few orbits in our orbit libraries. We note that taking too few kinematic bins, will not be satisfactory as well, as discussed in section 4.2.



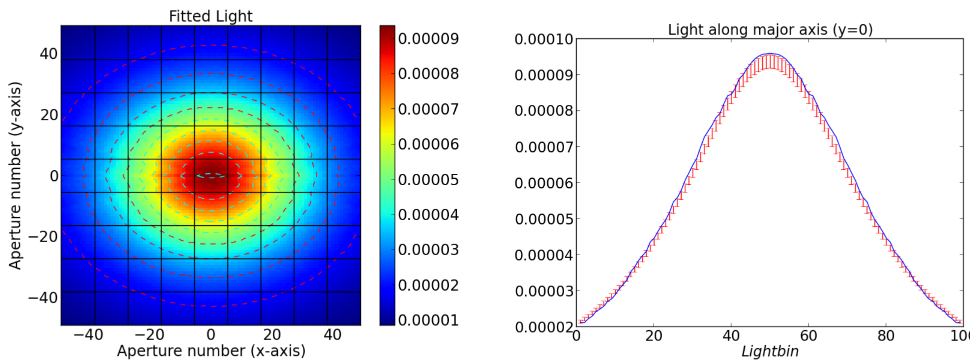
(a) (b)

Figure 4.16: The results after fitting the true q80v20 library of attempt 3 to our mock data consisting of 10^5 stars in our field of view, assuming an edge-on view. This time the light is fitted in 99x99 bins, the kinematics in 9x9 bins (see grid). In panel (a) we show the fitted surface brightness profile, where you can clearly see an abnormal pattern along the major axis. In panel (b) we show the fit along the major axis of the system.



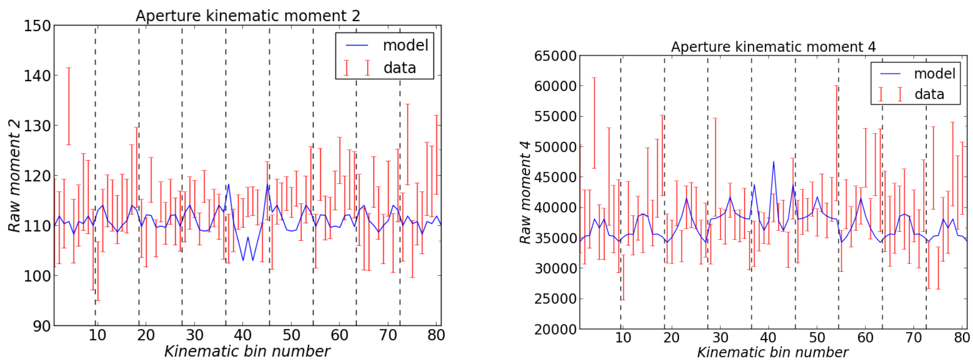
(a) (b)

Figure 4.17: As in figure 4.16, but now showing the fitted second (a) and fourth (b) moment. Both fits are reasonable.



(a) (b)

Figure 4.18: As in figure 4.16, but now showing the results for the best fit model q96v20. Even in this model the light is not fitted well.



(a) (b)

Figure 4.19: As in figure 4.18, but now showing the fitted second (a) and fourth (b) moment. The quality of the kinematic fit is similar to the one of the q80v20 model.

Chapter 5

NFW models: constraining the mass of our mock galaxy

We have shown that the Schwarzschild method can constrain the correct flattening and mass when the true functional form of the potential is known. Here we tackle the problem more realistically by allowing a different potential functional form. We modify the potential corresponding to a axisymmetric NFW-profile, following Vogelsberger et al. (2008):

$$\Phi(\tilde{r}) = -4\pi G\rho_0 R_s^3 \left[\frac{\ln(1+\tilde{r}/R_s)}{\tilde{r}}\right] \tag{5.1}$$

With respect to the spherical NFW-profile (Navarro et al. 1996), the radius $r = \sqrt{x^2+y^2+z^2}$ is being replaced by a newly defined radius $\tilde{r} = \frac{(r_a+r)r_E}{r_a+r_E}$, in which $r_E = \sqrt{(\frac{x}{a})^2+(\frac{y}{b})^2+(\frac{z}{c})^2}$ is the triaxial ellipsoidal radius, R_s the scale radius and r_a a transition radius. We require that $a^2+b^2+c^2 = 3$, such that choosing equal axes lenghts, i.e. $a = b = c = 1$, will result in the spherical NFW profile. For $r >> r_a$, $\tilde{r} \to r$, whereas for $r << r_a$, $\tilde{r} \to r_E$. Since we use $a = b$, we obtain an axisymmetric system in the central regions, whereas the potential becomes spherical in the outer regions. We use $r_a = 10$ kpc and since our mock field of view is 3x3 kpc, centered on the central parts of our mock galaxy, we might say that in this region $\tilde{r} \simeq r_E$. We note that assuming a potential Φ does not garantee that $\rho > 0$ everywhere and that this must be checked. The transition radius allows us to ensure that the total density, computed from the Poisson equation, is positive up to at least the orbits possessing the highest energies. We verified that choosing $r_a = 10$ kpc does satisfy this criterion as long as the flattening $c/a \geq 0.70$, see apendix B. Simply replacing r by r_E in stead of $\tilde{r}$, and thus ignoring the transition radius, only gives positive total densities inside the field covered by all orbits for the most round models: $c/a \geq 0.92$ (assuming a scale radius of 1.0 kpc). The total density corresponding to the triaxial Vogelsberger potential is given by:

$$\rho(x,y,z) = \frac{\nabla^2\Phi(x,y,z)}{4\pi G} \tag{5.2}$$

where

$$\nabla^2\Phi = \left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right]\Phi \tag{5.3}$$

such that

$$\begin{aligned}\rho(x,y,z) = -A\sum_{i=1}^{3}\Bigg\{ & x[i]^2\left(C1 + \frac{C2}{a[i]^2}\right)^2 B1 \\ & + B2\left[C1 + \frac{C2}{a[i]^2} + x[i]^2\,C3 + \left(\frac{x[i]}{a[i]}\right)^2 C4 + \left(\frac{x[i]}{a[i]^2}\right)^2 C5\right]\Bigg\}\end{aligned} \tag{5.4}$$

where (x[1],x[2],x[3]) = (x,y,z) and (a[1],a[2],a[3]) = (a,b,c) respectively and in which:

$$A = \rho_0 R_s^3 \tag{5.5}$$

$$B1 = \frac{1}{\tilde{r}}\left[\frac{2(D2 - D1)}{\tilde{r}} - \frac{1}{(R_s + \tilde{r})^2}\right] \tag{5.6}$$

$$B2 = \frac{1}{\tilde{r}}\left[D1 - D2\right] \tag{5.7}$$

$$C1 = \frac{1}{r_a + r_E}\left(\frac{r_E}{r}\right) \tag{5.8}$$

$$C2 = \frac{r_a + r}{(r_a + r_E)^2}\left(\frac{r_a}{r_E}\right) \tag{5.9}$$

$$C3 = -\frac{1}{r^2(r_a + r_E)}\left(\frac{r_E}{r}\right) \tag{5.10}$$

$$C4 = \frac{2}{r(r_a + r_E)^2}\left(\frac{r_a}{r_E}\right) \tag{5.11}$$

$$C5 = \frac{r_a + r}{r_E(r_a + r_E)}\left[\frac{2}{(r_a + r_E)^2} - \frac{1}{{r_E}^2} - \frac{1}{r_E(r_a + r_E)}\right] \tag{5.12}$$

$$D1 = \frac{1}{R_s + \tilde{r}} \tag{5.13}$$

$$D2 = \frac{1}{\tilde{r}}\ln(1 + \tilde{r}/R_s) \tag{5.14}$$

In this chapter we search for the best-fit parameters for scale radius R_s, mass M_{1kpc} and flattening c/a. We define that mass M_{1kpc}, expressed in units of $M_\odot$ of the model, such that it resembles the total enclosed mass within 1 kpc from the center in a spherical NFW profile with scale radius R_s. The mass of a spherical NFW profile is divergent and is given by:

$$M_{NFW}(R_{max}) = 4\pi G \rho_0 {R_s}^3\left[\ln\left(\frac{R_s + R_{max}}{R_s}\right) - \left(\frac{R_{max}}{R_s + R_{max}}\right)\right] \tag{5.15}$$

Therefore:

$$M_{1kpc} \equiv M_{NFW}(R_{max} = 1 \text{ kpc}) \tag{5.16}$$

Given a mass M_{1kpc} and scale radius R_s we compute ρ_0 by equation 5.15. It is this value that is used in the corresponding parameter of the axisymmetric Vogelsberger potential (equation 5.1) when making the orbit libraries. We note that the parameter ρ_0 of a NFW profile does not represent the central density, as it did in the composite model of our mock galaxy (equation 2.2). In an NFW profile the density diverges towards the center.

