The MadMapper: Milky Way Action Distribution of Mono-Abundance Populations and Potential Exploration and Recovery

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

Max-Planck-Institute for Astronomy, Heidelberg

trick@mpia.de

and

J. Bovy[3,4]

Institute for Advanced Study, Princeton, NJ

ABSTRACT

Starting point for abstract: my old poster abstract. [TO DO] We aim to recover the Milky Way's gravitational potential using action-based dynamical modeling (cf. Bovy & Rix 2013, Binney & McMillan 2011, Binney 2012). This technique works by modeling the observed positions and velocities of disk stars with an equilibrium, three-integral quasi-isothermal distribution function. In preparation for the application to stellar phase-space data from Gaia, we create and analyze a large suite of mock data sets and we develop qualitative "rules of thumb" for which characteristics and limitations of data, model and code affect constraints on the potential most. We investigate sample size and measurement errors of the data set, size and shape of the observed volume, numerical accuracy of the code and action calculation, and deviations of the data from the assumed family of axisymmetric model potentials and distribution functions. This will answer the question: What kind of data gives the best and most reliable constraints on the Galaxy's potential?

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

---

[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

Contents

1. Introduction

[TO DO]

Collection of thoughts for the introduction: *(Text is not yet perfect or concise, but should serve as a starting point to setup a basic structure for the introduction. The text will then have to be shortened, redundant formulations have to be removed, phrasing has to be improved and everything has to be supported with appropriate references.)*

- **Our modelling method in a nutshell:** We fit simultaneously a model for the Galaxy's gravitational potential and an orbit distribution function (df) to stellar phase-space data. To turn a star's position and velocity into a full orbit, we need the gravitational potential in which the star moves. We assume that we know a family of orbit distribution functions that are close enough to the real distribution of orbits. In this case the stellar orbits calculated within a proposed potential will only follow such a df, if this potential model is close enough to the true potential.
  Or in other words: We need the potential to calculate orbits. At the same time, if we *know* the true orbits, we can deduce the true potential from them. To find the true orbits, we make use of the predictive power of an orbit distribution function.

- **Motivation to use this modelling technique in the Milky Way:** Bovy et al. 2012 .... [TO DO]

- **Introducing orbits and actions:** There are different ways to describe stellar orbits. The most obvious is to give the stars position and velocity vector at each point in time,

by evaluating the potential forces that act on the star in each time step. Most orbits in realistic galaxy potentials are however not closed, so we would have to integrate the orbit forever. Another, much more convenient way to describe orbits, are so called integrals of motion. These integrals are functions of the star's time-dependent position and velocity, but are themselves constants in time, i.e. conserved quantities. The most obvious integral in static potentials is the energy of the orbit. Symmetries in potentials frequently allow more than one integral: In spherical potentials all three components of the angular momentum are conserved. In many axisymmetric potentials there is, in addition to the energy E and vertical component of the angular momentum L_z, a third non-classical integral of motion I_3, which has however no easy physical meaning. (Binney & Tremaine, Galactic Dynamics)
Because any function of integrals is an integral of motion itself, it is possible to construct integrals that have both very convenient properties and intuitive physical meanings. One such a set are the so-called actions. In axisymmetric potentials they are frequently called the radial action J_R, the vertical action J_z and the ϕ-action, which is simply the vertical component of the angular momentum, L_z. The radial action and vertical action quantify the amount of oscillation in radial and vertical direction that the orbit exhibits. Actions are constructed in such a way, that they are not only integrals, but also correspond to the momenta in a set of canonical coordinates. The canonical conjugate positions of the actions are the so-called angles, which have the convenient properties, that they increase strictly linearly in time while the star moves along the orbit. They are periodic in 2π and the frequencies by which they change are functions of the actions. In the action-angle coordinate system, the only thing we need to fully describe an orbit in an axisymmetric potential are therefore just three fixed numbers, the actions.

- **Using actions for distribution functions:** Actions are therefore the natural coordinates of orbits and each point in action space corresponds to one specific orbit in a given potential. It is often used in dynamical modelling, e.g. in the Schwarzschild superposition method (source???), to reconstruct a galaxy by superimposing different orbits and populating them with stars. In this way these kind of methods construct orbit distribution functions for galaxies, which are at the same time distribution functions in action space. Because angles increase linearly in time, when a star moves along its orbit, stars are uniformly distributed in angle space. Therefore a orbit distribution function in terms of actions and a uniform distribution of stars in angle-space can be directly mapped to a distribution of stars in canonical configuration phase-space, measurable stellar positions and velocities. While a stellar distribution in configuration space is six-dimensional, the distribution in action-angle space is effectively three-dimensional,

because of the uniformity in angles. (Rewrite, too verbose...)

- **Why should we care about actions in realistic galaxies?** In reality galaxies have rarely perfectly static and axisymmetric potentials, which drastically reduces the number of conserved quantites along orbits. In static non-axisymmetric potentials there can still be two integrals of motion, angular momentum however is no longer conserved. The Milky Way's disk might have an overall axisymmetric appearance, but is perturbed by spiral arms. The strongest deviation from axisymmetry in the Galaxy is the bar, which also causes the Galactic potential to vary slowly in time. The stirrs up the stars of the disk and the potential and causes radial migration of the orbits (Reference???), orbits change and with them the actions. One could wonder if, under such non-axisymmetric, non-static potential conditions, the assumption and treatment of globally conserved actions in the Milky Way is still a sensible approach. First of all, actions are the natural way to treat orbits and they can be locally defined, even if they might not be globally conserved. As long as we care about orbits, we should care about actions. An orbit carries information about the star's past, about where the star was born and which tidal processes might have carried it away from its inital orbit. Together with the chemistry of the stars, which determined by their place of birth, their current orbits are valuable diagnostics for the evolution and structure of the Milky Way. Secondly, gravitational processes do only in the most extreme cases completely change the actions. In a slowly changing potential, where orbits adapt adiabatically to those changes, actions are conserved (Binney & Tremaine, Galactic Dynamics). And even during bar-induced radial migration at least the vertical actions are conserved and will continue to carry some amount of information about the stars' inital orbit distribution.

