The RoadMapping Code: How to deal with "Real World" Issues in Action-based Dynamical Modelling the Milky Way

W. Trick[1,2], J. Bovy[3], and H.-W. Rix[1]

trick@mpia.de

ABSTRACT

We present *RoadMapping* , a dynamical modelling machinery that aims to recover the Milky Way's (MW) gravitational potential and the orbit distribution of stellar ~~mono-abundance~~ populations (~~MAP~~) in the Galactic disk ~~(Bovy & Rix 2013; Binney & McMillan 2011; Binney 2012)~~. *RoadMapping* is a full likelihood analysis that models the observed positions and velocities of ~~MAP~~ stars with an equilibrium, three-integral ~~quasi-isothermal~~ distribution function (qDF) in an axisymmetric potential. In preparation for the application to the large data sets of ~~modern Galactic~~ surveys like Gaia, we create and analyze a large suite of mock data sets and develop qualitative "rules of thumb" which characteristics and limitations of data, model and machinery affect constraints on the potential and qDF most. We find that, while the precision of the recovery increases with the number of stars, the numerical accuracy of the likelihood normalisation becomes increasingly important and dominates the computational efforts. The modelling has to account for the survey's selection function, but *RoadMapping* seems to be very robust against small misjudgments of the data completeness. Large radial and vertical coverage of the survey volume gives in general the tightest constraints. But no observation volume of special shape or position and stellar population should be clearly preferred, as there seems to be no stars that are on manifestly more diagnostic orbits. [TO DO: Write about results on measurement errors] We investigate how small deviations of the stars' distribution from the assumed qDF affect the modelling: Having less stars at small Galactocentric

[1]Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

[2]Correspondence should be addressed to trick@mpia.de.

[3]Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA

~~[4]Hubble fellow~~

radii R and with lower velocities as predicted by the qDF recovers the potential more reliably than having too many stars at small R and with higher velocities. The deviations of the true orbit distribution from the qDF introduced by binning stars into *MAP*s in the ([Fe/H],[α/Fe]) plane and due to abundance errors do not matter much, as long as the qDF parameters of two neighbouring *MAP*s do not vary more than 20% [TO DO: CKECK]. As the modelling has to assume a parametric form for the gravitational potential, deviations from the true potential have to be expected. We find, that in the axisymmetric case we can still hope to find a potential that is indeed a reliable best fit within the limitations of the assumed potential. Overall *RoadMapping* works as a reliable and unbiased estimator, and is robust against small deviations between model and the "real world".

Subject headings: Galaxy: disk — Galaxy: fundamental parameters — Galaxy: kinematics and dynamics — Galaxy: structure

Contents

1. Introduction

Stellar dynamical modelling is the fundamental tool to infer the gravitational potential of the Milky Way from the positions and motions of its stars (Rix & Bovy 2013; Binney 2011b) [TO DO]. The observational information on the phase-space coordinates of stars are currently growing at a rapid pace, and will be taken to a whole new level by the upcoming Gaia data. Yet, rigorous and practical modelling tools that turn this information into constraints both on the gravitational potential and on the distribution function (DF) of stellar orbits, are scarce (Rix & Bovy 2013) [TO DO: more references] [TO DO: References that explain that the modelling is scarce, or previous modelling approaches???]

[TO DO: Modelling tools for the MW: a) Made-to-measure: De Lorenzi et al. (2007)(based on Syer & Tremaine (1996) , best application to to bulge Bissantz et al. (2004), Hunt & Kawata (2014) also have a tool for Gaia at hand, b) ~~Streams: Johnston et al.~~ (1999) c) Action-based distribution function modelling Sanders & Binney (2015) Piffl et al. (2014) d) torus modelling e) Jeans modelling Büdenbender et al. (2015) ~~Loebman et al. (2012)~~)]

Accurately determining the Galactic gravitational potential is fundamental for understanding its dark matter and baryonic structure [REF]. Accurately determining the stellar-population dependent orbit distribution function is a fundamental contraint on the Galaxy's formation history.

Open questions about the MW's potential and structure, on which future modelling attempts will hopefully give more definite answers are: What is the local dark matter density (Zhang et al. 2013; Bovy & Tremaine 2012)? Is the Milky Way's dark matter halo flattened ([REF])? Is the MW disk maximal (Sackett 1997) and, to be able to disentangle halo and disk contribution (Dehnen & Binney 1998), what is the disk's overall mass scale length (Bovy & Rix 2013)?

Open questions about the star's distribution within the MW, which dynamical modelling can help to constrain, are: How are stellar kinematics and their chemical abundances are related (Sanders & Binney (2015),[REF])? In particular, does the disk have a thin/thick disk dichotomy (Gilmore & Reid (1983)) or is it a continuum of many exponential disks (Bovy et al. (2012d))? How does radial migration affect the orbit distribution (Sellwood & Binney 2002; Roškar et al. 2008a,b; Schönrich & Binney 2008; Minchev et al. 2011) [TO DO: These are References from Rix & Bovy 2013 - should I use all of them?]?
To address these questions, observed stellar positions and motions need to be turned into full orbits - which stresses again the importance of having a reliable model for the MW's

gravitational potential.

In the era of big Galactic surveys all of this could soon be within our reach. Not only will there be full 6D stellar phase-space coordinates for a ~~thousand million~~ billion of stars measured by Gaia to unprecedented precision by ~~the end of 2016.~~ early 2017. But already with existing surveys (e.g., SEGUE ~~(Beers et al. 2006),~~ RAVE (Steinmetz et al. 2006), LAMOST (Newberg et al. 2012), APOGEE (Majewski 2012), Gaia-ESO (Gilmore et al. 2012), GALAH (Freeman 2012) [TO DO: I just copied this from Melissas Cannon paper. Should I reference all of them??? Not in reference list yet.]) and sophisticated machine-learning tools (e.g. ~~The Cannon by~~ Ness et al. (2015)) to combine them, we will soon have huge data sets at our disposal.

Yanny et al. (2009)

In this work we present a rigorous, robust and reliable dynamical modelling machinery, strongly building on previous work by Binney & McMillan (2011); Binney (2012); Bovy & Rix (2013); Bovy (2015) and explicitly developed to exploit and deal with these large data sets in the future.

There is a variety of practical approaches to dynamical modelling of discrete collisionless tracers (such as the stars in the Milky Way) [REF]. Most of them – explicitly or implicitly – describe the stellar distribution through a distribution function. Actions are good ways to describe orbits, because they are canonical variables with their corresponding angles, have immediate physical meaning, and obey adiabatic invariance [Binney 2011abcdefg???].

Recently, Binney (2012) and Bovy & Rix (2013) [TO DO: are these the correct references???] proposed to combine parametrized axisymmetric potentials with DF's that are simple analytic functions of the three orbital actions to model discrete data. Binney (2010) and Binney & McMillan (2011) had proposed a set of simple action-based (quasi-isothermal) distribution functions (qDF). ? and Bovy & Rix (2013) showed that these qDF's may be good descriptions of the Galactic disk, when one only considers so-called mono-abundance populations (*MAP*), i.e. sub-sets of stars with similar [Fe/H] and [α/Fe] (Bovy et al. (2012b), Bovy et al. (2012c), Bovy et al. (2012d)).

Bovy & Rix (2013) implemented a modelling approach that put action-based DF modelling of the Galactic disk in an axisymmetric potential in practice. Given an assumed potential and an assumed DF, they directly calculated the likelihood of the observed $(\vec{x}, \vec{v})$ for each sub-set of *MAP* among SEGUE ~~Gdwarf~~ (Yanny et al. 2009). This modelling also accounted for the complex, but known selection function of the kinematic tracers. For each

Do not italicize "MAP". Change everywhere.

G-dwarf stars

MAP , the modelling resulted in a constraint of its DF, and an independent constraint on the gravitational potential, which members of all *MAP*s feel the same way.
Taken as an ensemble, the individual *MAP* models constrained the disk surface mass density over a wide range of radii ($\sim 4-9$ kpc), and proved a powerful constraint on the disk mass scale length (~~~ 2 kpc~~) and on the disk-to-dark-matter ratio at the Solar radius [~~TO DO: quote number???~~].

Yet, these recent models still leave us poorly prepared with the wealth and quality of the existing and upcoming data sets. This is because Bovy & Rix (2013) made a number of quite severe and idealizing assumptions about the potential, the DF and the knowledge of observational effects (such as the selection function). All these idealizations are likely to translate into systematic error on the inferred potential or DF, well above the formal error bars of the upcoming data sets.

In this work we present *RoadMapping* ("Recovery of the Orbit Action Distribution of Mono-Abundance Populations and Potential INference for our Galaxy") - an improved and refined version of the original modelling machinery by Bovy & Rix (2013), making extensive use of the *galpy* python package (Bovy (2015)). *RoadMapping* relaxes some of the restraining assumptions Bovy & Rix (2013) had to made, is more flexible and more adept in dealing with large data sets. In this paper we set out to explore the robustness of *RoadMapping* against the breakdowns of some of the most important assumptions of DF-based dynamical modelling. What is it about the data, the model and the machinery itself, that limits our recovery of the true gravitational potential?

In the light of Gaia we explicitly analyze how well the modelling machinery behaves in the limit of large data. For a huge number of stars three statistical aspects become important, that are hidden behind Poisson noise for smaller data sets: (i) We have to make sure that our modelling is an un-biased and asymptotically normal estimator (§3.1). (ii) Numerical inaccuracies in the actual modelling machinery start to matter and need to be avoided (§??). (iii) Parameter estimates become so precise, that we start to be able to distinguish between similar models. We therefore want more flexibility and more free parameters in the potential and DF model. The modelling machinery itself needs to be flexible and fast in effectively finding the best fit parameters for a large set of parameters. The improvements made to the machinery used in Bovy & Rix (2013) are presented in §2.6.

Different characteristics of the data might influence the success of the parameter recov-

ery. (i) In an era where we can choose data from different MW surveys, it might be worth to explore if different regions within the MW (i.e. differently shaped or positioned survey volumes) are especially diagnostic to recover the potential (§??). (ii) What happens if our knowledge about the selection function, specifically the completeness of the data set within the survey volume, is not perfect (§3.3)? (iii) How to account for measurement errors in the modelling (§3.4)?

One of the strongest assumptions is to restrict the dynamical modelling to a certain family of parametrized models. We investigate how well we can we hope to recover the true potential, when our potential and DF models deviate from the true potential and DF. For the DF we specifically investigate two of our assumptions in §3.5: First, what would happen if the stars within *MAP*s do intrinsically not follow a single qDF as assumed by Ting et al. (2013); Bovy & Rix (2013). Second, and assuming *MAP*s do indeed follow the qDF, what would be the effect of pollution of *MAP*s through stars from neighbouring *MAP*s in the ([Fe/H],[α/Fe]) plane due to too big abundance errors or bin sizes.
And last but not least we test in §?? how well the modelling works, if our assumed potential family deviaties from the true potential.

too colloquial

For all of these aspects we show some plausible and illustrative examples on the basis of investigating mock data. The mock data is generated from galaxy models presented in §2.1-2.3 following the procedure in §2.4, analysed according to the description of the machinery in §2.5-2.6 and the results are presented in §3 and discussed in §4.