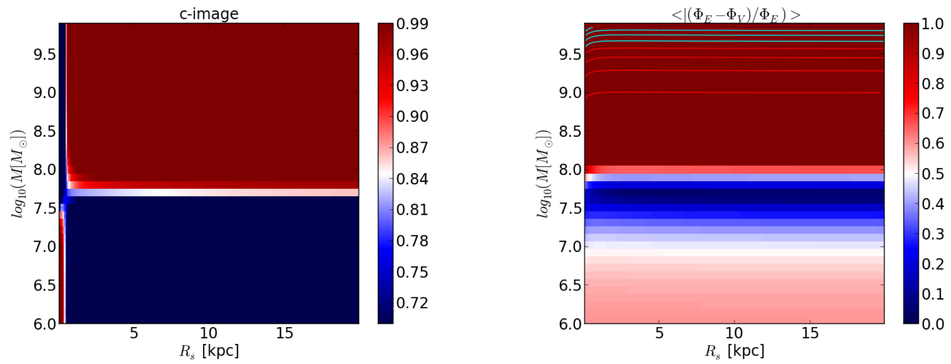
5.1 Setting up a grid of models

As in chapter 4 we will set up a grid of models in which the characteristic parameters of the Vogelsberger potential are varied. Therefore, we first need to develop an intuition for what model will resemble most closely the true potential. In order to estimate for what parameter range we should make our Vogelsberger models, we do a number of fits.

5.1.1 Fitting the potential

First, we search for what parameters of scale radius, mass and flattening the Vogelberger potential fits best the true logarithmic potential of our mock galaxy ($q = 0.8$, $R_c = 1.0$ kpc, $v_0 = 20$ km/s in equation 2.1). On a grid where we set $y = 0$ and where both $R = x$ and z range from 0.2 to 2.0 kpc, where consecutive coordinates are spaced by 0.1 kpc, we compute the true potential of our mock galaxy. Then we find the best-fit parameters for the Vogelsberger potential on the same grid. We do this in two different ways: either fixing the parameters R_s and M_{1kpc} or fixing R_s and c/a.

In the first approach we fix the scale radius and mass of the system. After setting a coordinate grid of R_s and M_{1kpc}, we use a fitting routine to find the optimal value for the flattening of the potential and the same for an offset parameter Φ_0 for the potential. The latter parameter has no physical meaning and only shifts the potential upwards or downwards. This extra freedom is needed in order to make a fair comparison between the fits. For every combination of parameters the value of Φ_0 will change. We put constraints on the values of the flattening such that values smaller than 0.7 (negative densities) and larger than 0.99 (round model) are not allowed. In panel (a) of figure 5.1 we show the best-fit parameters of the flattening on a grid ranging from $R_s = 0.1$ to $R_s = 20$ kpc ($\triangle R_s = 0.1$ kpc) and from $\log_{10}(M_{1kpc}) = 6.0$ to $\log_{10}(M_{1kpc}) = 9.9$ ($\triangle \log_{10}(M_{1kpc}) = 0.1$). We see that the values of the flattening reach our boundaries for nearly all models, except in the models with mass $\log_{10}(M_{1kpc}) \simeq 7.7$. In panel (b) we compute the relative difference of the fitted Vogelsberger potential with respect to the true logarithmic potential. The errors are smallest for $\log_{10}(M_{1kpc}) \simeq 7.7$ but we do not find a constraint on the scale radius. In figure 5.2 we show how well the best-fit parameters ($\log_{10}(M_{1kpc}) = 7.7$, $R_s = 4.3$ kpc, $c/a = 0.84$ and $\Phi_0 = 80.7$ km^2/s^2) recover the logarithmic potential at the (x,z)-coordinate grid. The potential at low values of z is fitted the worst.



(a) (b)

Figure 5.1: Panel (a) shows the best-fit values for the flattening. We set boundaries to the flattening at 0.7 (ensures positive densities everywhere) and 0.99 (nearly round model). Only models with $\log_{10}(M_{1kpc})$ equal to 7.7 or 7.8 result in values for c/a in between 0.8 and 0.95, when using a scale radius of 1 kpc as minimum. In panel (b) the relative error on the potential is shown, after fitting c/a and Φ_0. A mass that satisfies $\log_{10}(M_{1kpc}) = 7.7$ is giving the best results.

Because panel (a) of figure 5.1 did not show strong constraints on the best-fit flattening (they vary very quickly around models with best-fit mass $\log_{10}(M_{1kpc}) \simeq 7.7$), we decided to fix R_s and c/a in stead of R_s and M_{1kpc}. In this approach we fit the mass $\log_{10}(M_{1kpc})$ and offset parameter Φ_0, such that we can get more detailed information about the flattening parameter. We sampled the flattening from 0.7 to 0.98 with steps of 0.02 and R_s in the same way as in our first approach. In panel (a) of figure 5.3 we show that the best-fit mass is roughly $\log_{10}(M_{1kpc}) \simeq 7.66$ for all different combinations of R_s and c/a (except the models with $R_s < 0.5$ kpc). This is in agreement with our first approach. In panel (b) we show for every combination of R_s and c/a the relative errors on the recovered potential with respect to the true logarithmic potential after fitting $\log_{10}(M_{1kpc})$ and Φ_0. A flattening of $c/a = 0.78$, which is lower than what we found in our first approach ($c/a = 0.84$), and scale radii larger than 2 kpc are giving the best results.

5.1.2 Fitting circular velocities

Since the potential form is very different from the logarithmic potential the best fit still depicts relatively large differences (see figure 5.2). To gather more information, we also fit the circular velocities, the first derivative of Φ, along the major axis of the system. We define the circular velocity along the major axis of the system by:

$$\frac{v_c^2(R, z=0)}{R} = \nabla\Phi|_{R,z=0} \tag{5.17}$$

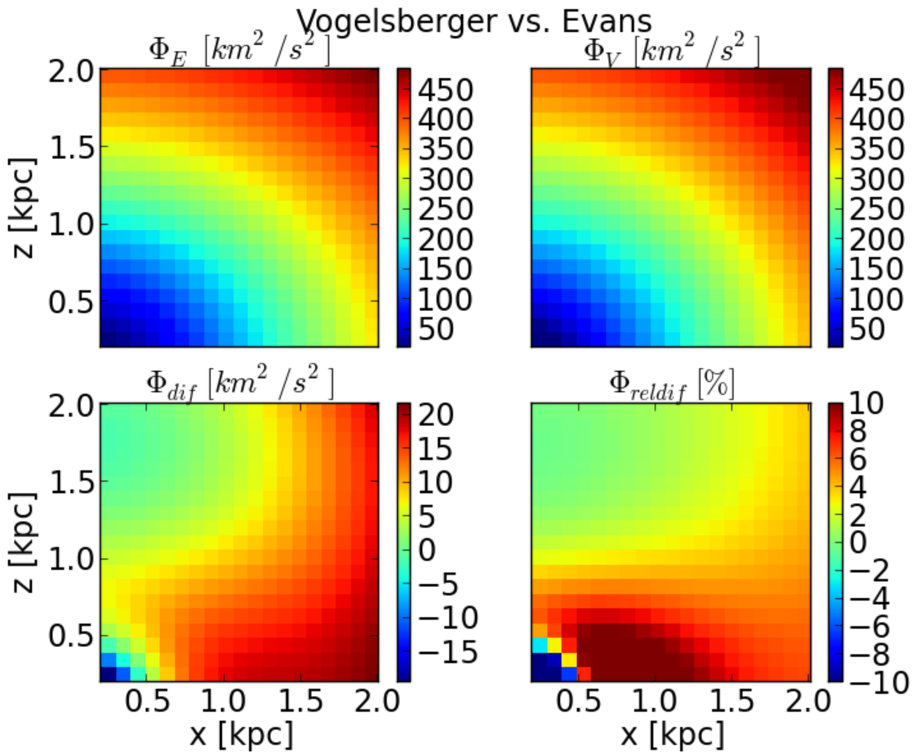


Figure 5.2: Figure showing how well the potential is fitted at the location of our best-fit parameters: $\log_{10}(M_{1kpc}) = 7.7$, $R_s = 4.3$ kpc, $c/a = 0.84$ and $\Phi_0 = 80.7$ km^2/s^2. From top left to bottom right: the logarithmic potential, the best-fit Vogelsberger potential, the absolute differences, the relative difference. In the fit the relative differences are minimized. For low z the relative errors are highest.

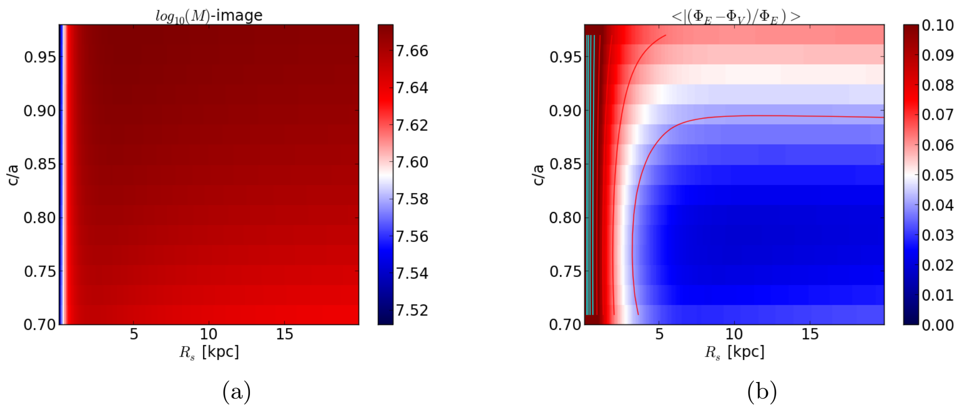


Figure 5.3: Panel (a) shows the best-fit values for the mass given a combination of scale radius and flattening. The best-fit mass equals $\log_{10}(M_{1kpc}) \simeq 7.66$, when using a scale radius of 1 kpc as minimum. In panel (b) the relative error on the potential is shown, after fitting the mass and Φ_0. A flattening around $c/a = 0.78$ is giving the best results.

where

$$\nabla\Phi = \left[\frac{\partial}{\partial R}\hat{R} + \frac{1}{R}\frac{\partial}{\partial\phi}\hat{\phi} + \frac{\partial}{\partial z}\hat{z}\right]\Phi \tag{5.18}$$

Then, for the logarithmic potential this becomes:

$$v_c^2(R, z=0) = \frac{v_0^2 R^2}{R_c^2 + R^2} \tag{5.19}$$

For the Vogelsberger potential the circular velocity equals:

$$v_c^2(R, z=0) = -4\pi G \rho_0 R_s^3 \left[\frac{1}{R_s + R} - \frac{\ln(1 + R/R_s)}{R} \right] \tag{5.20}$$

These circular velocities are independent of the flattening of the potential but give information about the mass and scale radius of the system. In figure 5.4 we show the mean absolute difference between the circular velocity corresponding to the logarithmic potential and the Vogelsberger potential, computed at all radii ranging from 0.01 kpc to 2.0 kpc, spaced by 0.01 kpc. We did this for the same combination of R_s- and $\log_{10}(M_{1kpc})$-values as in the first approach of section 5.1.1. We find that based on the first derivative of the potentials, we get the best results for $\log_{10}(M_{1kpc}) \simeq 7.7$. This is in agreement with the results from section 5.1.1 in which we fitted the potential itself. From fitting the circular velocities we also do not get much information about the scale radius of the system. We only see that $R_s < 0.5$ kpc does result in a worse fit of the circular velocities.