  [TO DO] (Maybe cite Potzen 2015, who showed that analysing aspherical systems in spherical actions can still be a powerful tool, when used with care...)

- **Why should we care about an axisymmetric "best fit" model for the Milky Way disk?** One of the key assumptions of our modelling technique is the assumed axisymmetry of the Milky Way's gravitational potential, especially its disk. As we discussed already in the previous paragraph, this assumption is indeed only an approximation to the real disk, which has a much richer structure and more complicated potential, with spiral arms and ring-like structures (like the Monocerros ring), with a warp and a flare in the outer disk (references????). Also the Milky Way's halo has substructure, a multitude of streams (references???) and shell-like overdensities (reference???). The ultimate goal will be to find and identify substructures observationally and describe theoretically the structure and evolution of potential perturbations. Our method and efforts to extract information about the axisymmetric Milky Way potential

from disk stars aims to create a reliable and well-constraint basis for these endevours: The best possible axisymmetric approximation to the Milky Way's potential could serve as a realistic equilibrium model from which a description of non-axisymmetric tidal perturbations can be theoretically established by perturbation theory. It will also help a great deal to identify sub-structures, e.g. to find and orbitally connect tidal streams, which in return will then give better constraints on the deviations from axisymmetry. Many modelling and techniques, both purely gravitational, but also chemo-dynamical, can greatly profit from a good axisymmetric model for the galaxy: While we are still far away from knowing the MW's potential all over the place, an axisymmetric model will be the best reference to turn phase-space coordiantes into whole orbits. And orbits are the diagnostics that carry information from everywhere in the galaxy into the solar neighbourhood, where we can hope to exploit them. (Some overlap with section before. How to better structure these two sections and assign the arguments more clearly to "axisymmetric disk" or "actions"?)

- **Previous results with this modelling technique:** Bovy & Rix (2013) ... [TO DO]
  - disk scale length $R_d = 2.15 \pm 0.14$ kpc (Bovy & Rix 2013)
  - disk is maximal (Bovy & Rix 2013)
  - slope of dark matter halo $\alpha < 1.53$ (Bovy & Rix 2013)
- **What do we already know about the axisymmetric MW disk (from other references)?** [TO DO]
  - rotation curve is well-known (reference???)
- **What is there left to learn about the axisymmetric MW disk?** (as Jo asked at the Santa Barbara conference... [TO DO]
  - separation of different MW component is still unclear: individual density profiles, contributions to total pot
  - thin/thick disk vs. continuum of exponential disks
  - dark matter at smaller radii
  - slope & shape of dark matter halo (current state of knowledge?)
- **Other modelling approaches using DF's similar to Binney:**
  - Piffl et al. (2014) used a slightly different DF-based modelling approach to constrain the MW's vertical density profile near the sun. They fitted a superposition of "quasi-isothermal" DFs for thick and thin disk, and a DF for the halo to

˜200,000 giant stars from the RAVE survey (RAdial Velocity Experiment, Steinmetz et al. (2006)). They didn't use any chemical information of the stars. To account for different populations within the thin disk, they weighted the corresponding DF's with an assumed star-formation rate instead. To circumvent the use of RAVE's non-trivial spatial selection function, they separated stars into spatial bins in (R, z) and fitted the velocity distribution predicted by their DF and potential model at the mean (R, z) of each bin to the observed velocities only. They're result for their radial profile of the vertical force within $|z| = 1.1$ kpc and $R > 6.6$ kpc agrees well with the previous results from our method by Bovy & Rix (2013). By not using chemical information and hiding the spatial distribution of stars by binning to circumvent a complicated selection function, Piffl et al. (2014) is however rejecting a lot of valuable information in the data set. ([TO DO: Look at other useful references in this paper: Bienayme et al. 2014, Zhang et al. 2013, Binney et al. 2014a, Binney 2012b, McMillan & Binney 2013])

- **Motivating this method characterization in anticipation of GAIA:** [TO DO]

2. Model & Mock Data

2.1. Actions

[TO DO]

2.2. Distribution function

[TO DO]

2.3. Potential models

[TO DO] Mention different ways to calculate actions in different potentials.

2.4. Selection function: observed volume and completeness

[TO DO]

2.5. 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 obvious procedure would be to draw $\boldsymbol{J}_i$ from $\mathrm{qDF}(\boldsymbol{J}_i \mid p_{DF})$ and $\boldsymbol{\theta}_i$ between 0 and 2π, transforming this to $(\mathrm{x}_i, \mathrm{v}_i)$ in the given potential and rejecting all stars that are outside the observed volume. 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 $(\boldsymbol{x}_i, \boldsymbol{v}_i) \longrightarrow (\boldsymbol{J}_i, \boldsymbol{\theta}_i)$. The observed volume is also much smaller than the whole galaxy and the fraction of rejected stars would be enormous. We propose a fast and simple two-step method for drawing mock data from an action distribution function in a given observed volume. [TO DO]