The strongest assumption that goes into this kind of dynamical modelling might be the idealization of the Galaxy to be axi-symmetric and being in steady state. We do not investigate this within the scope of this paper but strongly suggest a systematic investigation of this for future work.

2. Dynamical Modelling

2.1. Actions and Potential Models

Actions. Orbits in axisymmetric potentials are best described and fully specified by the three actions J_R, J_z and $J_\phi = L_z$. They are integrals of motion and generally defined as

$$J_i = \frac{1}{2\pi} \oint_{\text{orbit}} p_i \mathrm{d}x_i \qquad (1)$$

and depend on the potential via the connection between position x_i and momentum p_i along the orbit. Actions have a clear physical meaning: They quantify the amount of oscillation in each coordinate direction of the full orbit [REF]. The position of a star along the orbit is denoted by a set of angles, which form together with the angles a set of canonical conjugate phase-space coordinates (Binney & Tremaine 2008). Even though actions are the optimal choice as orbit labels and arguments for stellar distribution functions, their computation is very expensive.

Action calculation. The action calculation depends on the choice of potential in which the star moves: The spherical isochrone (Binney & Tremaine 2008) is the only potential for which Eq. (1) takes an analytic form. For axisymmetric Stäckel potentials actions can be calculated exactly by the (numerical) evaluation of a single integral. In all other potentials numerically calculated actions will always be approximations, unless Eq. (1) is integrated up to infinity. A computational fast way to get actions for arbitrary axisymmetric potentials is the "Stäckel fudge" by Binney (2012), which locally approximates the potential by a Stäckel potential. To speed up the calculation even more, an interpolation grid for J_R and J_z in energy E, angular momentum L_z and [TO DO: what else???] can be build out of these Stäckel fudge actions, as described in Bovy (2015).[1]

Potential models. In our modelling we assume a family of parametrized potentials with a fixed number of free parameters. We use different kinds of potentials: Besides the Milky Way like potential from Bovy & Rix (2013) ("MW13-Pot") with bulge, disk and halo, we also extensively use the spherical isochrone potential ("Iso-Pot") in our test suites to make use of the analytic (and therefore exact and fast) way to calculate actions. In addition we use the 2-component Kuzmin-Kutuzov Stäckel potential by Batsleer & Dejonghe (1994) ("KKS-Pot"), which displays a disk and halo structure and also provides exact actions. Table 1

[1] [TO DO: Write which numerical accuracy I needed for the grid, as the default values were not good enough.]

summarizes all reference potentials together used in this work with their free parameters p_Φ. The density distribution of these potentials is illustrated in Fig. 1.

2.2. Distribution Function

Distribution Function. Motivated by the findings of Bovy et al. 2012??? and Ting et al. (2013) about the simple phase-space structure of *MAP*s , and following Bovy & Rix (2013) and their successful application, we also assume that each *MAP* follows a single qDF of the form given by Binney & McMillan (2011). This qDF is a function of the actions $\boldsymbol{J} = (J_R, J_z, L_z)$ and has the form

$$\mathrm{qDF}(\boldsymbol{J} \mid p_{\mathrm{DF}}) = f_{\sigma_R}(J_R, L_z \mid p_{\mathrm{DF}}) \times f_{\sigma_z}(J_z, L_z \mid p_{\mathrm{DF}}) \quad (2)$$

$$\text{with } f_{\sigma_R}(J_R, L_z \mid p_{\mathrm{DF}}) = n \times \frac{\Omega}{\pi\sigma_R^2(R_g)\kappa}\left[1 + \tanh\left(L_z/L_0\right)\right] \exp\left(-\frac{\kappa J_R}{\sigma_R^2(R_g)}\right) \quad (3)$$

$$f_{\sigma_z}(J_z, L_z \mid p_{\mathrm{DF}}) = \frac{\nu}{2\pi\sigma_z^2(R_g)} \exp\left(-\frac{\nu J_z}{\sigma_z^2(R_g)}\right) \quad (4)$$

$$(5)$$

Here $R_g \equiv R_g(L_z)$ and $\Omega \equiv \Omega(L_z)$ are the (guidig-center) radius and the circular frequency of the circular orbit with angular momentum L_z in a given potential. $\kappa \equiv \kappa(L_z)$ and $\nu \equiv \nu(L_z)$ are the radial/epicycle (κ) and vertical (ν) frequencies with which the star would oscillate around the circular orbit in R- and z-direction when slightly perturbed (Binney & Tremaine 2008). The term $[1 + \tanh(L_z/L_0)]$ suppresses counter-rotation for orbits in the disk with $L \gg L_0$ which we set to a ~~random~~ small value ($L_0 = 10 \times R_\odot/8 \times v_{\mathrm{circ}}(R_\odot)/220$).
For this qDF to be able to incorporate the findings by Bovy et al. 2012??? about the

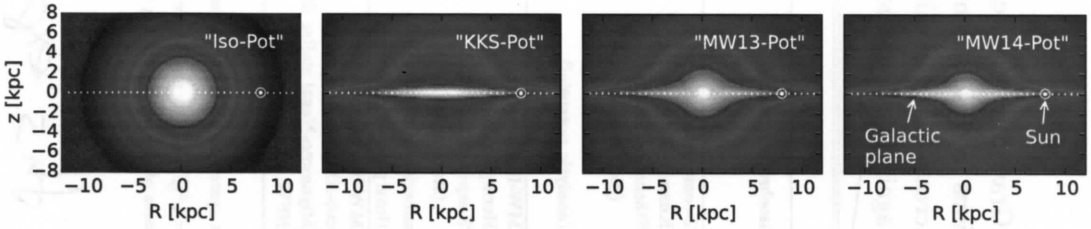

Fig. 1.— Density distribution of the four reference galaxy potentials in Table 1, for illustration purposes. These potentials are used throughout this work for mock data creation and potential recovery. [TO DO: Halo sichtbarer machen, evtl. mit isodensity contours]

Table 1. Gravitational potentials of the reference galaxies used troughout this work and the respective ways to calculate actions in these potentials. All four potentials are axisymmetric. The potential parameters are fixed for the mock data creation at the values given in this table. In the subsequent analyses we aim to recover these potential parameters again. The parameters of "MW13-Pot" and "KKS-Pot" were found as direct fits to the "MW14-Pot".

name	potential type	potential parameters p_Φ		action calculation	reference for potential type
"Iso-Pot"	isochrone potential	circular velocity at the sun isochrone scale length	$v_{\mathrm{circ}} = 230$ km s^{-1} $b = 0.9$ kpc	***analytical and exact*** J_r, J_ϑ, L_z; use $J_r \to J_R, J_\vartheta \to J_z$ in eq. (???)	Binney & Tremaine (2008)
"KKS-Pot"	2-component Kuzmin-Kutuzov- Stäckel potential (disk + halo) (analytic potential)	circular velocity at the sun focal distance of coordinate system[a] axis ratio of the coordinate surfaces[a] ... ...of the disk component ...of the halo component relative contribution of the disk mass to the total mass	$v_{\mathrm{circ}} = 230$ km s^{-1} $\Delta = 0.3$ $\left(\frac{a}{c}\right)_{\mathrm{Disk}} = 20$ $\left(\frac{a}{c}\right)_{\mathrm{Halo}} = 1.07$ $k = 0.28$	***exact*** J_R, J_z, L_z using "Stäckel Fudge" (Binney 2012) and interpolation on action grid (Bovy 2015)	Batsleer & Dejonghe (1994)
"MW13-Pot"	MW-like potential with Hernquist bulge, 2 exponential disks (stars + gas), spherical power-law halo (interpolated potential)	circular velocity at the sun stellar disk scale length stellar disk scale height relative halo contribution to $v^2_{\mathrm{circ}}(R_\odot)$ "flatness" of rotation curve	$v_{\mathrm{circ}} = 230$ km s^{-1} $R_d = 3$ kpc $z_h = 0.4$ kpc $f_h = 0.5$ $\frac{\mathrm{d}\ln(v_{\mathrm{circ}}(R_\odot))}{\mathrm{d}\ln(R)} = 0$	***approximate*** J_R, J_z, L_z using "Stäckel Fudge" (Binney 2012) and interpolation on action grid (Bovy 2015)	Bovy & Rix (2013)
"MW14-Pot"	MW-like potential with cut-off power-law bulge, Miyamoto-Nagai stellar disk, NFW halo	-	-	***approximate*** J_R, J_z, L_z (see "MW13-Pot")	Bovy (2015)

[a]The coordinate system of each of the two Stäckel-potential components is $\frac{R^2}{\tau_{i,p}+\alpha_p} + \frac{z^2}{\tau_{i,p}+\gamma_p} = 1$ with $p \in \{\mathrm{Disk}, /\mathrm{Halo}\}$ and $\tau_{i,p} \in \{\lambda_p, \nu_p\}$. Both components have the same focal distance $\Delta = \sqrt{\gamma_p - \alpha_p}$, to make sure that the superposition of the two components itself is still a Stäckel potential. The axis ratio of the coordinate surfaces $\left(\frac{a}{c}\right)_p := \sqrt{\frac{\alpha_p}{\gamma_p}}$ describes the flattness of the corresponding Stäckel component.

Is this explained anywhere? I'm surprised that $V_c = 230$ because MW14 has $V_c = 220$

Hénon 6X?

issue here

phase-space structure of *MAP*s summarized in §1, we set the functions n, σ_R and σ_z, which indirectly set the stellar number density and radial and vertical velocity dispersion profiles,

$$n(R_g \mid p_{\rm DF}) \propto \exp\left(-\frac{R_g}{h_R}\right) \tag{6}$$

$$\sigma_R(R_g \mid p_{\rm DF}) = \sigma_{R,0} \times \exp\left(-\frac{R_g - R_\odot}{h_{\sigma_R}}\right) \tag{7}$$

$$\sigma_z(R_g \mid p_{\rm DF}) = \sigma_{z,0} \times \exp\left(-\frac{R_g - R_\odot}{h_{\sigma_z}}\right). \tag{8}$$

The qDF for each *MAP* has therefore a set of five free parameters $p_{\rm DF}$: the density scale length of the tracers h_R, the radial and vertical velocity dispersion at the solar position $R_\odot$, $\sigma_R, 0$ and $\sigma_z, 0$, and the scale lengths h_{σ_R} and h_{σ_z}, that describe the radial decrease of the velocity dispersion. The *MAP*s we use for illustration through out this work are summarized in Table 2.