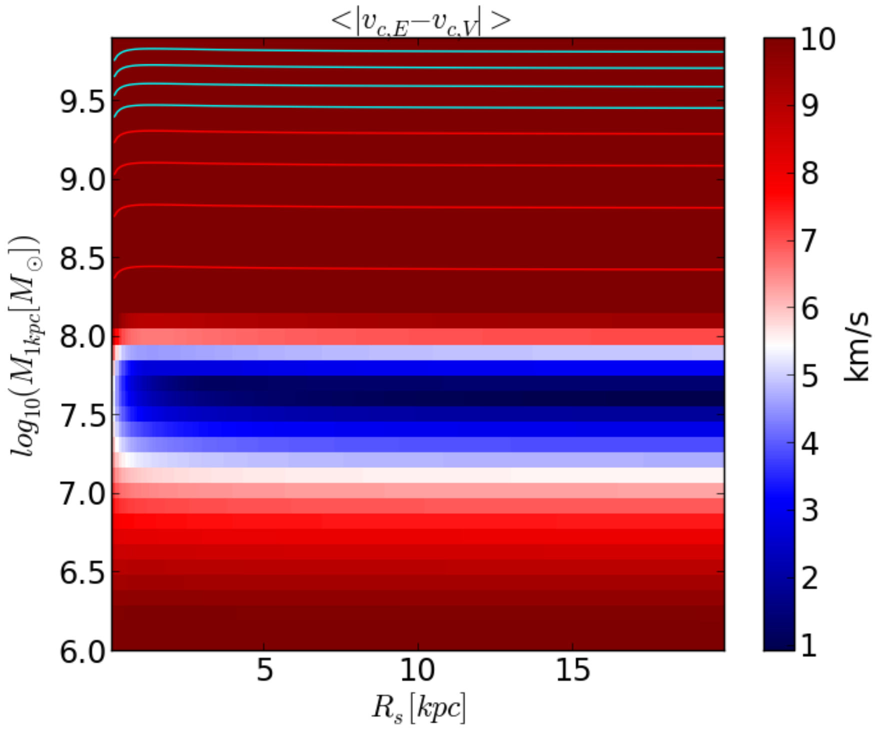


Figure 5.4: Figure showing how well the circular velocities can be fitted as function of the scale radius and the enclosed mass at 1kpc (in a spherical NFW profile). At each coordinate the mean difference between the logarithmic circular velocity and the Vogelsberger circular velocity is computed. The mean is computed after comparing the velocities at all radii from $R = 0.01$ to $R = 2.0$ kpc with $\triangle R = 0.01$ kpc. Fitting the circular velocity does seem to put a constraint on the mass, but not on the scale radius.

5.1.3 Comparing the flattening of the total density

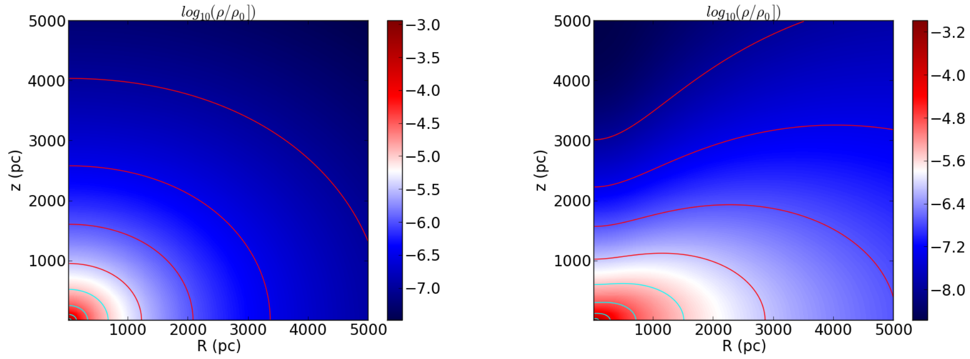
Because both approaches from section 5.1.1 resulted in slightly different best-fit values for the flattening, we decided to check for what values of c/a the flattening in the total density (see equation 5.4), the second derivative of Φ, is similar to the flattening of the total density corresponding to the logarithmic potential (see equation 2.5). We compute this axial ratio at different radii on the

c/a	0.1 kpc	0.5 kpc	1 kpc	3 kpc	5 kpc	10 kpc	30 kpc	50 kpc
E	0.73	0.72	0.69	0.57	0.51	0.47	0.43	0.43
	1kpc							
0.70	0.47	0.44	0.42	0.36	0.31	0.28	0.55	0.70
0.75	0.55	0.53	0.51	0.45	0.41	0.42	0.64	0.75
0.80	0.63	0.61	0.60	0.55	0.54	0.55	0.72	0.80
0.85	0.72	0.70	0.69	0.66	0.65	0.68	0.80	0.85
0.90	0.81	0.80	0.79	0.77	0.77	0.79	0.87	0.90
0.95	0.90	0.90	0.89	0.88	0.88	0.90	0.93	0.95
	5kpc							
0.70	0.47	0.46	0.46	0.47	0.49	0.50	0.73	0.80
0.75	0.55	0.54	0.54	0.56	0.57	0.63	0.77	0.83
0.80	0.63	0.63	0.63	0.64	0.66	0.71	0.82	0.87
0.85	0.72	0.71	0.72	0.73	0.74	0.78	0.87	0.90
0.90	0.81	0.81	0.81	0.82	0.83	0.86	0.91	0.93
0.95	0.90	0.90	0.90	0.91	0.91	0.93	0.96	0.97

Table 5.1: Flattening of the total density profile at different radii (columns 2-9) corresponding to an axial ratio of c/a in the potential (column 1). Row 1 shows the flattening of the total density of our mock galaxy (E=Evans). The flattening is higher in the center and converges to roughly 0.42 at very large radii. Rows 2-7 show the results when using the Vogelsberger potential and a scale radius of 1 kpc. Rows 8-13 show the results when using the Vogelsberger potential and a scale radius of 5 kpc. All results derived from the Vogelsberger potential are computed after setting the transition radius to 10 kpc. The Vogelsberger total density becomes rounder at larger radii. We verified that the flattening of the total density approaches 1.0 when choosing radii of order 1000 kpc. Changing the scale radius to 5 kpc does not change the flattening in the inner regions ($\leq$1 kpc), but does increase the flattening of the outer regions. For the inner regions a model with c/a = 0.85 would result in similar values for the flattening, whereas for the outer regions, models with a low value for the scale radius and a very low value for c/a would be needed.

major axis. In table 5.1 we compare the flattening of the Vogelsberger total density (rows 2-13) to the flattening of the total density corresponding to the true logarithmic potential of our mock galaxy (row 1). We set the transition radius to 10 kpc and investigate both models with a scale radius of 1 kpc and 5 kpc. We show that in the inner regions ($\leq$1 kpc), the flattening in the total density is similar when choosing a flattening of $c/a = 0.85$ in the Vogelsberger potential. At a distance of 3 to 5 kpc a flattening in the potential of $0.7 < c/a < 0.8$ (depending on the choice of the scale radius) would be needed, wheres for the larger radii even lower values for flattening are needed. We note that choosing higher scale radii result in larger values for the flattening in the total density. As the total density following the potential of our mock galaxy has rather low values for its flattening we do not need to increase the scale radius even further. Because the results from this section and section 5.1.1 show no strong constraints on the best-fit flattening of the Vogelsberger potential, we decided to vary it over the physically accepted range of values ($c/a \geq 0.7$). In figure 5.5 we show the total densities of the Vogelsberger potential corresponding to $c/a = 0.9$ (panel (a)) and $c/a = 0.7$ (panel (b)). Decreasing the flattening even more would results in negative densities, as described in appendix B.

In this section we found that most likely the best-fit Vogelsberger potential will have a mass equal to $\log_{10}(M_{1kpc}) \simeq 7.7$, a scale radius below 5 kpc (although based on section 5.1.3 only), but larger than 0.5 kpc (see section 5.1.1) and probably larger than 2 kpc (see section 5.1.2) and a flattening c/a in between 0.7 and 0.85 from this entire section.



(a) (b)

Figure 5.5: Figure showing the (positive) total densities when using the Vogelsberger profile with a scale radius of 1 kpc, transition radius of 10 kpc and a flattening of 0.9 in panel (a) and 0.7 in panel (b) in a field from $R = z = 0$ kpc to $R = z = 5$ kpc. The difference in the flattening of the total density is clearly visible. Models with $c/a < 0.7$ and $R_c = 1$ kpc would result in negative total densities within the spatial orbital range of our libraries (≤ 50 kpc).