2.5.1. Preparation: Tracer density

One crucial point in creating the mock data, as well as in the analysis, is to calculate the spatial tracer density $\rho_{\mathrm{DF}}(\boldsymbol{x} \mid p_\Phi, p_{DF})$ for a given distribution function with parameters p_{DF} in a potential with parameters p_Φ. We do this by integrating the axisymmetric distribution function over the velocity at $N_{\mathrm{spatial}} \times N_{\mathrm{spatial}}$ regular grid points in the (R, z) plane, using a Gauss-Legendre quadrature of order N_{velocity} in each of the three velocity components. The velocity distribution according to the qDF at a given (R_i, z_i) looks approximately Gaussian for v_R and v_z. The v_T distribution peaks somewhere around $v_{circ}(R_\odot)$ and $v_T > 0$ (cf. §?????). We approximate the integration over the velocity therefore as

$$\begin{aligned} \rho_{\mathrm{DF}}(R, |z| \mid p_\Phi, p_{DF}) &= \int_{-\infty}^{\infty} \mathrm{qDF}(J[R, z, \boldsymbol{v} \mid p_\Phi] \mid p_{DF})\, \mathrm{d}^3\boldsymbol{v} && (1) \\ &\approx \int_{-N_{\mathrm{sigma}}\sigma_R(R|p_{DF})}^{N_{\mathrm{sigma}}\sigma_R(R|p_{DF})} \int_{-N_{\mathrm{sigma}}\sigma_z(R|p_{DF})}^{N_{\mathrm{sigma}}\sigma_z(R|p_{DF})} \int_0^{1.5 v_{circ}(R_\odot)} \\ &\qquad \mathrm{qDF}(J[R, z, \boldsymbol{v} \mid p_\Phi] \mid p_{DF})\, \mathrm{d}v_T\, \mathrm{d}v_z\, \mathrm{d}v_R, && (2) \end{aligned}$$

where $\sigma_R(R \mid p_{DF})$ and $\sigma_z(R \mid p_{DF})$ are the star's radial and vertical velocity dispersion according to the qDF and given by eq. (???) and (???) and N_{sigma} has to be large enough. The size of the $N_{\mathrm{spatial}} \times N_{\mathrm{spatial}}$ grid in (R, z) is chosen cover the extent of the observed volume for $z > 0$. We interpolate then over this grid to be able to evaluate the density

at each (R, z) within the observed volume. The total number of grid points and therefore actions that need to be calculated are $N^2_{\rm spatial} \cdot N^3_{\rm velocity}$. Fig. ??? shows the importance of choosing $N_{\rm spatial}$, $N_{\rm velocity}$ and $N_{\rm spatial}$ sufficiently large in order to get the density with an acceptable numerical accuracy. For the creation of the mock data we use $N_{\rm spatial} = 20$, $N_{\rm velocity} = 40$ and $N_{\rm sigma} = 5$.

2.5.2. Step 1: Drawing positions from the selection function

To get $\boldsymbol{x}_i$ for our mock data stars, we first sample random positions (R_i, z_i, ϕ_i) uniformly from the observed volume. Making use of the geometric shapes of our simple observed volumes (spheres, cylinders, wedges...) we can do this efficiently with inverse transform Monte Carlo sampling. In a second step we then apply a rejection Monte Carlo method to these positions using the pre-calculated, interpolated density grid $\rho_{\rm DF}(R, |z| \mid p_\Phi, p_{DF})$. 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 on the set of positions we got from the last step. This procedure can be repeated until the required number of positions are reached. The sample then follows the required distribution

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

2.5.3. 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) found in step 1 we now sample velocities directly from the $\mathrm{qDF}(R_i, z_i, \boldsymbol{v} \mid p_{Phi}, p_{DF})$ 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. The envelope Gaussian peaks at $(v_R = 0, v_T = max(\mathrm{qDF}(R_i, z_i, 0, vT, 0 \mid p_{Phi}, p_{DF})), v_z = 0)$ and has a standard deviation of $(2\sigma_R(R_i), 2\sigma_R(R_i), 2\sigma_z(R_i))$. We then pick one of the sampled velocities $\boldsymbol{v}_i$ at and for each (R_i, z_i) and arrive at our desired mock data sample

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

2.5.4. Step 3: Introducing measurement errors

[TO DO]

Collection of possible tests and plots

- *Diagram 1*: schematic flow chart of how to sample mock data (could be helpful for people, who want to sample mock data in action space and didn't know how to start, like me)
- *Plot 2:* 2 triangle plots with (jr, lz, jz) on the axes to show the distribution of stars in action space within mock data sets - for a large sphere and a small sphere. (I thought it was very instructive to see how the spatial selection function shapes the distribution of actions, it also helped me understand which orbits have which actions.)
- *Plot 3:* distribution of mock data set in real space: z vs. R. and vz vs. R, maybe for a hot and cold population? (maybe a bit boring? Would be however illustrative, that the mock data sampled from the qdf is indeed similar to something we could observe. Also: could make a 4 sigma contour in the vz vs. R plot, to show, that the choice of integration limits is important but 4 sigma should be sufficient.)

3. Analysis

The idea behind our modeling approach is that the orbits of the stars belonging to one MAP [TO DO: explain MAP???], calculated from a phase-space observation for each star within a proposal potential, will only follow a distribution function from the family of qDFs (cf. §2.2) if this propsal potential is (close to) the true potential in which the stars move. This opens up the possibility to fit the qDF and the potential simultaneously to the stellar phase-space data of one MAP, using the orbits of the stars.

3.1. Likelihood

3.1.1. Data and Selection Function

We're fitting the potential and the qDF to the data

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

where $\boldsymbol{x}_i$ and $\boldsymbol{v}_i$ are the position and velocity of one star. The phase-space volume within which stars are observed by a given survey is defined by the survey's selection function $\mathrm{sf}(\boldsymbol{x}, \boldsymbol{v})$, which is in general a function of the position only, $\mathrm{sf}(\boldsymbol{x})$. To first order the shape of the selection function ("observed volume") is limited by the directions in which the survey is pointed and the sensitivity down to which limiting magnitude it can detect stars. In the simplest case, if all stars had the same brightness, the selection function is 1 everywhere inside the observed volume and 0 outside. Because stars have different brightness the selection function will usually decrease from 1 close to the sun to 0 at the edges of the observed volume ("completeness"). [TO DO: Explain selection function somewhere else????] Only stars for which the selection function is non-zero are contained in the data set D_j.
Our modeling takes place in the Galactocentric rest-frame with cylindrical coordinates $\boldsymbol{x} = (R, \phi, z)$ and velocity components in the corresponding coordinate directions $\boldsymbol{v} = (v_R, v_\phi, v_z)$.[1]