Tracer Density. One crucial point in our dynamical modelling technique (§???), as well as in creating mock data (§2.4), is to calculate the (axisymmetric) spatial tracer density $\rho_{\rm DF}(\boldsymbol{x} \mid p_\Phi, p_{\rm DF})$ for a given qDF and potential . We do this by integrating the qDF at a given (R, z) over all three velocity components, using a $N_{\rm velocity}$-th order Gauss-Legendre quadrature for each integral:

$$\rho_{\rm DF}(R, |z| \mid p_\Phi, p_{\rm DF}) = \int_{-\infty}^{\infty} {\rm qDF}(\boldsymbol{J}[R, z, \boldsymbol{v} \mid p_\Phi] \mid p_{\rm DF})\, {\rm d}^3\boldsymbol{v} \tag{9}$$

$$\approx \int_{-N_{\rm sigma}\sigma_R(R|p_{\rm DF})}^{N_{\rm sigma}\sigma_R(R|p_{\rm DF})} \int_{-N_{\rm sigma}\sigma_z(R|p_{\rm DF})}^{N_{\rm sigma}\sigma_z(R|p_{\rm DF})} \int_0^{1.5 v_{\rm circ}(R_\odot)} {\rm qDF}(J[R, z, \boldsymbol{v} \mid p_\Phi] \mid p_{\rm DF})\, {\rm d}v_T\, {\rm d}v_z\, {\rm d}v_R, \tag{10}$$

where $\sigma_R(R \mid p_{\rm DF})$ and $\sigma_z(R \mid p_{\rm DF})$ are given by eq. (7) and (8) and the integration ranges are motivated by Fig. 2. For a given p_Φ and $p_{\rm DF}$ we explicitly calculate the density on $N_{\rm spatial} \times N_{\rm spatial}$ regular grid points in the (R, z) plane; in between grid points the density is evaluated with a bivariate spline interpolation. The grid is chosen to cover the extent of the observations for $z > 0$. The total number of actions that need to be calculated to set up the density interpolation grid is $N^2_{\rm spatial} \cdot N^3_{\rm velocity}$. Fig. ??? shows the importance of choosing $N_{\rm spatial}$, $N_{\rm velocity}$ and $N_{\rm sigma}$ sufficiently large in order to get the density with an acceptable numerical accuracy.

2.3. Selection Function

Galactic Coordinate System. Our modelling takes place in the Galactocentric rest-frame with cylindrical coordinates $\boldsymbol{x} \equiv (R, \phi, z)$ and corresponding velocity components $\boldsymbol{v} \equiv (v_R, v_\phi, v_z)$. If the stellar phase-space data is given in observed coordinates, position $\tilde{\boldsymbol{x}} \equiv (\alpha, \delta, m - M)$ in right ascension α, declination δ and distance modulus $(m - M)$, and velocity $\tilde{\boldsymbol{v}} \equiv (\mu_\alpha, \mu_\delta, v_{\rm los})$ as proper motions $\boldsymbol{\mu} = (\mu_\alpha, \mu_\delta)$ [TO DO: cos somwhere???] and line-of-sight velocity $v_{\rm los}$, the data $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{v}})$ has to be converted first into the Galactocentric rest-frame coordinates $(\boldsymbol{x}, \boldsymbol{v})$ using the sun's position and velocity. For simplicity we assume for the sun

$$\begin{aligned} (R_\odot, \phi_\odot, z_\odot) &= (8 \text{ kpc}, 0°, 0 \text{ kpc}) \\ (v_{R,\odot}, v_{T,\odot}, v_{z,\odot}) &= (0, 230, 0) \text{ km s}^{-1}. \end{aligned}$$

Selection Function. A survey's selection function can be understood as a subvolume in the space of observables: e.g. position on the plane of the sky (limited by the pointing of the survey), distance from the sun (limited by the brightness of the stars and the sensitivity of the detector), colors and metallicity of the stars (limited by survey mode and targeting). Within the framework of this paper, using only mock data for testing purposes, we ignore target cuts in colors and metallicity and simply use spatial selection functions, which we define as

$$\text{sf}(\boldsymbol{x}) \equiv \begin{cases} \text{completeness}(\boldsymbol{x}) & \text{if } \boldsymbol{x} \text{ within observed volume} \\ 0 & \text{outside} \end{cases}$$

It's value describes the probability to observe a star at $\boldsymbol{x}$.
For the observed volume we use simple geometrical shapes: Either a sphere of radius $r_{\rm max}$ with the sun at its center, or a "wedge", which we define as the angular segment of an cylindrical annuli, i.e. the volume with $R \in [R_{\rm min}, R_{\rm max}], \phi \in [\phi_{\rm min}, \phi_{\rm max}], z \in [z_{\rm min}, z_{\rm max}]$ within the model galaxy. The sharp outer cut of the survey volume could be understood as the detection limit in apparent brightness in the case, where all stars have the same luminosity.
The completeness is, in our framework, a function of position with $0 \leq \text{completeness}(\boldsymbol{x}) \leq 1$ everywhere inside the observed volume. It could be understood as a position-dependent detection probability. Unless explicitly stated otherwise, we use everywhere

$$\text{completeness}(\boldsymbol{x}) = 1.$$

Table 2. Reference distribution-function parameters for the qDF in eq. (2)-(8). These qDFs describe the phase-space distribution of stellar *MAP*s for which mock data is created and analysed throughout this work for testing purposes. The parameters of the "cooler" & "colder" ("hotter" & "warmer") *MAP*s were chosen such that the they have the same σ_R/σ_z ratio as the "hot" ("cool") *MAP* . The "colder" and "warmer" *MAP*s have a free parameter X that governs how much colder/warmer they are then the reference "hot" and "cool" qDFs. Hotter populations have shorter tracer scale lengths (Bovy et al. 2012d) and the velocity dispersion scale lengths were fixed according to Bovy et al. (2012c).

name of *MAP*	qDF parameters $p_{\rm DF}$				
	h_R [kpc]	σ_R [km s^{-1}]	σ_z [km s^{-1}]	h_{σ_R} [kpc]	h_{σ_z} [kpc]
"hot"	2	55	66	8	7
"cool"	3.5	42	32	8	7
"cooler"	2 +50%	55-50%	66-50%	8	7
"hotter"	3.5-50%	42+50%	32+50%	8	7
"colder"	2 +X%	55-X%	66-X%	8	7
"warmer"	3.5-X%	42+X%	32+X%	8	7

2.4. Mock Data

One goal of this work is to test how the loss of information in the process of measuring stellar phase-space coordinates can affect the outcome of the modelling. To investigate this, we assume first that our measured stars do indeed come from our assumed families of potentials and distribution functions and draw mock data from a given true distribution. In further steps we can manipulate and modify these mock data sets to mimick observational effects.

The distribution function is given in terms of actions and angles. The transformation $(\boldsymbol{J}_i, \boldsymbol{\theta}_i) \longrightarrow (\boldsymbol{x}_i, \boldsymbol{v}_i)$ is however difficult to perform and computationally much more expensive than the transformation $(\boldsymbol{x}_i, \boldsymbol{v}_i) \longrightarrow (\boldsymbol{J}_i, \boldsymbol{\theta}_i)$. We propose a fast and simple two-step method for drawing mock data from an action distribution function, which also accounts effectively for a given survey selection function.

Preparation: Tracer density. We first setup the interpolation grid for the tracer density $\rho(R, |z| \mid p_\Phi, p_{\mathrm{DF}})$ generated by the given qDF and according to §2.2 and Eq. 10. For the creation of the mock data we use $N_{\mathrm{spatial}} = 20$, $N_{\mathrm{velocity}} = 40$ and $N_{\mathrm{sigma}} = 5$.

Step 1: Drawing positions from the selection function. To get positions $\boldsymbol{x}_i$ for our mock data stars, we first sample random positions (R_i, z_i, ϕ_i) uniformly from the observed volume. Then we apply a rejection Monte Carlo method to these positions using the precalculated $\rho_{\mathrm{DF}}(R, |z| \mid p_\Phi, p_{\mathrm{DF}})$. In an optional third step, if we want to apply a non-uniform selection function, $\mathrm{sf}(\boldsymbol{x}) \neq$ const. within the observed volume, we use the rejection method a second time. The sample then follows

$$\boldsymbol{x}_i \longrightarrow p(\boldsymbol{x}) \propto \rho_{\mathrm{DF}}(R, z \mid p_\Phi, p_{\mathrm{DF}}) \times \mathrm{sf}(\boldsymbol{x}).$$

Step 2: Drawing velocities according to the distribution function. The velocities are independent of the selection function and observed volume. For each of the positions (R_i, z_i) we now sample velocities directly from the $\mathrm{qDF}(R_i, z_i, \boldsymbol{v} \mid p_{\Phi,i}, p_{\mathrm{DF}})$ using a rejection method. To reduce the number of rejected velocities, we use a Gaussian in velocity space as an envelope function, from which we first randomly sample velocities and then apply the rejection method to shape the Gaussian velocity distribution towards the velocity distribution predicted by the qDF. We now have a mock data set according to the required:

$$(\boldsymbol{x}_i, \boldsymbol{v}_i) \longrightarrow p(\boldsymbol{x}, \boldsymbol{v}) \propto \mathrm{qDF}(\boldsymbol{x}, \boldsymbol{v} \mid p_\Phi, p_{\mathrm{DF}}) \times \mathrm{sf}(\boldsymbol{x}).$$

Example: Fig. 2 shows examples of mock data sets in configuration space $(\boldsymbol{x}, \boldsymbol{v})$ and action space. The qDF represents realistic stellar distributions in position-velocity space: More stars are found at smaller R and $|z|$, and are distributed uniformly in ϕ according to our assumption of axisymmetry. The distribution in radial and vertical velocities, v_R and v_z, is approximately Gaussian with the (total projected) velocity dispersion being $\sim \sigma_{R,0}$ and $\sim \sigma_{z,0}$ (see Table 2). The distribution of tangential velocities v_T is skewed because of asymmetric drift [TO DO: Find out, if we need an explanation for asymmetric drift here] No!
The distribution in action space demonstrates the intuitive physical meaning of actions: The stars of the "cool" *MAP* have in general lower radial and vertical actions, as they are on more circular orbits. The different relative distributions of the radial and vertical actions J_R and J_z of the "hot" and "cool" *MAP* is due to them having different velocity anisotropy $\sigma_{R,0}/\sigma_{z,0}$. The different ranges of angular momentum L_z in the two volumes reflect $L_z \sim R v_{\rm circ}$ and the different radial extent of both volumes. The volume above the plane contains more stars with higher J_z, because stars with small J_z can't reach that far above the plane. Circular orbits with $J_R = 0$ and $J_z = 0$ can only be observed in the Galactic mid-plane. An orbit with L_z much smaller or larger than $L_z(R_\odot)$ can only reach into a volume located around $R_\odot$, if it is more eccentric and has therefore larger J_R. This together with the effect of asymmetric drift can be seen in the asymmetric distribution of J_R in the top central panel of Fig. 2. [TO DO: Part of this could also be mentioned in the figure caption.] Better here, bc otherwise very long caption!