5.2 Modeling axisymmetric Vogelsberger potentials

Following section 5.1 we decided to model 6 different values for the flattening ranging from 0.70 to 0.95 with steps of 0.05. For each value we made a (R_s-$\log_{10}(M_{1kpc})$)-grid that ranges from 1 to 5 kpc in scale radius ($\triangle R_s = 1$kpc) and from $\log_{10}(M_{1kpc})$=7.2 to $\log_{10}(M_{1kpc})$=8.0 for its mass [$\triangle \log_{10}(M_{1kpc}) = 0.2$].

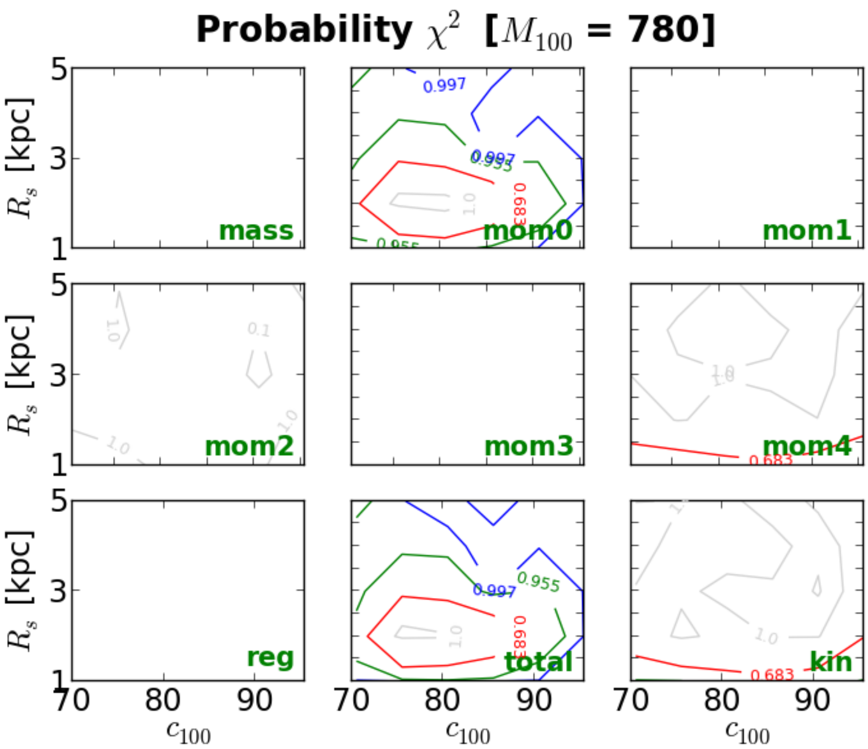


Figure 5.6: The results after fitting our mock data consisting of 10^5 stars inside our field of view, in a known edge-on view. We show the ($c100$, R_s)-2D-slice of probabilities at the location of the best fit model.

To be more efficient we here decrease the number of orbits to $N_{ener} = 24$, $N_{I_2} = 16$ and

$N_{I_3} = 8$. We have found this also gives good results in terms of recovery of the light profile and kinematics. Generating the initial conditions for both the orbits and the box orbits still takes approximately 2-3 hours, but making one orbit library now takes 1-2 days. The libraries each take roughly 1.0 GB disk space. The fitting routine needs 6-12 hours to compute the best fit orbital weights, when regularization is not included. The fitting routine needs 0.7 GB memory usage.

We present the results in the same way as we have done in chapter 4. The only difference is that we have a third degree of freedom in finding the best model, since we have started varying the scale radius in this case. Therefore, we have a cubic grid of models, in which the σ-confidence intervals are described by surfaces. For three degrees of freedom the 1σ-, 2σ- and 3σ-confidence intervals are defined by $\triangle\chi^2$ equal to 3.53, 8.02 and 14.2 respectively (Press et al. 1992). The Vogelsberger models are named by MxxxRsyyyczzz, in which xxx $= M_{100} \equiv 100\log_{10}(M_{1kpc}[M_\odot])$, yyy $= 100R_s$[kpc] and zzz $= c_{100} \equiv 100c/a$.

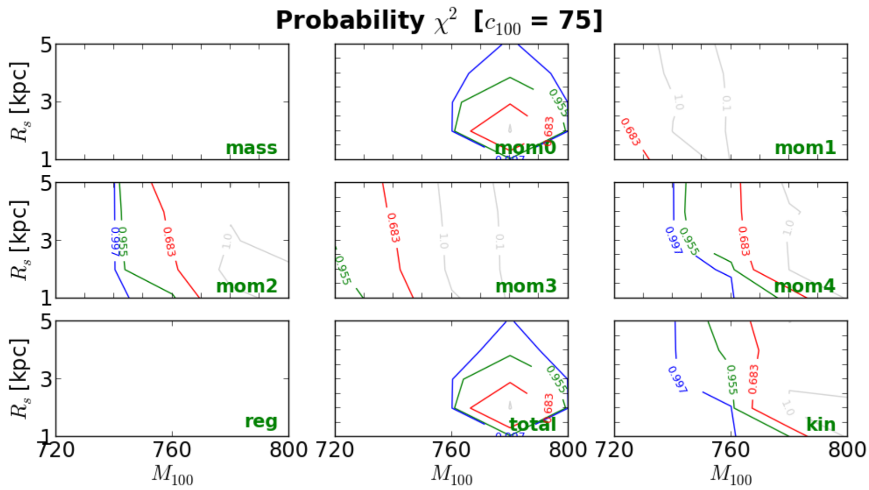


Figure 5.7: As in figure 5.6, but now showing the ($M100$, R_s)-2D-slice. In addition to the light, the kinematics clearly favor high mass models.

5.2.1 100000 stars

In this section we show how well we can recover the characteristic parameters of the Vogelsberger potential when pretending to observe 10^5 stars in a known edge-on view. From the cubic grid of models we show 2 perpendicular slices at the location of the best fit model, model M780Rs200c075 in which χ^2_{tot} has the lowest value ($\chi^2_{red} = 0.0244$). In figure 5.6 we show the ($c100$, R_s)-slice and conclude that we can constrain the scale radius of the true Vogelsberger potential to a value of $R_s = 2 \pm 1$ kpc and the flattening to $c/a = 0.75^{+0.15}_{-0.05}$ within a 1σ-confidence interval. In figure 5.7 we show the ($M100$, R_s)-slice and conclude that we can constrain the mass of the system to $\log_{10}(M_{1kpc}[M_\odot]) = 7.8 \pm 0.2$. In figure 5.8 we show that the M780Rc200c075 model does indeed fit the surface brightness profile very well and in figure 5.9 we show that the same holds for the second and fourth velocity moments.

5.2.2 10000 stars

We repeat the procedure to investigate how downsampling the number of stars affects the probability contours. Similar to what we have done in our previous chapter we decreased to number of stars in our field of view by a factor of 10, towards 10^4. Decreasing the number of stars towards 10^4 gives similar best-fit parameters compared to observing 10^5 stars. We did not fold the data as we did in section 4.2 for 10^4 stars. Model M780Rs200c080 is the best-fit model ($\chi^2_{red} = 0.00572$), which differs only in its flattening compared to the fit using 10^5 stars, although not significantly different as the 2σ-confidence intervals do not rule out models having a flattening c/a ranging from 0.7 to 0.9 for both 10^5 and 10^4 stars. In figure 5.10 we show the ($c100$, R_s)-slice and conclude that we can constrain the scale radius of the true Vogelsberger potential to a value of

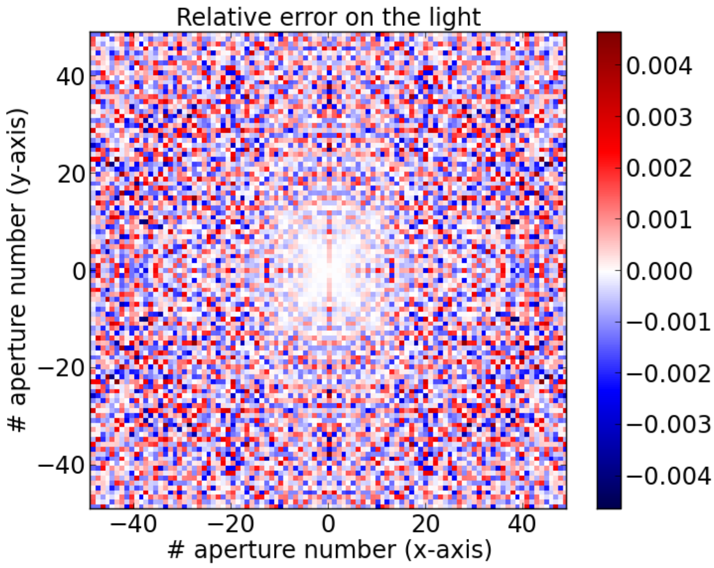
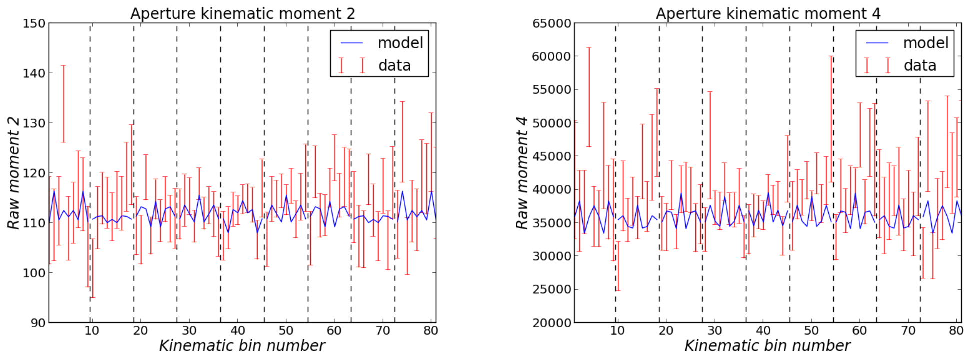


Figure 5.8: The relative errors on the light distribution after fitting the Vogelsberger M780Rs200c075 library to our mock data consisting of 10^5 stars in our field of view, assuming an edge-on view. The light is fitted in 99x99 bins, the kinematics in 9x9 bins.