3.1.2. Model Parameters

We fit the five free parameters of the qDF family, h_R, σ_R, σ_z, h_{σ_R} and h_{σ_z}, in logarithmic scale, which corresponds to a logarithmically flat prior in the framework of Bayesian statistics. The set of qDF fit parameters is therefore

$$p_{DF} := \{\ln(h_R/8\mathrm{kpc}), \ln(\sigma_R/220\mathrm{km/s}), \ln(\sigma_z/220\mathrm{km/s}), \ln(h_{\sigma_R}/8\mathrm{kpc}), \ln(h_{\sigma_z}/8\mathrm{kpc})\}.$$

To be able to control the number of degrees of freedom in the potential fit, we have to assume a certain family of potential models, parametrized by the parameters p_Φ (cf. §2.3).
The total set of model parameters to fit is then

$$M = \{p_{DF}, p_\Phi\},$$

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 $\mathrm{qDF}(\boldsymbol{J}_i \mid M) := \mathrm{qDF}(\boldsymbol{J}[\boldsymbol{x}_i, \boldsymbol{v}_i \mid p_\Phi] \mid p_{DF})$.

---

[1] If the phase-space data is given in observed coordinates, position $\tilde{\boldsymbol{x}} = (\alpha, \delta, m - M)$ in right ascension α, declination δ and distance modulus $m - M$ and velocity $\tilde{\boldsymbol{v}} = (\mu_\alpha, \mu_\delta, v_{\mathrm{los}})$ as proper motions $\boldsymbol{\mu} = (\mu_\alpha, \mu_\delta)$ [TO DO: cos somwhere???] and line-of-sight velocity v_{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 (cf. §???).

3.1.3. Form of the Likelihood

The likelihood of the model given the data $\mathscr{L}(M \mid D_j)$ 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.

$$\mathscr{L}(M \mid D_j) = \prod_i^{N_j} P(\boldsymbol{x}_i, \boldsymbol{v}_i \mid M), \tag{3}$$

where N_j is the number of stars in the data set D_j. This probability is, properly normalized,

$$\begin{aligned} P(\boldsymbol{x}_i, \boldsymbol{v}_i \mid M) &= \frac{\mathrm{qDF}(\boldsymbol{J}_i \mid M) \cdot \mathrm{sf}(\boldsymbol{x}_i)}{\int \mathrm{d}^3x\, \mathrm{d}^3v\ \mathrm{qDF}(\boldsymbol{J} \mid M) \cdot \mathrm{sf}(\boldsymbol{x})} \\ &\propto \frac{\mathrm{qDF}(\boldsymbol{J}_i \mid M)}{\int \mathrm{d}^3x\ \rho_{\mathrm{DF}}(R, |z| \mid M) \cdot \mathrm{sf}(\boldsymbol{x})}. \end{aligned} \tag{4}$$

In the second step we used eq. (1). The factor $\prod_i \mathrm{sf}(\boldsymbol{x}_i)$ is independent of the model parameters, so we use simply eq. (4) in the likelihood calculation. We find the best set of model parameters by maximising the likelihood.

3.1.4. Numerical accuracy in calculating the likelihood

[TO DO: Consistent capitals in section titles. ????]

To evaluate the likelihood at a given set of (p_Φ, p_{DF}) we proceed in principle in the following way: The numerator in eq. (4) can be calculated straightforward by calculating the actions of each star in the given potential (cf. §???) and then evaluating the qDF at each action. For the normalisation of the likelihood we first have to calculate the density $\rho_{\mathrm{DF}}(R, |z| \mid M)$ on a grid as described in §2.5.1. The density is then interpolated using bivariate spline interpolation. In the case of $\mathrm{sf}(\boldsymbol{x}) = 1$ everywhere inside the observed volume and $\mathrm{sf}(\boldsymbol{x}) = 0$ outside, i.e. for a complete sample, the integral in the normalisation in eq. (4) is essentially two-dimensional in R and z and we can use the shape of the observed volume to set finite integration limits. We perform this integral over the interpolated tracer density by using Gauss Legendre integration of order 40 in each R and z direction. The integration over ϕ is done analytically.

Unfortunately the evaluation of the likelihood for only one set of model parameters is already very computationall expensive. The computation speed is set by the number of action calculations needed, i.e. the number of stars and the numerical accuracy of the integrals in the normalisation, which requires $N^2_{\mathrm{spatial}} \times N^3_{\mathrm{velocity}}$ action calculations. The numerical

accuracy has to be choosen high enough, such that the integrals in the normalisation are mostly converged and the error introduced by this does not dominate in the likelihood, i.e.

$$\begin{aligned} \log \mathscr{L}(M \mid D_j) &= \sum_i^{N_j} \log \mathrm{qDF}(\boldsymbol{J_i} \mid M) \\ &\quad -N_j \log(\text{true normalisation}) - N_j \log(1 + \text{rel. error}), \\ \text{with} &\quad N_j \log(1 + \text{rel. error}) \lesssim 1. \end{aligned}$$

[TO DO: Don't understand why 1 is the threshold here. ???] For data sets as large as $N_j =$ 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. 4 demonstrates that the numerical accuracy we use in the analysis, $N_{\mathrm{spatial}} = 16$, $N_{\mathrm{velocity}} = 24$ and $N_{sigma} = 5$, does satisfy this requirement.[2] [TO DO: Should we also show that 40th order GL integraiton over interpolated density is enough? as this is really a lot and well converged, I would simply state that this is enough, but not show anythin.????] [TO DO: Look up what McMillan & Binney 2013 have to say about the numerical accuracy of the normalisation. Sanders & Binney (2015) are quoting them on that matter.]