~~Introducing measurement~~ errors. If we want to add measurement errors to the mock data, we need to apply two modifications to the above procedure.
First, ~~measurement errors~~ are best described in the phase-space of observables. We use the heliocentric coordinate system right ascension and declination (α, δ) and distance modulus $(m-M)$ as proxy for the distance from the sun, the proper motion in both α and δ direction (μ_α, μ_δ) and the line-of-sight velocity $v_{\rm los}$. For the conversion between these observables and the Galactocentric cylindrical coordinate system in which the analysis takes place, we need the position and velocity of the sun, which we set for simplicity in this study to be $(R_\odot, z_\odot) = (8, 0)$ kpc and $(v_R, v_T, v_z) = (0, 230, 0)$ km s^{-1}. We assume Gaussian measurement errors in the observables $\tilde{\boldsymbol{x}} = (\alpha, \delta, (m-M))$, $\tilde{\boldsymbol{v}} = (\mu_\alpha, \mu_\delta, v_{\rm los})$. no checked
Second, in the case of distance errors, stars can virtually scatter in and out of the observed volume. To account for this, we first draw "true" positions from a volume that is larger than the actual observation volume, perturb the stars positions according to the distance errors and then reject all stars that lie now outside of the observed volume. This procedure mirrors the Poisson scatter around the detection threshold for stars whose distances are determined from the apparent brightness and the distance modulus. [TO DO: Can I say it like this???]
We then sample velocities (given the "true" positions of the stars) as described above and

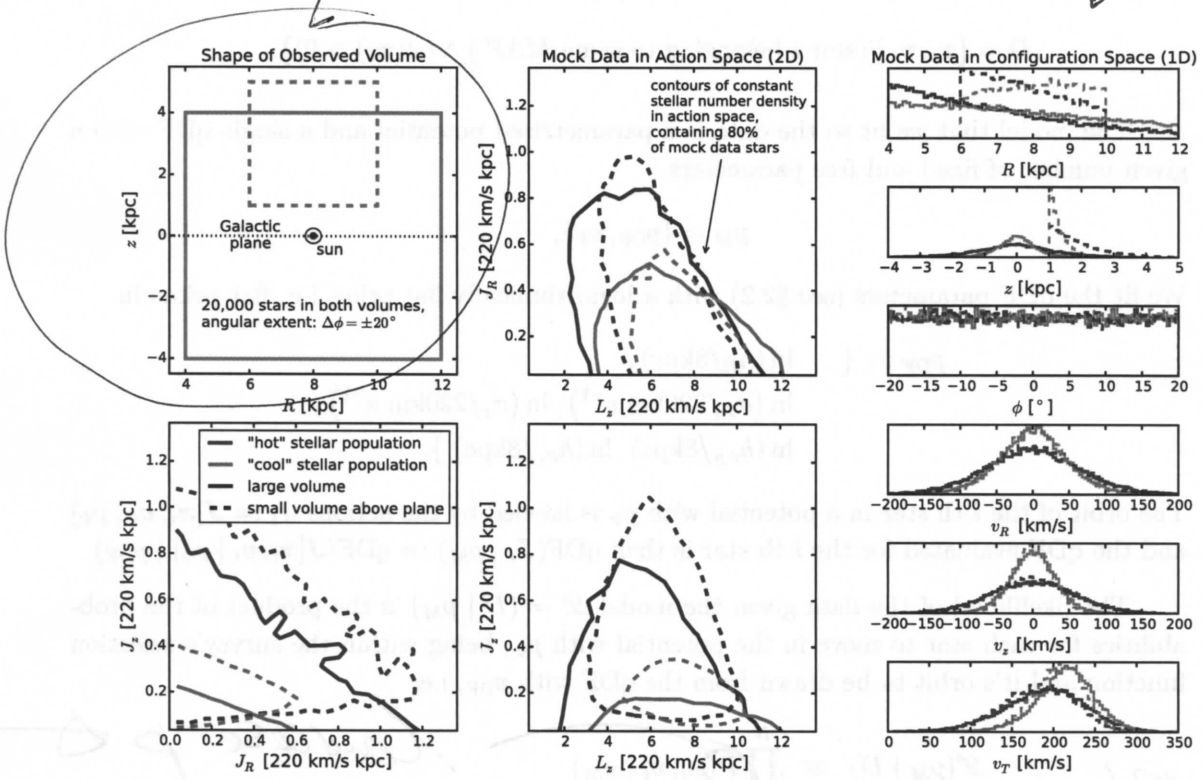

Fig. 2.— Distribution of mock data in action space (2D iso-density contours enclosing 80% of the stars, in the two central and the lower left panel) and configuration space (1D histograms in right panels), depending on shape and position of the survey observation volume and temperature of the stellar population. The parameters of the mock data model is given as Test ① in Table 3. In the upper left panel we demonstrate the shape of the two different "wedge"-like observation volumes within which we were creating each a "hot" (red) and "cool" (blue) mock data set: a large volume centered on the Galactic plane (solid lines) and a smaller one above the plane (dashed lines).

2.5. Likelihood

Form of the likelihood. As data we use the positions and velocities of stars coming from a given *MAP* and survey selection function $\mathrm{sf}(\boldsymbol{x})$,

$$D = \{\boldsymbol{x}_i, \boldsymbol{v}_i \mid (\text{star } i \text{ belonging to same } MAP\,) \wedge (\mathrm{sf}(\boldsymbol{x_i}) > 0)\}.$$

The model that we fit to the data is a parametrized potential and a single qDF with a given number of fixed and free parameters,

$$p_M = \{p_{\mathrm{DF}}, p_\Phi\},$$

We fit the qDF parameters (see §2.2) with a logarithmically flat prior, i.e. flat priors in

$$\begin{aligned} p_{\mathrm{DF}} := \{ \quad & \ln\left(h_R/8\mathrm{kpc}\right), \\ & \ln\left(\sigma_R/220\mathrm{km\ s^{-1}}\right), \ln\left(\sigma_z/220\mathrm{km\ s^{-1}}\right), \\ & \ln\left(h_{\sigma_R}/8\mathrm{kpc}\right), \ln\left(h_{\sigma_z}/8\mathrm{kpc}\right)\ \}. \end{aligned}$$

The orbit of the i-th star in a potential with p_Φ is labeled by the actions $\boldsymbol{J}_i := \boldsymbol{J}[\boldsymbol{x}_i, \boldsymbol{v}_i \mid p_\Phi]$ and the qDF evaluated for the i-th star is then $\mathrm{qDF}(\boldsymbol{J}_i \mid p_M) := \mathrm{qDF}(\boldsymbol{J}[\boldsymbol{x}_i, \boldsymbol{v}_i \mid p_\Phi] \mid p_{\mathrm{DF}})$.

The likelihood of the data given the model $\mathscr{L} = (D \mid p_M)$ is the product of the probabilities for each star to move in the potential with p_Φ, being within the survey's selection function and it's orbit to be drawn from the qDF with p_{DF}, i.e.

$$\begin{aligned} \mathscr{L}(p_M \mid D) \;&\equiv\; \prod_i^N P(\boldsymbol{x}_i, \boldsymbol{v}_i \mid p_M) \\ &= \prod_i^N \frac{1}{(r_o v_o)^3} \cdot \frac{\mathrm{qDF}(\boldsymbol{J}_i \mid p_M) \cdot \mathrm{sf}(\boldsymbol{x}_i)}{\int \mathrm{d}^3x\, \mathrm{d}^3v\ \mathrm{qDF}(\boldsymbol{J} \mid p_M) \cdot \mathrm{sf}(\boldsymbol{x})} \\ &\propto \prod_i^N \frac{1}{(r_o v_o)^3} \cdot \frac{\mathrm{qDF}(\boldsymbol{J}_i \mid p_M)}{\int \mathrm{d}^3x\ \rho_{\mathrm{DF}}(R, |z| \mid p_M) \cdot \mathrm{sf}(\boldsymbol{x})}, \end{aligned} \tag{11}$$

where N is the number of stars in the data set D. In the last step we used eq. (9). The factor $\prod_i \mathrm{sf}(\boldsymbol{x}_i)$ is independent of the model parameters, so we simply evaluate Eq. (11) in the likelihood calculation. We find the best set of model parameters by maximising the likelihood.

A word on units. We evaluate the likelihood in a scale-free potential within a Galactocentric coordinate system which is defined as $v_{\mathrm{circ}}(R = 1) = 1$. The circular velocity at the

sun's radius, $v_{\rm circ}(R_\odot = 8.{\rm kpc}) \sim 230{\rm km\ s}^{-1}$, determines the total mass amplitude of the galaxy potential. In the modelling all data and model parameters are re-scaled to spatial units of $r_o := R_\odot$ or velocitiy units of $v_o := v_{\rm circ}(R_\odot)$. The prefactor $1/\left(r_o v_o\right)^3$ in eq. (11) makes sure that the likelihood has the correct units to satisfy:

$$\int P(\boldsymbol{x}, \boldsymbol{v} \mid p_M)\, {\rm d}^3x\, {\rm d}^3v \propto 1$$

Including this prefactor is crucial when $v_{\rm circ}(R_\odot)$ is a free fitting parameter.

Numerical accuracy in calculating the likelihood. The normalisation in Eq. (11) is a measure for the total number of tracers inside the survey volume,

$$M_{\rm tot} \equiv \int {\rm d}^3x\ \rho_{\rm DF}(R, |z| \mid p_{model}) \cdot {\rm sf}(\boldsymbol{x}). \tag{12}$$

In the case of an axisymmetric galaxy model and ${\rm sf}(\boldsymbol{x}) = 1$ everywhere inside the observed volume (i.e. a complete sample as assumed in most tests in this work), the normalisation is essentially a two-dimensional integral in R and z of the interpolated tracer density ρ_{DF} (see Eq. (10) and surrounding text) over the survey volume times the observation volume's geometric angular contribution at each (R, z). We perform this integral as a Gauss Legendre quadrature of order 40 in each R and z direction.
Unfortunately the evaluation of the likelihood for only one set of model parameters is computationally expensive. The computation speed is set by the number of action calculations required, i.e. the number of stars and the numerical accuracy of the integrals in Eq. ??? needed for the normalisation, which requires $N^2_{\rm spatial} \times N^3_{\rm velocity}$ action calculations. The accuracy has to be chosen high enough, such that a resulting numerical error

$$\delta_{M_{tot}} \equiv \frac{M_{\rm tot}(N_{\rm spatial}, N_{\rm velocity}, N_{\rm sigma}) - M_{\rm tot,true}}{M_{\rm tot,true}} \tag{13}$$

does not dominate the likelihood, i.e.

$$\begin{aligned} \log \mathscr{L}(p_M \mid D) \;&=\; \sum_i^N \log {\rm qDF}(\boldsymbol{J_i} \mid p_M) - 3N \log\left(r_o v_o\right) \\ &\quad -N \log(M_{\rm tot,true}) - N \log(1 + \delta_{M_{tot}}), \\ &\quad \text{with} \qquad N \log(1 + \delta_{M_{tot}}) \lesssim 1. \end{aligned} \tag{14}$$

In other words, this error is only small enough, if it does not affect the comparison of two adjacent models whose likelihoods differ, to be clearly distinguishable, by a factor of 10. Otherwise numerical inaccuracies could lead to systematic biases in the potential and DF

fitting. For data sets as large as $N = 20{,}000$ stars in one *MAP* , which in the age of GAIA could very well be the case [TO DO: Really???], we would need a numerical accuracy of 0.005% in the normalisation. Fig. 3 demonstrates that the numerical accuracy we use in the analysis, $N_{\rm spatial} = 16$, $N_{\rm velocity} = 24$ and $N_{sigma} = 5$, does satisfy this requirement.