(a)

(b)

Figure 5.9: As in figure 5.8, but now showing the fitted second (a) and fourth (b) velocity moment.

$R_s = 2^{+2}_{-1}$ kpc and the flattening to $c/a = 0.80^{+0.05}_{-0.10}$ within a 1σ-confidence interval. In figure 5.11 we show the ($M100$, R_s)-slice and conclude that we can constrain the mass of the system to $\log_{10}(M_{1kpc}[M_\odot]) = 7.8 \pm 0.2$.

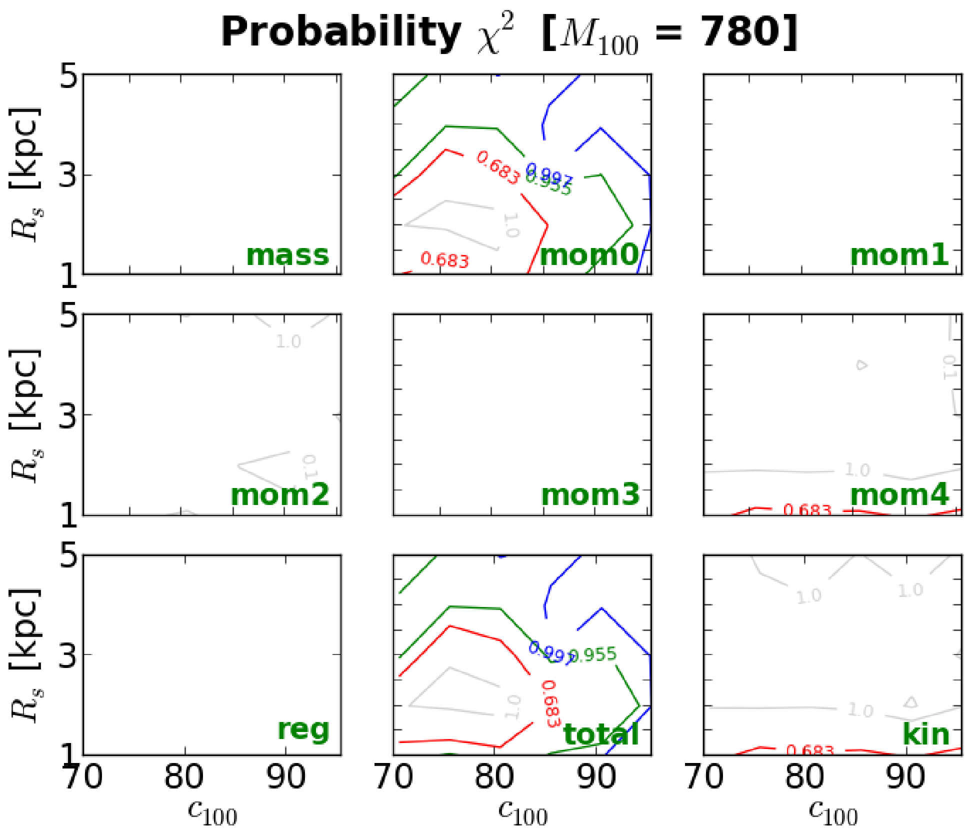


Figure 5.10: The results after fitting our mock data consisting of 10^4 stars inside our field of view, in a known edge-on view. We show the ($c100$, R_c)-2D-slice of probabilities at the location of the best fit model. The light seems to fully constrain the probability contours. Comparing the figure with figure 5.6, we see that the 2-σ contours do not change significantly.

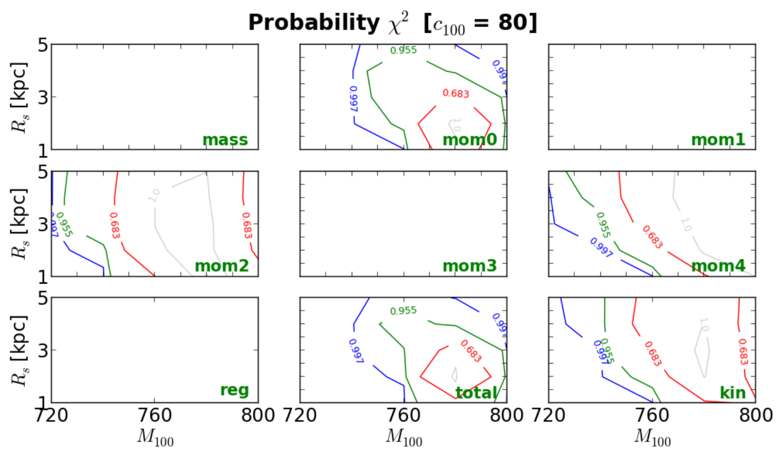


Figure 5.11: As in figure 5.10, but now showing the ($M100$, R_c)-2D-slice. In addition to the light, the kinematics clearly favor high mass models. Comparing the figure with figure 5.7, we see that the 2-σ contours of both the light and the fourth component do allow lower mass models to be fitted well, but combined the fit results do not change.

Chapter 6

Summary and conclusions

To test whether we can apply the axisymmetric Schwarzschild methods to real data of dwarf spheroidal galaxies meaningfully, we set up a Sculptor-like mock galaxy and tested whether we could recover its characteristic parameters.

In chapter 2, we set up the axisymmetric mock galaxy, whose global potential results in a logarithmic potential, that contains a flattened luminous and dark matter component. Both positions and velocities are drawn from the true distribution function with flattening parameter $q = 0.8$, mass parameter $v_0 = 20$ km/s, scale radius $R_c = 1.0$ kpc and slope parameter $p = 3.5$. We generated realistic kinematic datasets containing (1) 10^5 and (2) 10^4 stars by assuming a 2 km/s measurement error in the line-of-sight velocity for every star and assuming an edge-on view towards the system.

In chapter 3 we describe how we implemented the Schwarzschild modeling technique and in its last section, section 3.6, we tested the Schwarzschild method by making an ideal library, which contains the orbits of the mock stars themselves, such that all orbits should be represented in the model with equal weights. We verified that this is indeed the case and that the characteristic properties of our mock galaxy, like the light distribution and the velocity moments, are recovered.

In chapter 4 we made models that follow the logarithmic potential functional form but with varying the characteristic parameters v_0 and q. We show that we can recover the true characteristic parameters of the mock galaxy, assuming an edge-on view and observing the line-of-sight velocities, including realistic errors, for samples of either 10^5 or 10^4 stars. Within a 1σ-confidence interval we find for both cases that $q = 0.8^{+0.04}_{-0.04}$ and $v_0 = 20^{+3}_{-3}$ km/s, although we also can not rule out a model with $q = 0.96$ and $v_0 = 23$ km/s, values for which χ^2 is also minimized. In the case of 10^4 stars we folded the data into the positive quadrant of the sky by symmetry arguments to increase the S/N-ratio. As expected, the constraints that can be put on the parameters by observing 10^4 stars are less strong than in the case of 10^5 stars, as the 2σ-confidence intervals span over a larger area in that case.

In the same chapter we conclude that right choices need to be made in order to generate reliable Schwarzschild models. We show that we needed many light bins (99x99) in order to get enough information about the flattening of the system, not too many and not too few kinematic bins and high resolution models to constrain the parameters of the true potential and to recover the proper light distribution and velocity moments of the mock galaxy.

In chapter 5 we modeled an axisymmetric NFW potential functional form, the Vogelsberger potential. We performed fits to develop an intuition for what models will resemble the true potential of the mock galaxy the most closely. We found that most likely the best-fit model would satisfy a mass of $\log_{10}(M_{1kpc}[M_\odot]) \simeq 7.7$, a scale radius of $2 \leq R_s \leq 5$ kpc and a flattening of $0.70 \leq q \leq 0.85$. Then, we showed that we can recover the flattening, scale radius and mass of the system, assuming an edge-on view and observing the velocities including realistic errors of either 10^5 or 10^4 stars. In the case of 10^5 stars we conclude that we can constrain its characteristic scale radius to a value of $R_s = 2 \pm 1$ kpc, its flattening to $c/a = 0.75^{+0.15}_{-0.05}$ and the mass of the system to $\log_{10}(M_{1kpc}[M_\odot]) = 7.8 \pm 0.2$ within a 1σ-confidence interval. In the case of 10^4 stars we conclude that we can constrain its characteristic scale radius to a value of $R_s = 2^{+2}_{-1}$ kpc, its flattening to $c/a = 0.80^{+0.05}_{-0.10}$ and the mass of the system to $\log_{10}(M_{1kpc}[M_\odot]) = 7.8 \pm 0.2$ within a 1σ-confidence interval.

As we have shown in this work, it is possible to constrain the characteristic parameters of the logarithmic potential and the Vogelsberger potential when datasets consisting of 10^4 stars are available if one assumes an edge-on view. Most available samples are smaller, but it would be good to aim for such datasets in the near future (with e.g. WEAVE, 4MOST or other facilities).