3.2. Measurement Errors

[TO DO]

3.3. Fitting Procedure

We search the (p_Φ, p_{DF}) parameter space for the maximum of the likelihood in eq. (3). The most crucial part of our fitting procedure for finding the peak and width of the likelihood in the (p_Φ, p_{DF}) parameter space is therefore the reduction of computational costs while not introducing systematic errors due to numerical inaccuracies. We do this by 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 will either sample the shape of the likelihood (or rather the posterior probability distribution) using a Monte-Carlo Markov Chain (MCMC) or calculate the likelihood on a much finer grid.

[2] In case of the isochrone potential we already have high enough accuracy for $N_{\mathrm{spatial}} = 16$, $N_{\mathrm{velocity}} = 20$ and $N_{sigma} = 4$.

3.3.1. Fitting Step 1: Nested-grid search with a fiducial qDF

[TO DO: Make consistent: use of $\sigma_{R,0}$ and σ_R as profile or dispersion at sun. ???]

The (p_Φ, p_{DF}) parameter space can be high-dimensional and we do not necessarily have a good notion where to look for the likelihood peak initially. We use a nested-grid approach to minimize effectively the number of models for which we have to evaluate the likelihood. Using a grid is preferable to other optimizing methods in our case, because we have fast ($p_{\rm DF}$) and very slow (p_Φ) parameters. Each time the likelihood is evaluated for a new potential, we have to convert the observed phase-space data of the stars $(\boldsymbol{x}_i, \boldsymbol{v}_i)$ into the actions $\boldsymbol{J}_i$ in this given potential. As we need to calculate $N_j \times N_{\rm error} + N^2_{\rm spatial} \times N^3_{\rm velocity}$ [TO DO: explain $N_{\rm error}$???] actions, this consumes most of the computation time. A grid-search allows us to pre-calculate all actions for the k-th potential $p_{\Phi,k}$ first and then step through all $p_{\rm DF}$ parameters and evaluate $\mathscr{L}(p_{\Phi,k}, p_{\rm DF})$. The calculations in each of these two steps, first action calculation and then likelihood calculation for all $p_{\rm DF}$s, do not depend on each other and can therefore be easily parallelized on multiple cores. For this to work properly, the pre-calculation of the actions has to be completely independent of the $p_{\rm DF}$ parameters. The actions needed for the normalisation are related to the $p_{\rm DF}$ through the velocity integration limits in eq. (???). We therefore 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". [TO DO: Continue]
The function evaluations of the qDF at each of the $N_j \times N_{\rm error} + N^2_{\rm spatial} \times N^3_{\rm velocity}$ actions have a smaller, but non-negligible contribution to the total computation time. Therefore the number of qDFs we are investigating for each potential also plays a role. [TO DO]

3.3.2. Choosing integration ranges

[TO DO]

4. Results

4.1. Verification of the Method

Collection of possible tests and plots

- *Plot 1:* two panels:

  - a) convergence of the normalisation vs. ngl_vel (GL order of integrating the qdf over the velocities to get the density),
  - b) convergence of the normalisation vs. n_dens (number of grid points in each (R,z) at which the density is explicitely calculated, before interpolating and integrating over the volume to get the normalisation).
    This might not be a very exciting plot, but when we later show plots, that demonstrate e.g. how robust the method is against incompleteness, people might think, that in this case the normalization is not so important and time could be saved in calculating it. We know, that it is important to get the normalisation right. Plus, it proves, that possible biases when using the Stäckel approximation are not due to a wrong normalisation.

- *Test 1:* Isochrone potential, 2 different b, 2 different populations, 5 different SF (isoSph test suite)
  Plot 2: scatter plot (offset / stddev) vs. (stddev / true value [%]) for b and one qdf parameter –¿ and panel with normal distribution. This plot could show 3 things:
  - Central limit theorem is satisified –¿ method works.
  - Bigger volumes give better constraints.
  - hot populations seem to give tighter constraints on the potential.

- *Plot 3:* Would be cool to have a plot, that shows that for the Stäckel potential we don't get biases, but that there are some for the analytic Miyamoto-Nagai + power-law halo & interpolated MW potential and therefore this bias is probably due to incorrect action calculation.

- *Plot 4:* stddev 1/sqrt(N)

4.2. Do shape and position of the observation volume matter?

Collection of possible tests and plots *Test 1:* Compare results of wedges of same volume, but different positions and orientations.

- I guess, the ones that demonstrate, that phi-coverage is much less important than R and z coverage, is boring, right? And I already have a plot in 3.1 that shows, that larger volumes are better.

- This test suite was made with the MW-like potential and there seem to be biases, that are different for different volumes. If we say, okay, we have to deal with whatever biases we get, I could still include those volumes with good R AND z coverage, because for them the biases seem to be smaller.

- I might add a few more volumes, e.g. one with large vertical coverage at different positions

- Do we explicitely want to test, if it matters, if the radial coverage is larger or smaller the disk scale length, and the vertical coverage is larger or smaller than the disk scale height?
  Plot 1: a) cross section of volumes in R and z b) offset / stddev vs. stddev / true value [%], that demonstrates, that it doesn't matter much for the potential recovery, if we have more radial or vertical coverage, and the position within the galaxy.

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

Collection of possible tests and plots

- *Test 1:* isochrone potential, b=0.9 kpc, two populations, completeness(d) $= 1 - \epsilon \cdot d/Rmax$, where Rmax is radius of spherical selection function. Marginalize over vT in analysis.
  Plot 1: Violin plot: x-axis - ϵ. y-axis: b-parameter and one of the qdf parameters.