Dealing with measurement errors. We assume Gaussian errors in the observable space $\boldsymbol{y}_i \equiv (\tilde{\boldsymbol{x}}_i, \tilde{\boldsymbol{v}}_i) = (\alpha, \delta, (m-M), \mu_\alpha, \mu_\delta, v_{\rm los})$,

$$N[\boldsymbol{y}_i, \sigma_{y,i}](\boldsymbol{y}') = N[\boldsymbol{y}', \sigma_{y,i}](\boldsymbol{y}_i) \equiv \prod_k \frac{1}{\sqrt{2\pi\sigma_{y,k}^2}} \exp\left(-\frac{(y_{i,k} - y'_k)^2}{2\sigma_{y,k}^2}\right),$$

where $y_{i,k}$ are the coordinate components of $\boldsymbol{y}_i$. Observed stars follow the (quasi-isothermal) distribution function ($\mathrm{DF}(\boldsymbol{y}) \equiv \mathrm{qDF}(\boldsymbol{J}[\boldsymbol{y} \mid p_\Phi] \mid p_{\rm DF}]$) for short), convolved with the error distribution $N[0, \sigma_y](\boldsymbol{y})$. The selection function $\mathrm{sf}(\boldsymbol{y})$ acts on the space of (error affected) observables. Then the probability of one star coming from potential p_Φ, distribution function $p_{\rm DF}$ and being affected by the measurement errors $\sigma_{\boldsymbol{y}}$ becomes

$$\tilde{P}(\boldsymbol{y}_i \mid p_\Phi, p_{\rm DF}, \sigma_{\boldsymbol{y},i}) \equiv \frac{\mathrm{sf}(\boldsymbol{y}_i) \cdot \int \mathrm{d}^6 y' \ \mathrm{DF}(\boldsymbol{y}') \cdot N[\boldsymbol{y}_i, \sigma_{\boldsymbol{y},i}](\boldsymbol{y}')}{\int \mathrm{d}^6 y \ \mathrm{DF}(\boldsymbol{y}) \cdot \int \mathrm{d}^6 y' \ \mathrm{sf}(\boldsymbol{y}') \cdot N[\boldsymbol{y}, \sigma_{\boldsymbol{y},i}](\boldsymbol{y}')}.$$

In the case of errors in distance or position, the evaluation of this is computational expensive - especially if the stars' have heteroscedastic errors $\sigma_{\boldsymbol{y},i}$, for which the normalisation would have to be calculated for each star separately. In practice we apply the following approximation:

$$\tilde{P}(\boldsymbol{y}_i \mid p_\Phi, p_{\rm DF}, \sigma_{\boldsymbol{y},i}) \approx \frac{\mathrm{sf}(\boldsymbol{x}_i)}{\int \mathrm{d}^6 y \ \mathrm{DF}(\boldsymbol{y}) \cdot \mathrm{sf}(\boldsymbol{x})} \cdot \frac{1}{N_{\rm error}} \sum_n^{N_{\rm error}} \mathrm{DF}(\boldsymbol{x}_i, \boldsymbol{v}[\boldsymbol{y}'_{i,n}]) \tag{15}$$

$$\text{with } \boldsymbol{y}'_{i,n} \sim N[\boldsymbol{y}_i, \sigma_{\boldsymbol{y},i}](\boldsymbol{y}') \tag{16}$$

In doing so, we ignore errors in the star's position $\boldsymbol{x}_i$ altogether. This simplifies the normalisation drastically and makes it independent of measurement errors, including the velocity errors. Distance errors however are included, but only implicitly in the convolution over the stars' velocity errors in the Galactocentric restframe. We calculate the convolution using Monte Carlo integration with $N_{\rm error}$ samples drawn from the full error Gaussian in observable space, $y'_{i,n}$.

Should we show |error| on a logarithmic scale?

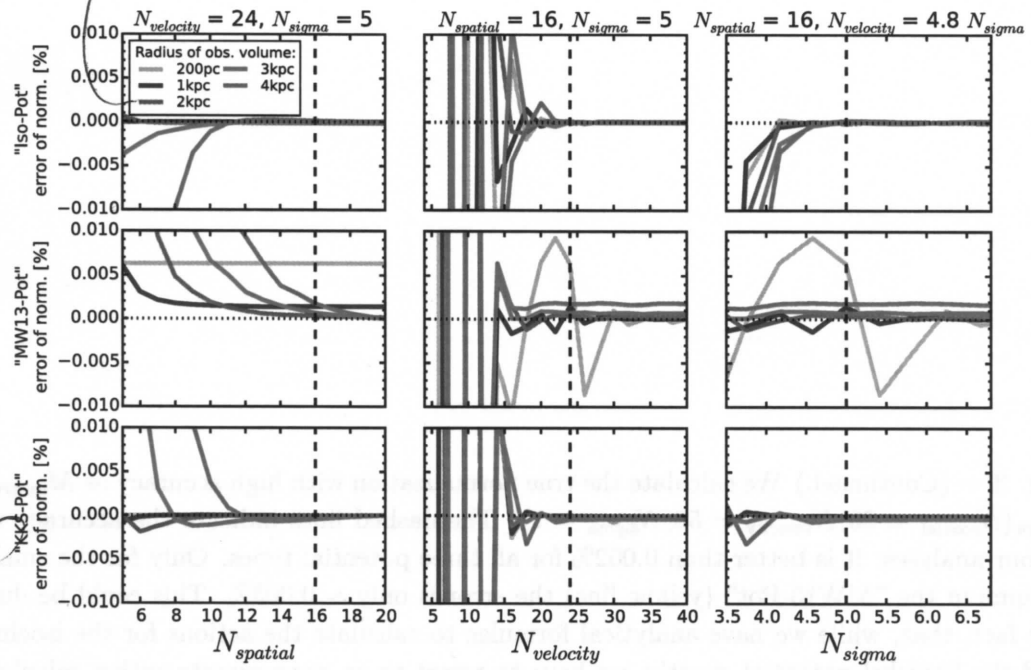

Fig. 3.— Relative error of the likelihood normalization in Eq. (13) depending on the accuracy of the density calculation in Eq. (10) (and sourrounding text). The different colors represent calculations for different radii of the spherical observation volume around the sun, as indicated in the legend. $N_{\rm spatial}$ is the number of regular grid points in each R and $z > 0$ within the observed volume on which the tracer density is evaluated according to Eq. (10). At each (R, z) a Gauss-Legendre integration of order $N_{\rm velocity}$ is performed over an integration range of $\pm N_{\rm spatial}$ times the dispersion in v_R and v_z and $[0, 1.5 v_{\rm circ}(R_\odot)]$ [TO DO: update if required] in v_T. To integrate the interpolated density over the observed volume to arrive at the likelihood normalization in Eq. (12), we perform a 40th-order Gauss-Legendre integration in each R and z direction. We compare the convergence of the normalisation for the "hot" qDF in three potentials, "Iso-Pot", "MW13-Pot" and "KKS-Pot" (see also test ⑨ in Table 3 for all other model details). In each column of plots we keep two of the accuracy parameters fixed (indicated on top), while the third parameter is varied. (Caption continues on next page.)

The different volumes make this figure difficult to understand. It seems like the distance resolution is ~~not~~ important in the left panel? That is, radius/$N_{spatial}$

2.6. Fitting Procedure

We search the $(p_\Phi, p_{\rm DF})$ parameter space for the maximum of the likelihood in Eq. (11) using a two-step procedure: The first step finds the approximate peak and width of the likelihood using a nested-grid search, while the second step samples the shape of the likelihood (or rather the posterior probability distribution) using a Monte-Carlo Markov Chain (MCMC) approach.

Fitting Step 1: Nested-grid search. The $(p_\Phi, p_{\rm DF})$ parameter space can be high-dimensional. To effectively minimizing the number of likelihood evaluations before finding its peak, we use a nested-grid approach:

- *Initialization.* For N free model parameters $M = (p_\Phi, p_{\rm DF})$, we set up a sufficiently large initial grid with 3^N regular grid points.[2]

- *Evaluation.* We evaluate the likelihood at each grid-point. Because of the many computationally expensive $\boldsymbol{x}, \boldsymbol{v} \xrightarrow{p_\Phi} \boldsymbol{J}$ transformations that have to be performed for each new set of p_Φ parameters, an outer loop iterates over the p_Φ parameters and pre-calculates the actions, while an inner loop evaluates the likelihood Eq. (11) for all qDF parameters $p_{\rm DF}$ with the actions in the given potential and (analogously to Fig. 9 in Bovy & Rix (2013)).

- *Iteration.* To find from the very sparse 3^N likelihood grid a new grid, that is more centered on the likelihood and has a width of order of the width of the likelihood, we proceed as follows: For each of the model parameter in M we marginalize the likelihood by summing over the grid. If the resulting 3 points all lie within 4σ of a Gaussian, we fit a Gaussian to the 3 points and determine a new 4σ fitting range. Otherwise the grid point with the highest likelihood becomes the new fitting range. We proceed with iteratively evaluating the likelihood on finer and finer grids, until we have found a 4-sigma fit range in each of the model parameter dimensions.

- *The fiducial qDF.* For the above strategy to work properly, the action pre-calculations have to be independent of the choice of qDF parameters. This is clearly the case for the $N_j \times N_{\rm error}$ [TO DO: explain $N_{\rm error}$???] stellar data actions $\boldsymbol{J}_i$. To calculate

[2]To get a better feeling where in parameter space the true $p_{\rm DF}$ parameters lie, we fit eq. (???) directly to the data. This gives a very good inital guess for $\sigma_{R,0}$ and $\sigma_{z,0}$. To improve the estimate for h_R, we fit eq. (???) only to stars within a thin wedge around $(R = 0, z = 0)$ and then apply the relation in fig. 5 in Bovy & Rix (2013) between the stars' measured scale length $h_R^{\rm out}$ and the qDF tracer scale length $h_R^{\rm in} = h_R$.

the normalisation in Eq. (11), $N^2_{\rm spatial} \times N^3_{\rm velocity}$ actions $\boldsymbol{J}_n$ are needed. Formally the spatial coordinates at which the $\boldsymbol{J}_n$ are calculated depend on the $p_{\rm DF}$ parameters via the integration ranges in eq. (10). To relax this dependence we instead use the same velocity integration limits in the likelihood calculations for all $p_{\rm DF}$s in a given potential. This set of parameters, that sets the velocity integration range globally, $(\sigma_{R,0}, \sigma_{z,0}, h_{\sigma_R}, h_{\sigma_z})$ in Eq. (???), is referred to as the "fiducial qDF". Using the same integration range in the density calculation for all qDFs at a given p_Φ makes the normalisation vary smoothly with different $p_{\rm DF}$. Choosing a fiducial qDF that is very off from the true qDF can however lead to large biases. The optimal values for the fiducial qDF are the (yet unknown) best fit $p_{\rm DF}$ parameters. We take care of this by setting, in each iteration step of the nested-grid search, the fiducial qDF simply to the $p_{\rm DF}$ parameters of the central grid point. As the nested-grid search approaches the best fit values, the fiducial qDF approaches automatically the optimal values as well. This is another advantage of the nested-grid search, because the result will not be biased by a poor choice of the fiducial qDF.