We like to mention that we did not try the fit the viewing angles towards the system, although in reality these are not known. We also have not confirmed that, when using different realizations of our dataset, the best fit models have characteristic parameters well inside our 3σ-confidence intervals.

Chapter 7

Acknowledgements

I would like to thank my supervisors Amina and Maarten. I learned a lot from you while working on this project. I thank Remco van den Bosch for letting me use the Schwarzschild code and for his help when I did not completely understand it. I also thank the computer group for their support over the past year, especially when my laptop crashed. In addition, I thank my fellow students for the nice discussions during the lunch breaks.

Appendix A

MGE parametrization

In order to convert a surface brightness towards an intrinsic mass we take three steps, using a MGE parametrization. First we fit 2-dimensional Gaussians to the surface brightness profile. Then, we deproject these Gaussians into 3-dimensional Gaussians, assuming a set of viewing angles towards the system, to model the luminous density of the system. Since we only observe the light from the stars, we can assume a mass-to-light ratio to determine the potential of the system.

A 2-dimensional MGE model is fitted to the observed surface brightness $I(R', \theta')$, such that (van den Bosch et al. 2008):

$$I(R', \theta') = \sum_{j=1}^{N} \frac{L_j}{2\pi\sigma_j'^2 q_j'} \exp\left[-\frac{1}{2\sigma_j'^2}\left(x_j'^2 + \frac{y_j'^2}{q_j'^2}\right)\right] \tag{A.1}$$

where $x_j' = R' \sin(\theta' - \psi_j')$, $y_j' = R' \cos(\theta' - \psi_j')$ and where R' and θ' are the polar coordinates on the sky plane. L_j is the luminosity of the Gaussian j and N Gaussians are used in total. Their axial ratios and dispersions along the major axis are denoted by q_j' and σ_j'. The position angle ψ_j allows an isophotal twist of each Gaussian.

To demonstrate the multi-Gaussian expansion, we show a 2D-MGE parametrization from the luminous density ρ_{lum} in the (x,z)-plane $(y = 0)$, using the software from Cappellari (2002). The software assumes four-fold symmetry for the observed image and therefore takes the average of all four quadrants, after obtaining the principle axes of the image (see panel (a) of figure A.2). The software is able to take care of the background signal and fits a 2D MGE model to a number of 1D-sectors $(N_{sectors})$, equally spaced in polar angle (see panel (b) of figure A.2). Here we show an example in which we fitted $N = 8$ Gaussians while using $N_{sectors} = 19$. We choose to fit the density at coordinates $-5 < x < 5$ kpc and $-4 < z < 4$ kpc and choose $\rho_0\, R_c^p = 1$, $R_c = 1.0$ kpc, $q = 0.8$ and $p = 3.5$, such that:

$$\rho_{lum}(x, 0, z) = 1/\left[R_c^2 + x^2 + \left(\frac{z}{q}\right)^2\right]^{p/2} \tag{A.2}$$

Our test image is shown in figure A.1 and consists of 2001x1601 pixels of data. In panel (a) of figure A.3 we see the relative errors after applying the MGE parametrization. Our image can be parametrized quite well by 2D Gaussians as the relative errors are smaller than 1%, although this might in general not apply to all sorts of functional forms of the surface brightness profile.

Once the 2D MGE model has been made, the projected profile can be decomposed in a 3D MGE model (van den Bosch et al. 2008):

$$\rho(x, y, z) = \sum_{j=1}^{N} \left(\frac{M}{L}\right) \frac{L_j}{\left(\sigma_j\sqrt{2\pi}\right)^3 p_j q_j} \exp\left[-\frac{1}{2\sigma_j^2}\left(x^2 + \frac{y^2}{p_j^2} + \frac{z^2}{q_j^2}\right)\right] \tag{A.3}$$

where $\left(\frac{M}{L}\right)$ is the mass-to-light ratio and where p_j and q_j are the intrinsic axial ratios and where σ_j is the dispersion along the x-axis and where (x,y,z) is the intrinsic coordinate system. In order to do this one has to assume the viewing angles of the system (Cappellari 2002). The viewing angles are defined such that the spherical angles θ and ϕ determine the transformation from intrinsic

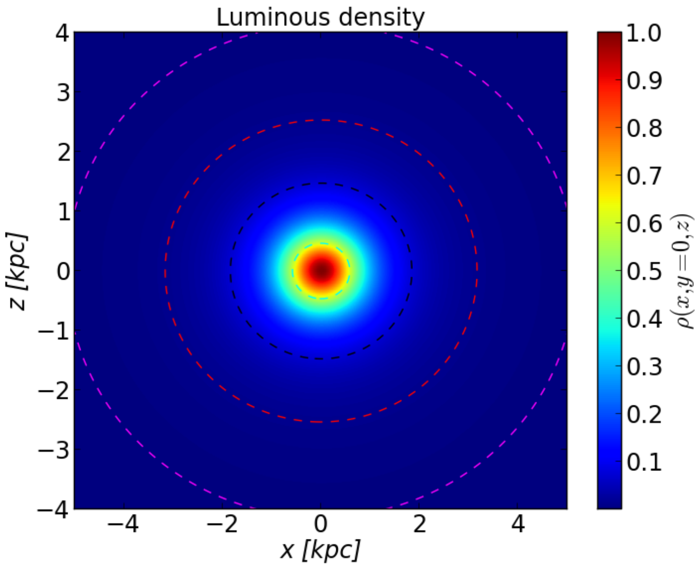
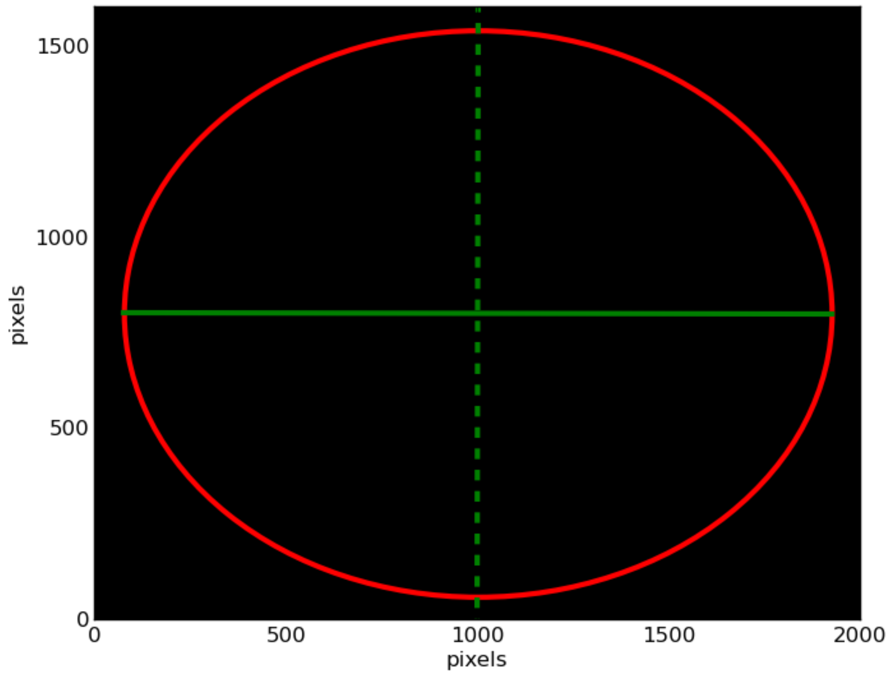


Figure A.1: The test profile to be fitted by the MGE parametrization.

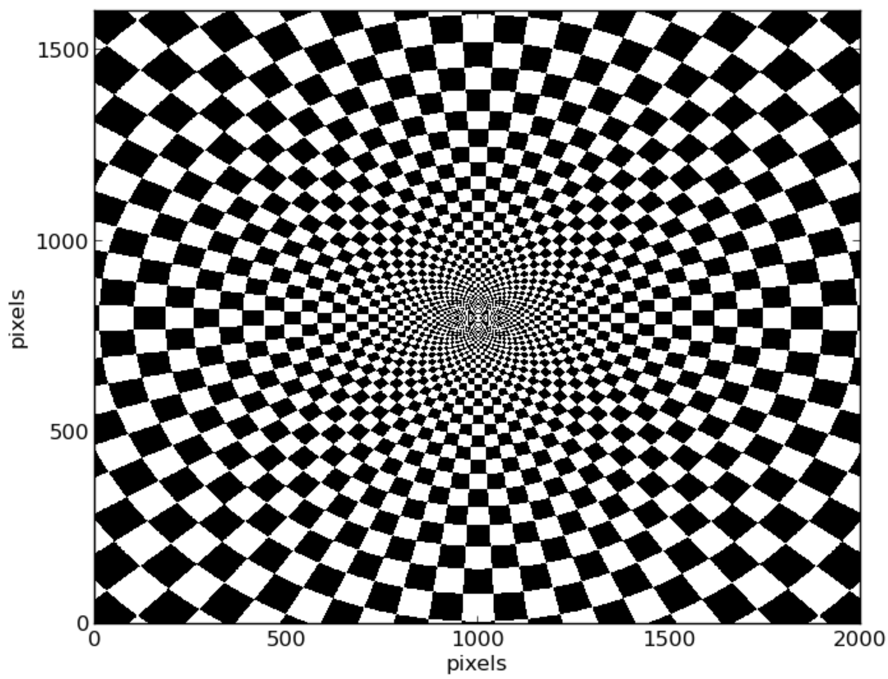
coordinates to the sky plane. They define the orientation of the line-of-sight with respect to the principal axes of the system. Therefore, (θ,ϕ) = (90,0) is a view down the major axis, (90,90) down the intermediate axis and $\theta = 0$ down the short axis of the system. In addition, an angle ψ specifies the rotation of the object around the line-of-sight. In this setup an oblate axisymmetric intrinsic shape satisfies $\psi = 90$. Increasing ψ will rotate the object clockwise in the sky plane. To demonstrate how these viewing angles are defined, we show the underlying surface brightness profile of our mock galaxy. For $\theta = 90$ and $\psi = 90$ we construct an edge on view. Decreasing ψ results in a counter-clockwise rotation of the image. Of course, changing ϕ has no impact on the projection, because we show an axisymmetric example. When viewing our object face-on ($\theta = 0$), a rotation in the sky plane is meaningless as well. The viewing angles are demonstrated in figure A.4.