- *Test 2:* isochrone potential, two populations, incompleteness function that depends only on z.
  Plot 2: violin plot

4.4. What if our assumed distribution function differs from the stars' DF?

Collection of possible tests and plots

- *Test 1:* mix hot and cold populations, 5 free qdf parameters in analysis!, use code that estimates the best velocity integration ranges. h_sigma_r & h_sigma_z are the same for both populations, sigma_r and sigma_z have the same ratio, but are 50% different for the two populations. h_R is also 50% different. Vary the fraction of pollution. Idea

behind this: What if the stellar distribution has a different shape, e.g. added "wings", or had a different tracer density decrease with R. Would be however great, if we could show how the mixture of qdf's quanlitatively changes the shape of the df. Any ideas? *Plot 1:* Violin plot: x-axis - fraction of pollution. y-axis: b-parameter and one or two qdf parameters.

- *Test 2:* same as Test 1, but this time vary the degree of difference and make it 50% pollution. Idea behind this: What happens, if we have errors in the abundances and mix different MAPs? For this it would be could to compare how much the qdf parameters of neighbouring MAPs differ and how big the difference between MAPs can be, such that it still can reproduce the potential.
  Plot 2: Violin plot: x-axis - difference in qdf parameters. y-axis: b-parameter and one or two qdf parameters.

4.5. What if our assumed potential model differs from the real potential?

Collection of possible tests and plots *Test 1:* Try to recover a Miyamoto-Nagai disk + power-law halo potential by fitting a 2-component Stäckel potential.
Plot 1:

- (R,z)-plane: color coding: difference between true potential's F_R and best fit potential F_R

- (R,z)-plane: color coding: difference between true potential's F_z and best fit potential F_z
  Any idea how to account for the error bars on the best fit potential?

4.6. Effect of measurement errors on recovery of potential?

Collection of possible tests and plots

- *Plot 1:* The plot I had on the poster, which shows the number of MC samples needed for given maximum error. However, we still haven't tested, if this plot depends on: * hotness of stars * number of stars

- *Plot 2:* Some plot that shows, that our approximation of ignoring distance errors works. Any ideas?

- *Test 1:* One selection function, one population, vary the size of the proper motion error (don't forget to adapt the number of MC samples needed)
  Plot 3: (width of pdf) vs. (maximum velocity error / temperature parameter)

5. Conclusion

[TO DO]

6. Questions that haven't been covered so far:

- What limits the overall code speed?
- What happens, when the errors are not uniform?
- What if errors in distance matter for selection?
- Deviations from axisymmetry: Take numerical simulations.

[TO DO: Check if all references are actually used in paper. ???]

REFERENCES

Binney, J. J., & McMillan, P. 2011, MNRAS, 413, 1889

Binney, J. J. 2012, MNRAS, 426, 1324

Bovy, J., Rix, H.-W., & Hogg, D. W. 2012b, ApJ, 751, 131

Bovy, J., Rix, H.-W., Hogg, D. W. et al., 2012c, ApJ, 755,115

Bovy, J., Rix, H.-W., Liu, C. et al., 2012d, ApJ, 753, 148

Bovy, J., & Rix, H.-W. 2003, ApJ, 779, 115

Piffl, T., Binney, J., & McMillan, P. J. et al., 2014, MNRAS, 455, 3133

Steinmetz, M. et al., 2006, AJ, 132, 1645

Ting, Y.-S., Rix, H.-W., Bovy, J., & van de Ven, G. 2013, MNRAS, 434, 652

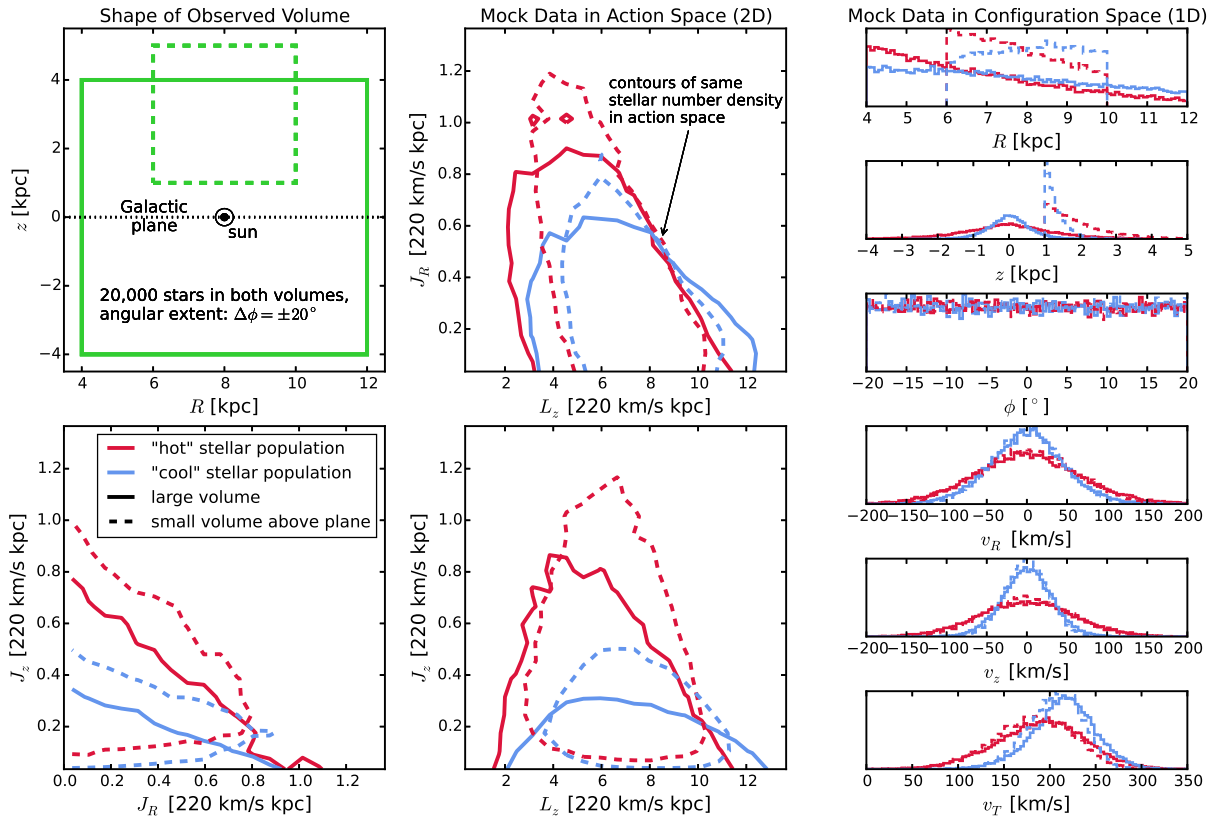
Shape of Observed Volume
Galactic plane
sun
20,000 stars in both volumes, angular extent: Δφ = ±20°
z [kpc]
R [kpc]
Mock Data in Action Space (2D)
contours of same stellar number density in action space
J_R [220 km/s kpc]
L_z [220 km/s kpc]
Mock Data in Configuration Space (1D)
R [kpc]
z [kpc]
φ [°]
"hot" stellar population
"cool" stellar population
large volume
small volume above plane
J_z [220 km/s kpc]
J_R [220 km/s kpc]
L_z [220 km/s kpc]
v_R [km/s]
v_z [km/s]
v_T [km/s]