- *Speed Limitations.* Overall the computation speed of this nested-grid approach is dominated (in descending order of importance) by a) the complexity of potential and action calculation, b) the number $N_j \times N_{\rm error} + N^2_{\rm spatial} \times N^3_{\rm velocity}$ of actions to calculate, i.e. the number of stars, error samples and numerical accuracy of the normalisation calculations, c) the number of different potentials to investigate (i.e. the number of free potential parameters and number of grid points in each dimension) and d) the number of qDFs to investigate. The latter is also non-negligible, because for such a large number of actions the number of qDF-function evaluations also take some time.

Fitting Step 2: MCMC. After the nested-grid search is converged, the grid is centered at the peak of the likelihood and it's extent contains the 4σ confidence interval. To actually sample the full shape of the likelihood, we could do a grid search with much finer grid spacing (e.g. $K = 11$ in each dimension). The number of grid points scales exponentially with number of free parameters N. For a large number of free parameters ($N > 4$) a Monte Carlo Markov Chain (MCMC) approach might sample the likelihood (or rather the posterior probability distribution, which is the likelihood times some priors, see §????) much faster. We use *emcee* by ? and release the walkers very close to the likelihood peak found by the nested-grid search, which will assure fast convergence in much less than K^N likelihood evaluations.

For a sufficiently high numerical accuracy in calculating the integrals in Eq. (10) the current qDF parameters as each values can be used as integration ranges. To get reasonable results also for slightly lower accuracy, a single fiducial qDF can be used for all likelihood evaluations

3. Results

We are now in a position to explore the questions about the ultimate limitations of action-based modelling, posed in the introduction:

- Can we still retrieve unbiased model parameter estimates p_M in the limit of large sample sizes?
- What role does the survey volume and geometry play, at given sample size?
- What if our knowledge of the sample selection function is imperfect, and potentially biased?
- How do the parameter estimates deteriorate if the individual errors on the phase-space coordinates become significant?

But we also consider the more fundamental limitations:

- What if the observed stars are not extacly drawn from the family of model distribution functions?
- What happens to the estimate of the potential and the DF, if the actual potential is not contained in the family of model potentials?

We do not explore the breakdown of the assumption that the system is axisymmetric and in steady state. ~~Except~~ of the test suite on measurement errors in §3.4, we assume that the phase-space errors are negligible.

Table 3. Summary of test suites in this work: The first column indicates the test suite, the second column the potential, DF and selection function model etc. used for the mock data creation, the third model the corresponding model assumed in the analysis, and the last column lists the figures belonging to the test suite. Parameters that are not left free in the analyis, are always fixed to their true value. Unless otherwise stated we calculate the likelihood by the nested-grid and MCMC approach outlined in §2.6 and use $N_{\rm spatial} = 16$, $N_{\rm velocity} = 24$, $N_{\rm sigma} = 5$ as numerical accuracy for the likelihood normalisation in Eq. ???. [TO DO: Change encircled numbers to proper order. Make sure the plot references are the right ones.]

Test		Model for Mock Data	Model in Analysis	Figures
①	*Potential:*	"KKS-Pot"	-	Mock data:
Influence of	*MAP :*	2 *MAP*s "hot" or "cold" qDF		Fig. 2
survey volume on	*Survey volume:*	a) $R \in [4, 12]$ kpc,$z \in [-4, 4]$ kpc, $\phi \in [-20°, 20°]$.		
mock data distribution,		b) $R \in [6, 10]$ kpc,$z \in [1, 5]$ kpc, $\phi \in [-20°, 20°]$.		
also in action space	*# stars per data set:*	20,000		
	# data sets:	4 (= 2 × 2 models)		
⑨	*Potential:*	"Iso-Pot", "MW13-Pot" & "KKS-Pot"	-	Convergence
Numerical accuracy	*MAP :*	"hot" qDF		of normalisation:
in calculation	*Survey volume:*	sphere around sun, $r_{\rm max} = 0.2, 1, 2, 3$ or 4 kpc		Fig. 3
of the likelihood	*Numerical accuracy:*	$N_{\rm spatial} \in [5, 20]$, $N_{\rm velocity} \in [6, 40]$, $N_{\rm sigma} \in [3.5, 7]$		
normalisation				
⑩	*Potential:*	"Iso-Pot"	"Iso-Pot", all parameters free	Fig. 4
*pdf*is a	*MAP :*	"hot" qdf	qDF, all parameters free	
multivariate	*Survey Volume:*	sphere around sun, $r_{\rm max} = 2$ kpc	(fixed & known)	
Gaussian	*# stars per data set:*	20,000		
for large data sets.	*# data sets:*	5 (only one is shown)		
	Numerical accuracy:		$N_{\rm velocity} = 20$ and $N_{\rm sigma} = 4$	
②	*Potential:*	"Iso-Pot"	"Iso-Pot", free parameter: b	Fig. 5
Width of the	*MAP :*	"hot" qDF	"hot" qDF, free parameters:	
likelihood scales			$\ln\left(\frac{h_R}{8\rm kpc}\right), \ln\left(\frac{\sigma_R}{230\rm km\ s^{-1}}\right), \ln\left(\frac{h_{\sigma,R}}{8\rm kpc}\right)$	
with number of stars	*Survey volume:*	sphere around sun, $r_{\rm max} = 3$ kpc	(fixed & known)	
by $\propto 1/\sqrt{N}$.	*# stars per data set:*	between 100 and 40,000		
	# data sets:	132		
	Analysis method:		likelihood on grid	
	Numerical accuracy:		$N_{\rm velocity} = 20$ and $N_{\rm sigma} = 4$ (for speed)	
③	*Potential:*	2 "Iso-Pot" with	"Iso-Pot", free parameter: b	Fig. 6
Parameter estimates		$b = 0.8$ kpc or $b = 1.5$ kpc		
are unbiased.	*MAP :*	2 *MAP*s , "hot" or "cool" qDF	"hot"/"cool" qDF, free parameters:	
			$\ln\left(\frac{h_R}{8\rm kpc}\right), \ln\left(\frac{\sigma_R}{230\rm km\ s^{-1}}\right), \ln\left(\frac{h_{\sigma,R}}{8\rm kpc}\right)$	
	Survey volume:	5 spheres around sun, $r_{\rm max} = 0.2, 1, 2, 3$ or 4 kpc	(fixed & known)	
	# stars per data set:	20,000		
	# data sets:	640 (= 2 × 2 × 5 models ×32 realisations)		
	Analysis method:		likelihood on grid	

3.1. Model parameter estimates in the limit of large data sets

The individual *MAP* in Bovy & Rix (2013) contained typically between 100 and 800 objects, so that each *MAP* implied a quite broad *pdf* for the model parameters $p_M = \{p_\Phi, p_{\rm DF}\}$. Here we explore what happens in the limit of very much larger samples for each *MAP*, say 20,000 objects. As outlined in §2.5 the immediate consequence of larger samples is given by the likelihood normalization requirement, $\log(1 + rel.error) \leq 1/N_{sample}$, (see Eq. 14), which is the modelling aspect that drives the computing time. This issues aside, we would, however, expect that in the limit of large data sets with vanishing measurement errors the *pdf*s of the p_M become Gaussian, with a *pdf* width (i.e. standard error SE of the Gaussian) that scales as $1/N_{\rm sample}$. Further, we must verify that any bias in the *pdf* expectation value is far less than SE, even for quite large samples.

Using sets of mock data, created according to §2.4 and with our fiducial model for p_M in Table 3, Tests ②, ③ and ⑩, we verified that *RoadMapping* satisfies all these conditions and expectations. Fig. 4 illustrates the joint *pdf*'s of all p_M. This figure illustrates that the *pdf*'s are multivariate Gaussians that project into Gaussians when considering the marginalized *pdf* for all the individual p_M. Note that some of the parameters are quite covariant, but the level of their actual covariance depends on the choice of the p_M from with the mock data were drawn. Figure5 then illustrates that the *pdf* width, SE, indeed scales as $1/N_{\rm sample}$. Fig.6 illustrates even more, that *RoadMapping* satisfies the central limit theorem. The average parameter estimates from many mock samples with identical underlying p_M are very close to the input p_M, and the distribution of the actual parameter estimates are a Gaussian around it.