Now we are able to make the triaxial MGE density, we can compute its potential by using the Chandrasekhar (1969) formulae. For $r = \sqrt{x^2 + y^2 + z^2} < 0.1\sigma_j$ and $r > 45\sigma_j$ an expansion is being used (de Zeeuw & Lynden-Bell 1985; van den Bosch et al. 2008), such that orbital integration speeds up significantly (van den Bosch et al. 2008). To compute the accelerations, the derivatives of these potentials are taken.

Since our composite system follows the logarithmic potential, we do not need to make a MGE parameterization. This is only needed if one wishes to assume that the total mass distribution follows the light distribution. One can still add dark matter potential forms, like a NFW profile.



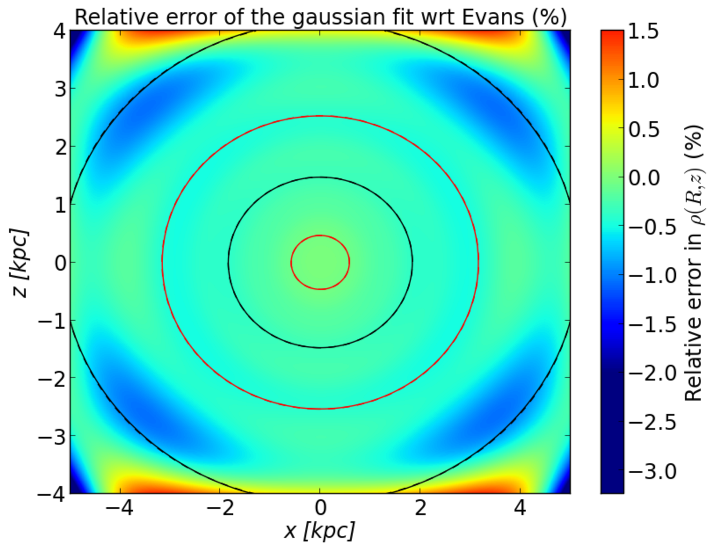
(a)



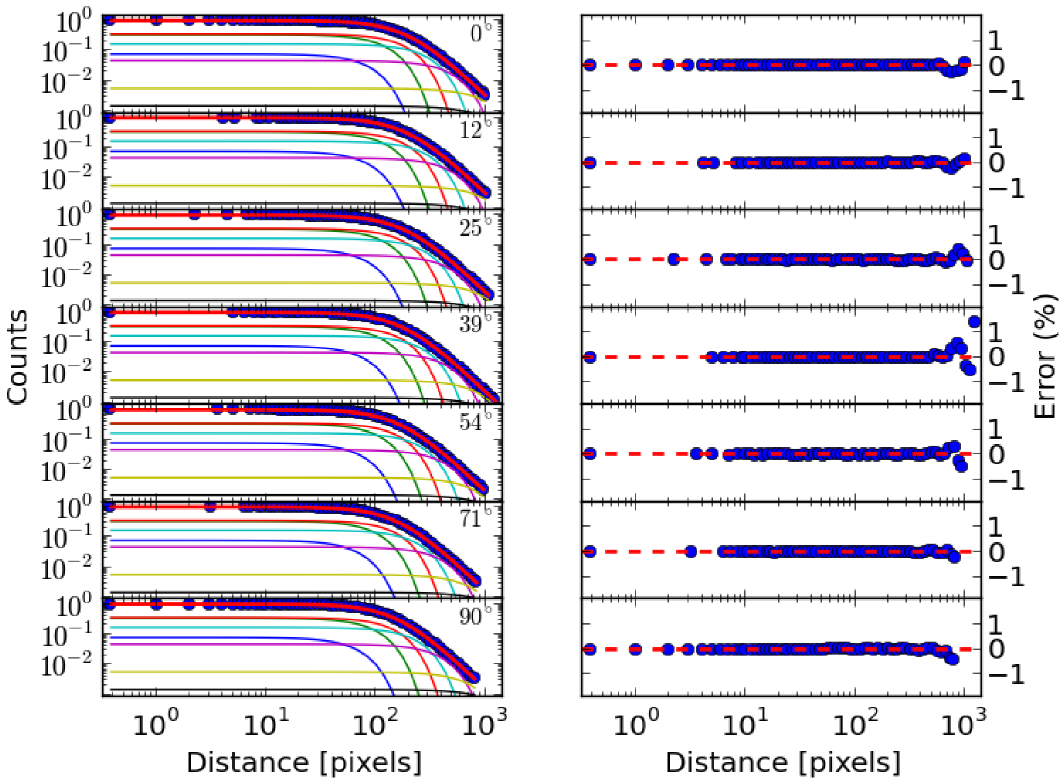
(b)

Figure A.2: Panel (a): Before doing the MGE parametrization, the MGE code searches for the principle axes of the observed system. Then it defines the corresponding sectors in which the surface brightness profile will be fitted (panel (b)). Both panels are produced using the software of Cappellari (2002).

.

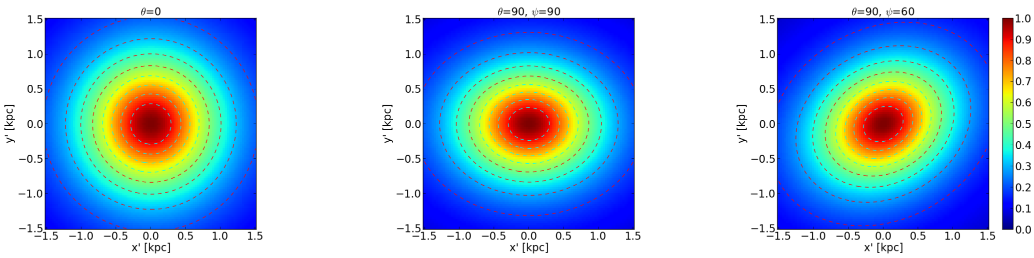


(a)



(b)

Figure A.3: Panel (a): Relative errors after doing a MGE parametrization. Dashed isodensity contours are drawn from the true density and the same full isodensity contours are drawn for the MGE model. The contours overlap. The relative errors are smaller than 1% for almost the entire image. Panel (b): visualization of the fit along several sectors. In the left column the figures show the image data as blue dots, the contribution of all N Gaussians as full coloured thin lines and the sum of all Gaussians, the model, as the red thick line. In the right column the relative errors between data and model are shown as function of radial distance. The figure is produced using the software of Cappellari (2002).



(a) (b) (c)

Figure A.4: Demonstration of the two most important viewing angles. Panel (a): surface brightness profile of our mock galaxy when viewing it face-on, θ=0. Since we have an axisymmetric system, the angle ϕ never changes the surface brightness profile. In this case the angle ψ has no influence as well. Panel (b): an edge-on view, θ=90, of our mock galaxy with ψ=90. Panel (c): same as panel (b), but now ψ=60. The image is effectively rotated 30 degrees anti-clockwise.

Appendix B

Choosing a NFW potential form

Changing the spherical NFW-profile towards a NFW profile in which the spherical radius $r = \sqrt{x^2 + y^2 + z^2}$ is replaced by the ellipsoidal radius $r_E = \sqrt{(\frac{x}{a})^2 + (\frac{y}{b})^2 + (\frac{z}{c})^2}$, where $a^2+b^2+c^2 = 3$, only gives positive densities for q≥0.92. To see this we first compute the density from the Poisson equation:

$$\rho(x, y, z) = -\frac{\rho_0 R_s^3}{r_E^2} \sum_{i=1}^{3} \frac{1}{a[i]^2} \Bigg\{ \left(\frac{x[i]}{a[i]}\right)^2 \left(\frac{3\ln(1 + r_E/R_s)}{r_E^3} - \frac{3}{(R_s + r_E)\, r_E^2} - \frac{1}{(R_s + r_E)^2\, r_E} \right) + \left(\frac{1}{R_s + r_E} - \frac{\ln(1 + r_E/R_s)}{r_E} \right) \Bigg\} \tag{B.1}$$

where (x[1],x[2],x[3]) = (x,y,z) and (a[1],a[2],a[3]) = (a,b,c) respectively and in which R_s is the scale radius. Then, when inserting $a = b$ to make the system axisymmetric, and choosing different values for the flattening c/a, we investigate the density of our system up to a distance of 50 kpc from the center, which corresponds to the distance that the highest energy orbits of our libraries can reach. Therefore we ensure that our orbits do not enter regions which result in negative densities. Using $R_s = 1$ kpc we find that negative densities show up for $c/a < 0.92$ (see figure B.1 for an example). Changing the scale radius to 5 kpc does give positive densities everywhere inside the 50 kpc boundaries for models with a flattening greater than 0.85.