Fig. 1.— [TO DO]

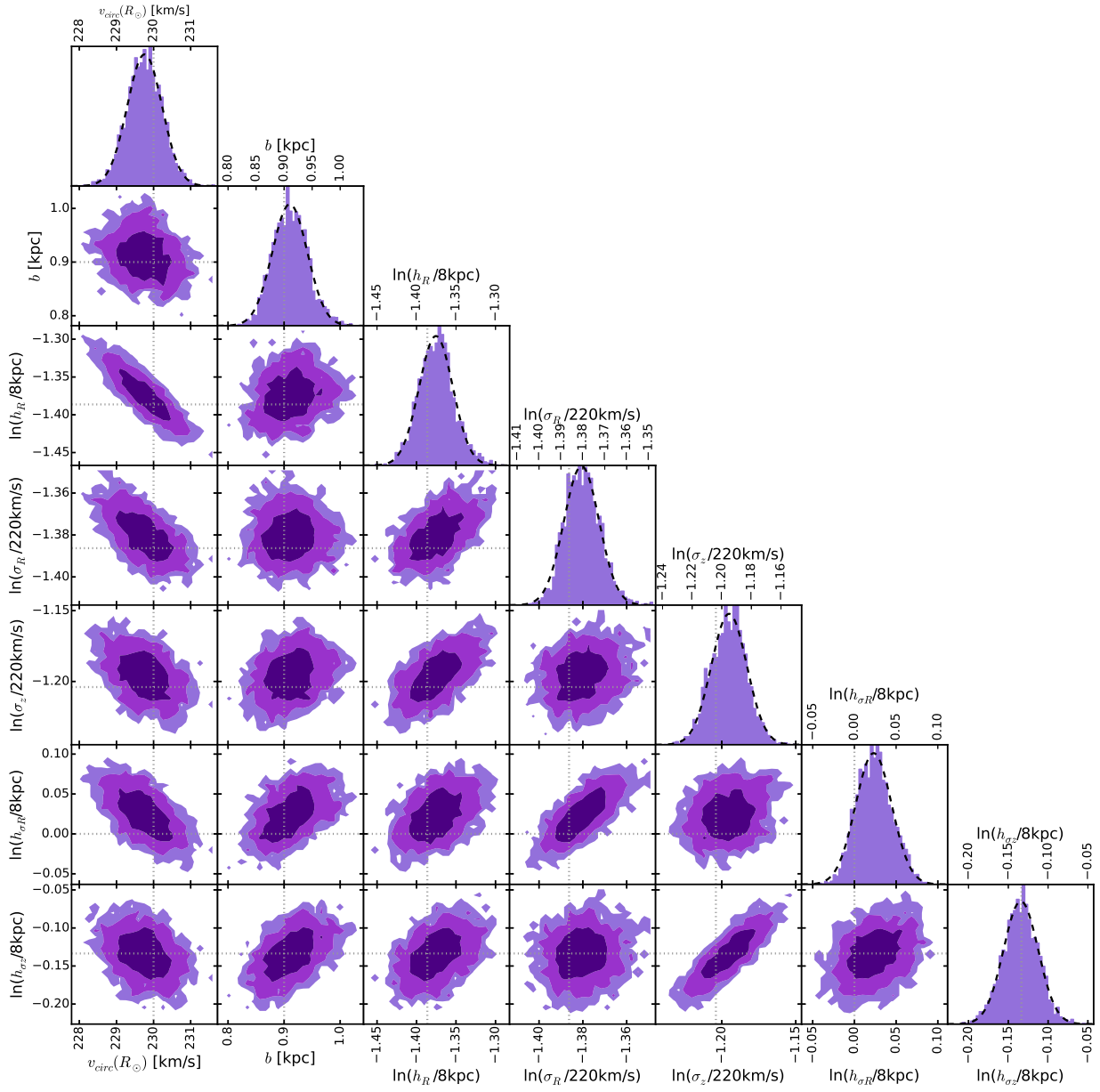

Fig. 2.— The likelihood in eq. (???) in the parameter space $\{p_\Phi, \ln(p_{DF})\}$ for one example mock data set. This mock data set has 20,000 stars and was created in a isochrone potential with $p_\Phi = \{v_{circ}, b\} = \{230$ km/s $, 0.9$ kpc $\}$, observed within a spherical volume around the sun of radius $r = 2$ kpc , and represents a rather hot stellar population with DF parameters $p_{DF} = \{h_R, \sigma_R, \sigma_z, h_{\sigma_R}, h_{\sigma_z}\} = \{2$ kpc $, 55$ km/s $, 66$ km/s $, 8$ kpc $, 7$ kpc $\}$. The true parameters are marked by dotted lines. The dark, medium and bright purple contours in the 2D distributions represent 1, 2 and 3 sigma confidence regions, respectively, and show weak or moderate covariances. The likelihood here was sampled using MCMC (with flat priors in p_Φ and $\ln(p_{DF})$ to turn the likelihood into a full posterior distribution function). 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.