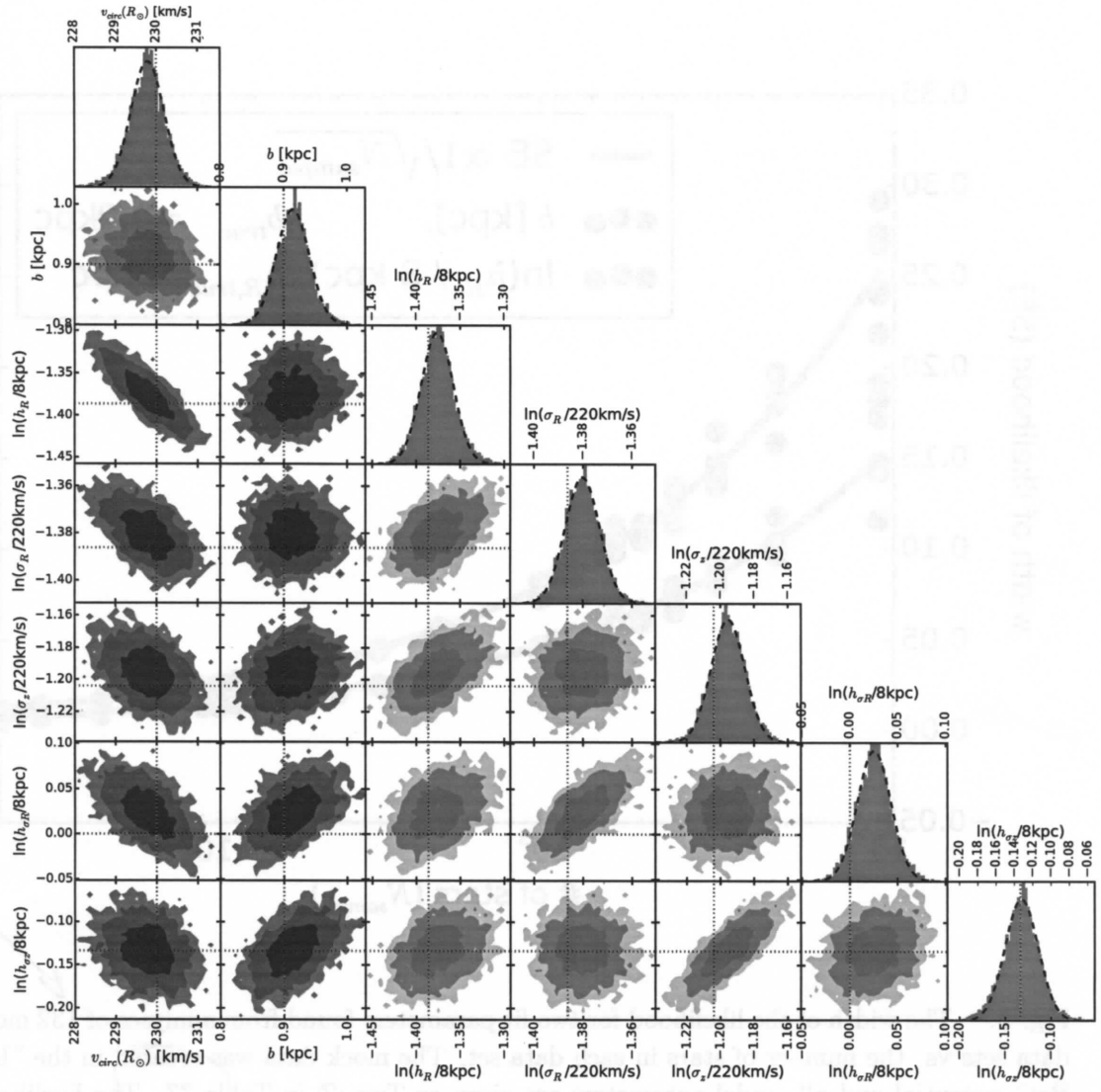

Fig. 4.— The likelihood in Eq. (11) in the parameter space $p_M = \{p_\Phi, \ln(p_{\rm DF})\}$ for one example mock data set created according to Test 10 in Table 3. Blue indicates the likelihood for the potential parameters, green the qDF parameters. The true parameters are marked by dotted lines. The dark, medium and bright contours in the 2D distributions represent 1, 2 and 3 sigma confidence regions, respectively, and show weak or moderate covariances. This analysis was picked among five similar analyses, to have all 1 sigma contours encompass the input values. The likelihood here was sampled using MCMC (with flat priors in p_Φ and $\ln(p_{\rm DF})$ to turn the likelihood into a full *pdf*). Because only 10,000 MCMC samples were used to create the histograms shown, the 2D distribution has noisy contours. The dashed lines in the 1D distributions are Gaussian fits to the histogram of MCMC samples. This demonstrates very well that for such a large number of stars, the likelihood approaches the shape of a multi-variate Gaussian, as expected from the central limit theorem.

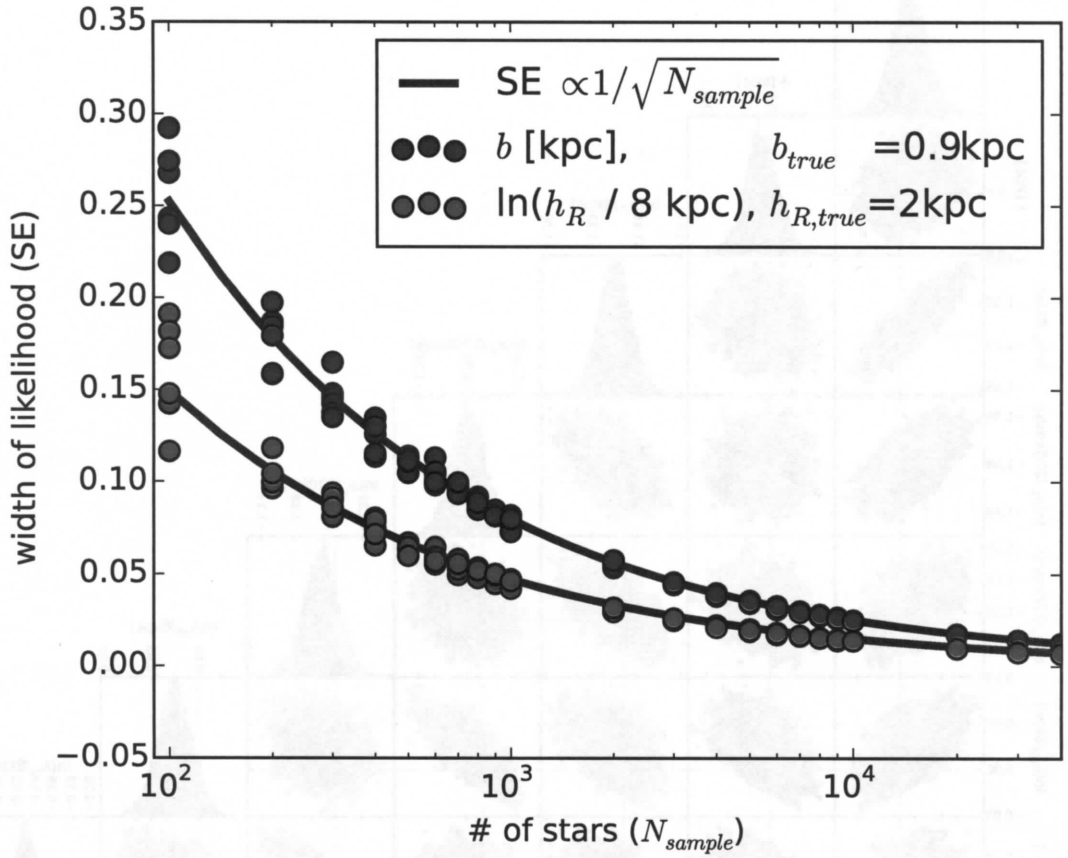

Fig. 5.— The width of the likelihood for two fit parameters found from analyses of 132 mock data sets vs. the number of stars in each data set. The mock data was created in the "Iso-Pot" potential and all model parameters are given as Test ② in Table ??. The likelihood in Eq. 11 was evaluated on a grid and a Gaussian was fitted to the marginalized likelihoods of each free fit parameter. The standard error (SE) of these best fit Gaussians is shown for the potential parameter b in kpc (red dots) and for the qDF parameter $\ln(h_R/8\text{kpc})$ in dimensionless units (blue). The black lines are fits of the functional form $\text{SE}(N_{\text{sample}}) \propto 1/\sqrt{N_{\text{sample}}}$ to the data points of both shown parameters. As can be seen, for large data samples the width of the likelihood behaves as expected and scales with $1/\sqrt{N_{\text{sample}}}$ as predicted by the central limit theorem.

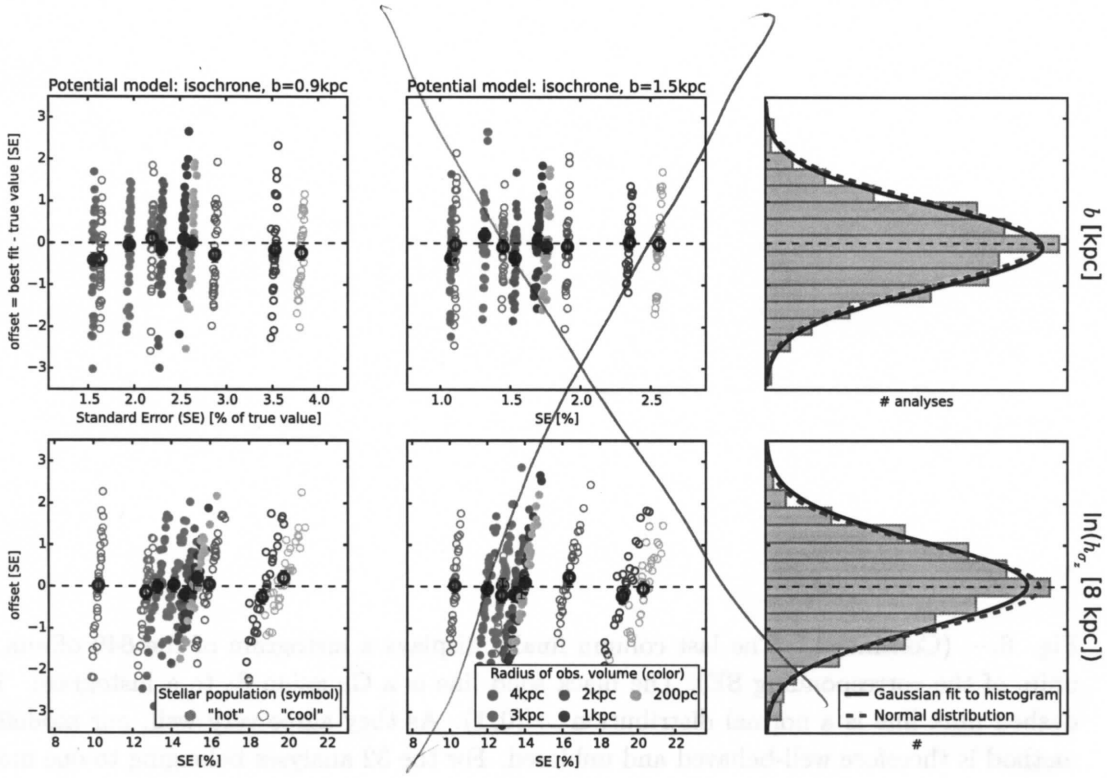

Fig. 6.— (Un-)bias of the parameter estimate. According to the central limit theorem the likelihood will follow a Gaussian distribution for a large number of stars. From this follows that also for a large number of data sets the corresponding best fit values for the model parameters have to follow a Gaussian distribution, centered on the true model parameters. That our method satisfies this and is therefore an unbiased estimator [TO DO: can I say that????] is demonstrated here. We create 640 mock data sets. They come from two different "Iso-Pot" potentials (first and second column), two different stellar populations ("hot" *MAP* (solid symbols) and "cool" *MAP* (open symbols)) and five spherical observation volumes of different sizes (color coded, see legend). All model parameters are summarized in Table 3 as test ③. We determine the best fit value and the standard error (SE) for each fit parameter by fitting a Gaussian to the marginalized likelihood. The offset is the difference between the best fit and the true value of each model parameter. In the first two columns the offset in units of the SE is plotted vs. the SE in % of the true model parameter. The first row shows the results for the isochrone scale length b and the second row the qDF parameter h_{σ_z}, which corresponds to the scale length of the vertical velocity distribution. [TO DO: rename isochrone potential in title to "Iso-Pot".]

3.2. The Role of the Survey Volume Geometry

Beyond the sample size, the survey volume *per se* must play a role; clearly, even a vast and perfect data set of stars within 100 pc of the Sun, has limited power to tell us about the potential at very different R. Intuitively, having dynamical tracers over a wide range in R suggests to allow tighter constraints on the radial dependence of the potential. To this end, we devise a number mock data sets, drawn from a one single p_M (see Test ??? in Table 3), but drawn from four different volume wedges (see §2.3), as illustrated in the right upper panel of fig. **??**. To make the parameter inference comparison very differential, the mock data sets are equally large (20,000) in all cases, and are drawn from identical total survey volumes (4.5 kpc^3, achieved by adjusting the angular width of the edges). The right panels of Fig.?? the illustrate the ability of *RoadMapping* to constrain model parameters (in this case two p_Φ parameters). The two top right panels of Fig.?? illustrate that the radial extent and the maximal height above the mid-plane matter. In the case shown, the standard error of the estimated parameters in twice as large for the volume with small ΔR and $\Delta|z|$; unsurprisingly, in the axisymmetic context the larger $\Delta\phi$ extent of that volume does not help to constrain the parameters. The panels in the bottom row explore whether the radial or vertical extent plays a dominant role: it appears that substantive radial and vertical extent are comparably important to constrain the parameters.