In order to be able to use models which allow lower values for the flattening, we started to use the potential functional form described by Vogelsberger et al. (2008). Now the spherical radius is replaced by a radius $\tilde{r} = \frac{(r_a+r)r_E}{r_a+r_E}$. For $r >> r_a$, $\tilde{r} \to r$ (spherical), whereas for $r << r_a$, $\tilde{r} \to r_E$ (ellipsoidal). This time the density is already given in equation 5.4. Choosing a transition radius of 10 kpc results in positive densities up to 50 kpc in potentials with flattening $c/a \geq 0.70$, while using a scale radius of 1 kpc. In figure B.2 we show the densities for potentials with $c/a = 0.9$ (panel (b)) and $c/a = 0.7$ (panel (b)). Decreasing the flattening even further would result in negative densities along the minor axis.

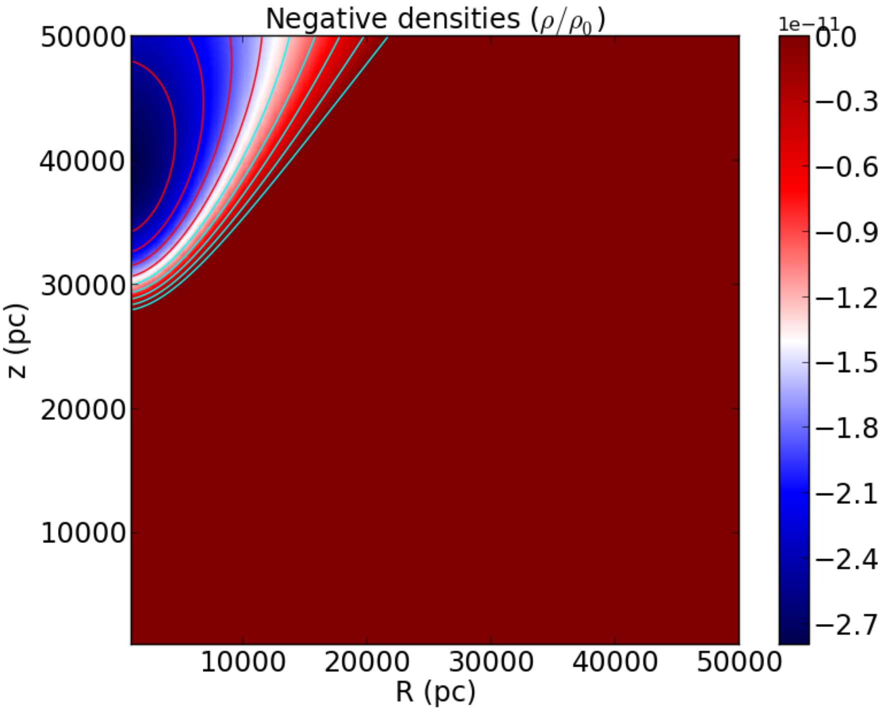
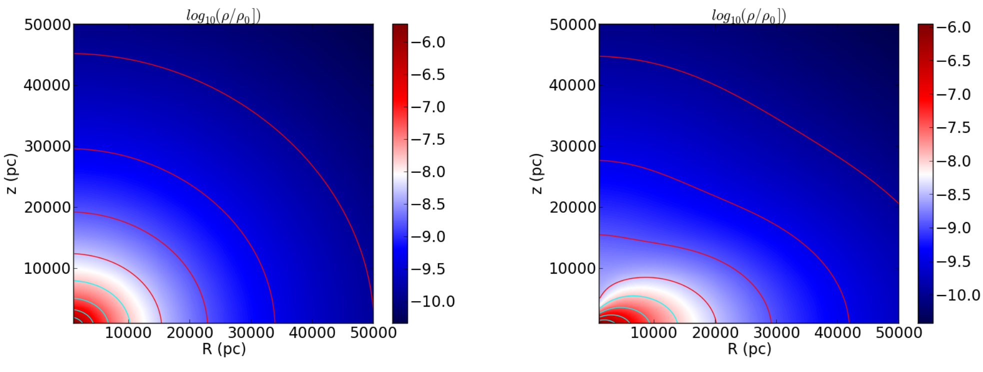


Figure B.1: Figure showing the negative densities when using the triaxial NFW profile with a scale radius of 1 kpc and a flattening of 0.9 inside a 50x50 kpc region in the (R,z)-plane.



(a)

(b)

Figure B.2: Figure showing the (positive) densities when using the Vogelsberger profile with a scale radius of 1 kpc, transition radius of 10 kpc and a flattening of 0.9 (panel(a)) and 0.7 (panel (b)) in the 50x50 kpc region.

Bibliography

Battaglia, G. 2007, PhD Thesis

Battaglia, G., Helmi, A., Tolstoy, E., et al. 2008, ApJ, 681, L13

Battaglia, G., Tolstoy, E., Helmi, A., et al. 2006, A&A, 459, 423

Bechtol, K., Drlica-Wagner, A., Balbinot, E., et al. 2015, ApJ, 807, 50

Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, ApJS, 208, 20

Binney, J. 1978, MNRAS, 183, 779

Binney, J. & Mamon, G. A. 1982, MNRAS, 200, 361

Binney, J. & Tremaine, S. 2008, Galactic Dynamics: Second Edition (Princeton University Press)

Boylan-Kolchin, M., Bullock, J. S., & Kaplinghat, M. 2011, MNRAS, 415, L40

Breddels, M. A. & Helmi, A. 2013, A&A, 558, A35

Breddels, M. A., Helmi, A., van den Bosch, R. C. E., van de Ven, G., & Battaglia, G. 2013, MNRAS, 433, 3173

Cappellari, M. 2002, MNRAS, 333, 400

Cappellari, M. 2015, in IAU Symposium, Vol. 311, IAU Symposium, ed. M. Cappellari & S. Courteau, 20–30

Cappellari, M., Bacon, R., Bureau, M., et al. 2006, MNRAS, 366, 1126

Chandrasekhar, S. 1969, Ellipsoidal figures of equilibrium

Cretton, N., de Zeeuw, P. T., van der Marel, R. P., & Rix, H.-W. 1999, ApJS, 124, 383

de Zeeuw, P. T. & Lynden-Bell, D. 1985, MNRAS, 215, 713

Einstein, A. 1936, Science, 84, 506

Evans, N. W. 1993, MNRAS, 260, 191

Freeman, K. C. 1970, ApJ, 160, 811

Hayashi, K. & Chiba, M. 2015, ApJ, 810, 22

Hubble, E. P. 1927, in Publications of the American Astronomical Society, Vol. 5, Publications of the American Astronomical Society, 261–264

Hui, L. 2001, Physical Review Letters, 86, 3467

Jin, S., Helmi, A., & Breddels, M. 2015, ArXiv e-prints, 1502.01215

Kant, I. 1755, Allgemeine Naturgeschichte und Theorie des Himmels

Klypin, A., Kravtsov, A. V., Valenzuela, O., & Prada, F. 1999, ApJ, 522, 82

Koposov, S. E., Belokurov, V., Torrealba, G., & Evans, N. W. 2015, ApJ, 805, 130

Levison, H. F. & Richstone, D. O. 1985, ApJ, 295, 340

Mateo, M. 2000, Dwarf Spheroidal Galaxies, ed. P. Murdin, 2616

Mateo, M., Olszewski, E., Welch, D. L., Fischer, P., & Kunkel, W. 1991, AJ, 102, 914

Mateo, M. L. 1998, ARA&A, 36, 435

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

Moore, B., Ghigna, S., Governato, F., et al. 1999, ApJ, 524, L19

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

Ollongren, A. 1962, Bull. Astron. Inst. Netherlands, 16, 241

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565

Peter, A. H. G. 2012, ArXiv e-prints, 1201.3942

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2015, ArXiv e-prints, 1502.01589

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical recipes in FORTRAN. The art of scientific computing

Renn, J., Sauer, T., & Stachel, J. 1997, Science, 275, 184

Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009

Rubin, V. C., Ford, W. K. J., & . Thonnard, N. 1980, ApJ, 238, 471

Schwarzschild, M. 1979, ApJ, 232, 236

Schwarzschild, M. 1993, ApJ, 409, 563

Shapley, H. 1938, Harvard College Observatory Bulletin, 908, 1

Shapley, H. & Curtis, H. D. 1921, Bulletin of the National Research Council, Vol. 2, Part 3, No. 11, p. 171-217, 2, 171

Strigari, L. E., Koushiappas, S. M., Bullock, J. S., et al. 2008, ApJ, 678, 614

van Agt, S. 1978, The Messenger, 12, 3

van den Bergh, S. 2008, MNRAS, 385, L20

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., Cretton, N., de Zeeuw, P. T., & Rix, H.-W. 1998, ApJ, 493, 613

Vera-Ciro, C. A., Helmi, A., Starkenburg, E., & Breddels, M. A. 2013, MNRAS, 428, 1696

Vogelsberger, M., White, S. D. M., Helmi, A., & Springel, V. 2008, MNRAS, 385, 236

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

Walker, M. G., Mateo, M., Olszewski, E. W., et al. 2007, ApJ, 667, L53

Walker, M. G., Mateo, M., Olszewski, E. W., et al. 2009b, ApJ, 704, 1274

Walsh, D., Carswell, R. F., & Weymann, R. J. 1979, Nature, 279, 381

Webster, D., Sutherland, R., & Bland-Hawthorn, J. 2014, ApJ, 796, 11

Wolf, J., Martinez, G. D., Bullock, J. S., et al. 2010, MNRAS, 406, 1220

Zwicky, F. 1933, Helvetica Physica Acta, 6, 110