This Figure also implies that for these cases volume offsets in the radial or vertical direction have at most modest impact. While we believe the argument for significant radial and vertical extent is generic, we have not done a full exploration of all combinations of p_M and volumina. Figure 6 amplifies the same point: it illustrates that at given sample size, drawing the data – more sparsely – from a larger volume provides better p_M constraints.

[TO DO] Stuff to explain about fig. 6: Mention also that bigger volumes give most of the time better constraints and that there is no clear answer, if a hot or cooler population gives better constraints. Depends on parameter considered, selection function etc. [TO DO] 'Larger is better' is also demonstrated in fig. 6 We demonstrate that for a given size of the observation volume the shape and position of the volume does not matter much as long as we have both large radial and/or vertical coverage. In an axisymmetric potential the coverage in angular direction does not matter, as long as there are enough stars in the observation volume.

Make these titles & bigger

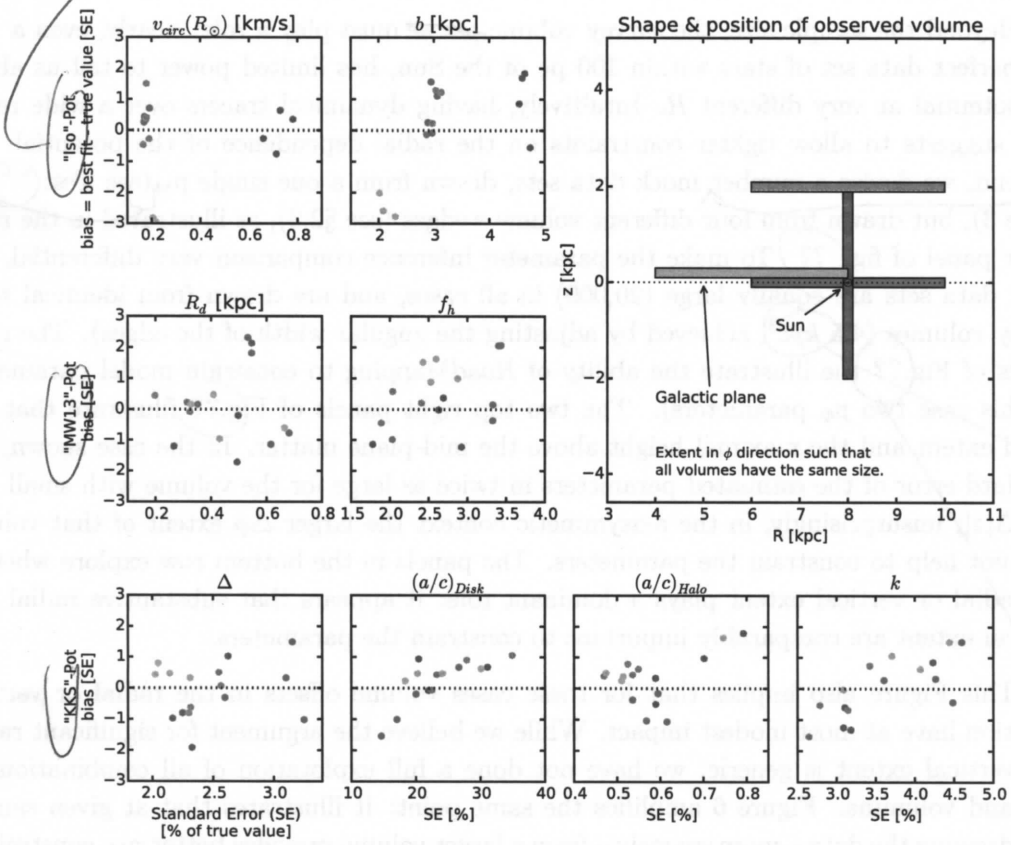

Fig. 7.— Bias vs. standard error in recovering the potential parameters for mock data stars drawn four different test observation volumes within the Galaxy (illustrated in the upper right panel) and three different potentials ("Iso-Pot", "MW13-Pot" and "KKS-Pot" from Table 1, top to bottom row). Standard error and offset were determined as in fig. 6. Per volume and potential we analyse four different mock data realisations; all model parameters are shown as Test ④ in Table 3. The colour-coding represents the different wedge-shaped observation volumes. The angular extent of each wedge-shaped observation volume was adapted such that all have the volume of 4.5 kpc^3, even though their extent in (R, z) is different. Overall there is no clear trend, that an observation volume around the sun, above the disk or at smaller Galactocentric radii should give remarkably better constraints on the potential the other volumes. [TO DO: MW-Pot and KKS-Pot analyses suffer from too low accuracy in action calculation (with StaeckelGrid). Used the StaeckelGrid for BOTH mock data and analysis, but the mock data distribution would actually not look exactly like the desired qDF distribution, i.e. this plot basically is created with a messed up DF. Don't know how higher accuracy would change the plot. The orange Iso-Pot analysis suffers from too small integration range in vT. More coding required, before redoing this.]

3.3. What if our assumptions on the (in-)completeness of the data set are incorrect?

The selection function of a survey is described by a spatial survey volume and a completeness function, which determines the fraction of stars observed at a given location within the Galaxy with a given brightness, metallicity etc (see §[TO DO CHECK]). The completeness function depends on the characteristics and mode of the survey, can be very complex and is therefore sometimes not perfectly known. We investigate how much an imperfect knowledge of the selection function can affect the recovery of the potential. We model this by creating mock data with varying incompleteness, while assuming constant completeness in the analysis. The mock data comes from a sphere around the sun and an incompleteness function that drops linearly with distance r from the sun (see ⑤, Example 1, in Table ?? and Fig. ??).

This could be understood as a model for the important effect of stars being less likely to be observed the further away they are. We demonstrate that the potential recovery with *RoadMapping* is very robust against somewhat wrong assumptions about the (in-)completeness of the data (see fig. ??). A lot of information about the potential comes from the rotation curve measurements in the plane, which is not affected by applying an incompleteness function. In Appendix §A.1 we also show that the robustness is somewhat less striking but still given for small misjudgements of the incompleteness in vertical direction, parallel to the disk plane (fig. ?? and ??). This could model the effect of wrong corrections for dust obscurement in the plane. We also investigate in Appendix §A.1 if indeed most of the information is stored in the rotation curve. For this we use the same mock data sets as in fig. ?? and ??, but this time were not including the tangential velocities in the modelling, rather marginalizing the likelihood over v_T. In this case the potential is much less tightly constrained, even for 20,000 stars. For only small deviations of true and assumed completeness ($\lesssim 10\%$) we can however still incorporate the true potential in our fitting result (see Fig. 18).

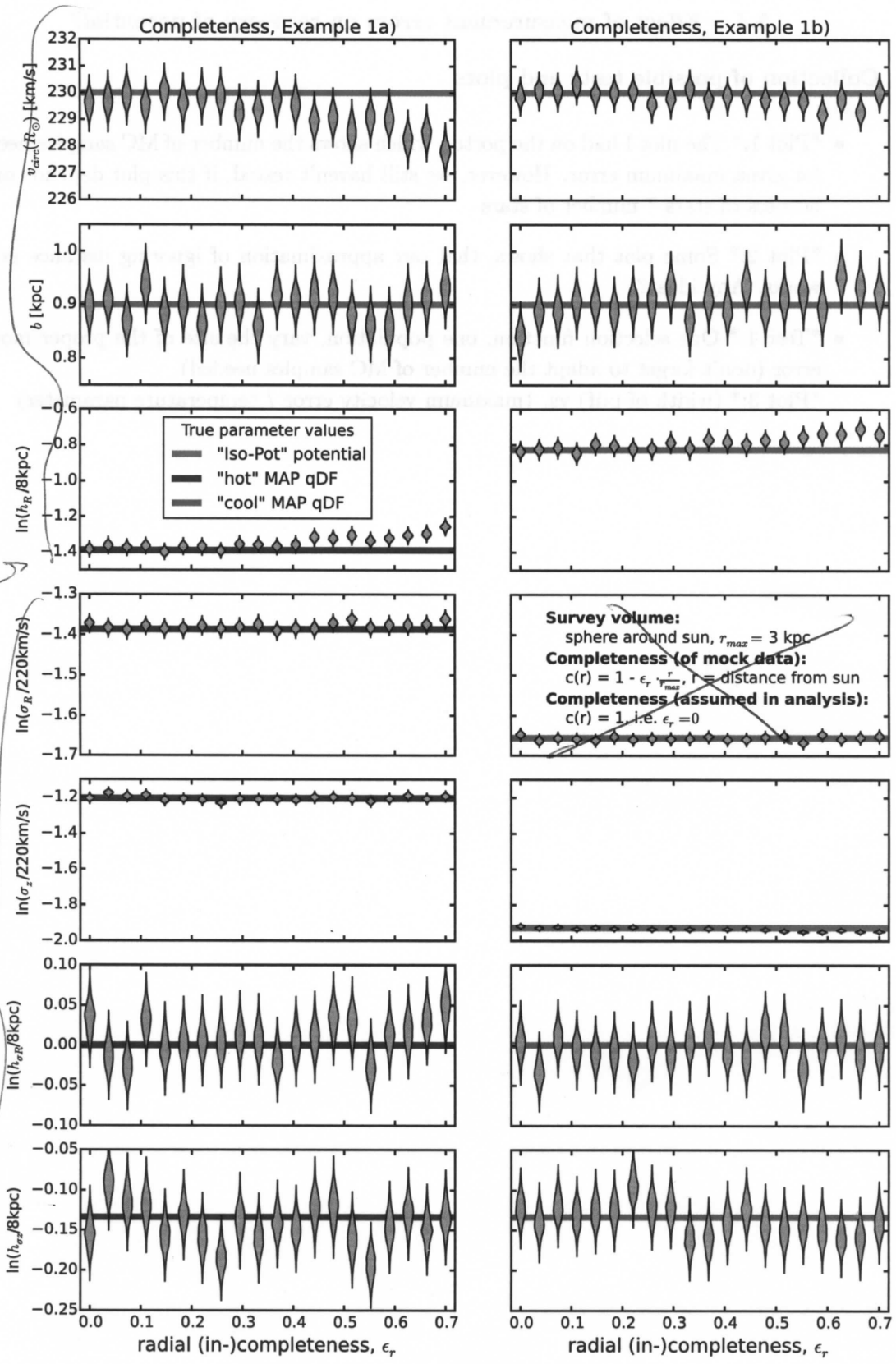

Fig. 9.— Influence of wrong assumptions about the radial incompleteness of the data on the