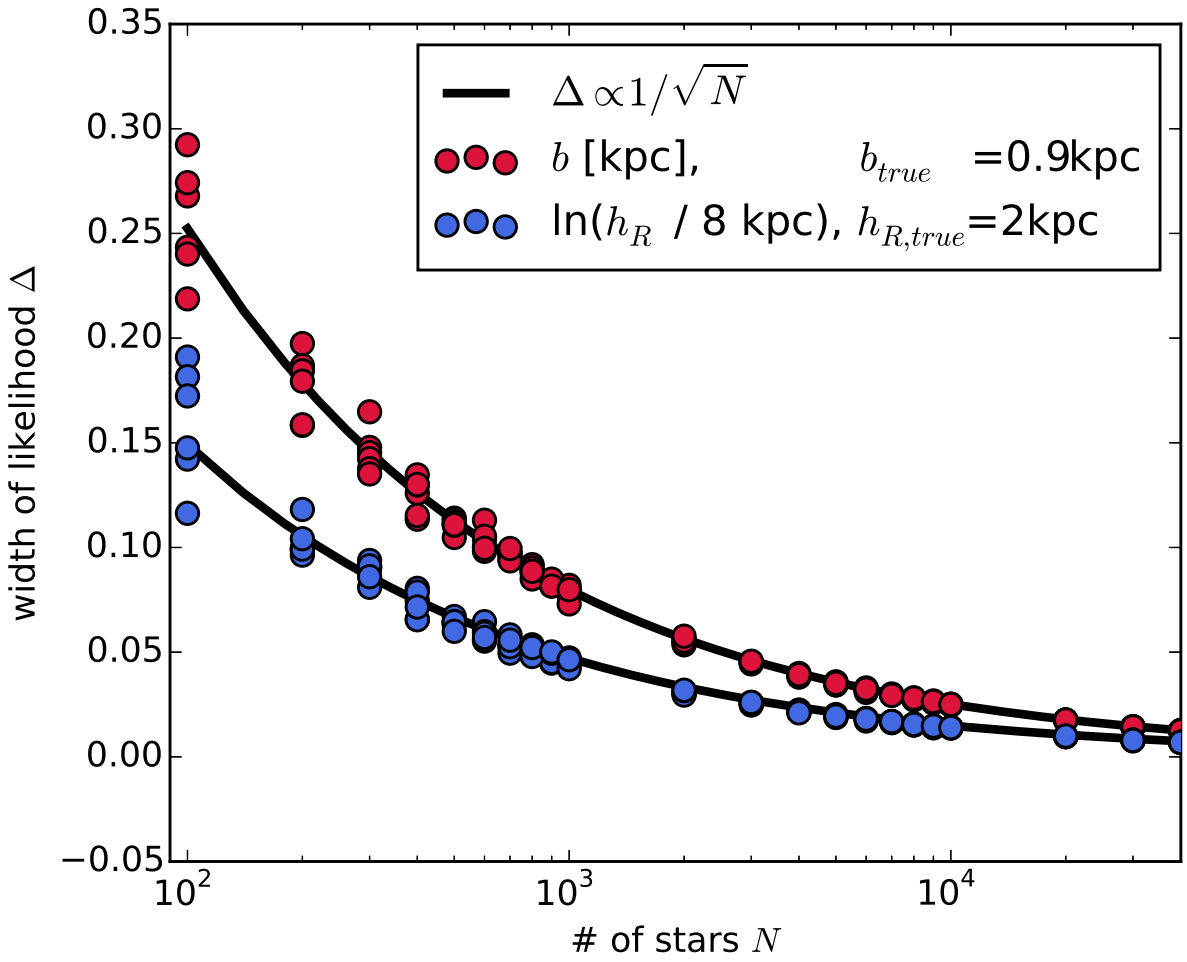

Fig. 3.— The width of the likelihood for two fit parameters found from analyses of 132 mock data sets vs. the number of stars in each data set. All mock data sets were created in an isochrone potential with $p_\Phi = \{v_{circ}, b\} = \{230 \text{ km/s }, 0.9 \text{ kpc }\}$, for a qDF with $p_{DF} = \{h_R, \sigma_R, \sigma_z, h_{\sigma_R}, h_{\sigma_z}\} = \{2 \text{ kpc }, 55 \text{ km/s }, 66 \text{ km/s }, 8 \text{ kpc }, 7 \text{ kpc }\}$ and within a spherical observation volume around the sun of radius $r = 3$ kpc . The data sets have different sample sizes and contain between 100 and 40,000 stars, as indicated on the x-axis. The likelihood was evaluated on a grid in the parameters $\{b, \ln(h_R/8\text{kpc}), \ln(\sigma_R/230\text{km/s}), \ln(h_{\sigma_R}/8\text{kpc})\}$, while all other parameters were assumed to be known and kept at their true values. For each fit parameter the likelihood was then marginalized by summing over the grid and then a Gauss curve was fitted to the marginalized likelihood. The standard deviation of these best fit Gaussians Δ is shown on the y-axis for b in kpc (red dots) and for $\ln(h_R/8\text{kpc})$ in dimensionless units (blue). The black lines are fits of the functional form $\Delta(N) \propto 1/\sqrt{N}$ 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 $1/\sqrt{N}$ as predicted by the central limit theorem.

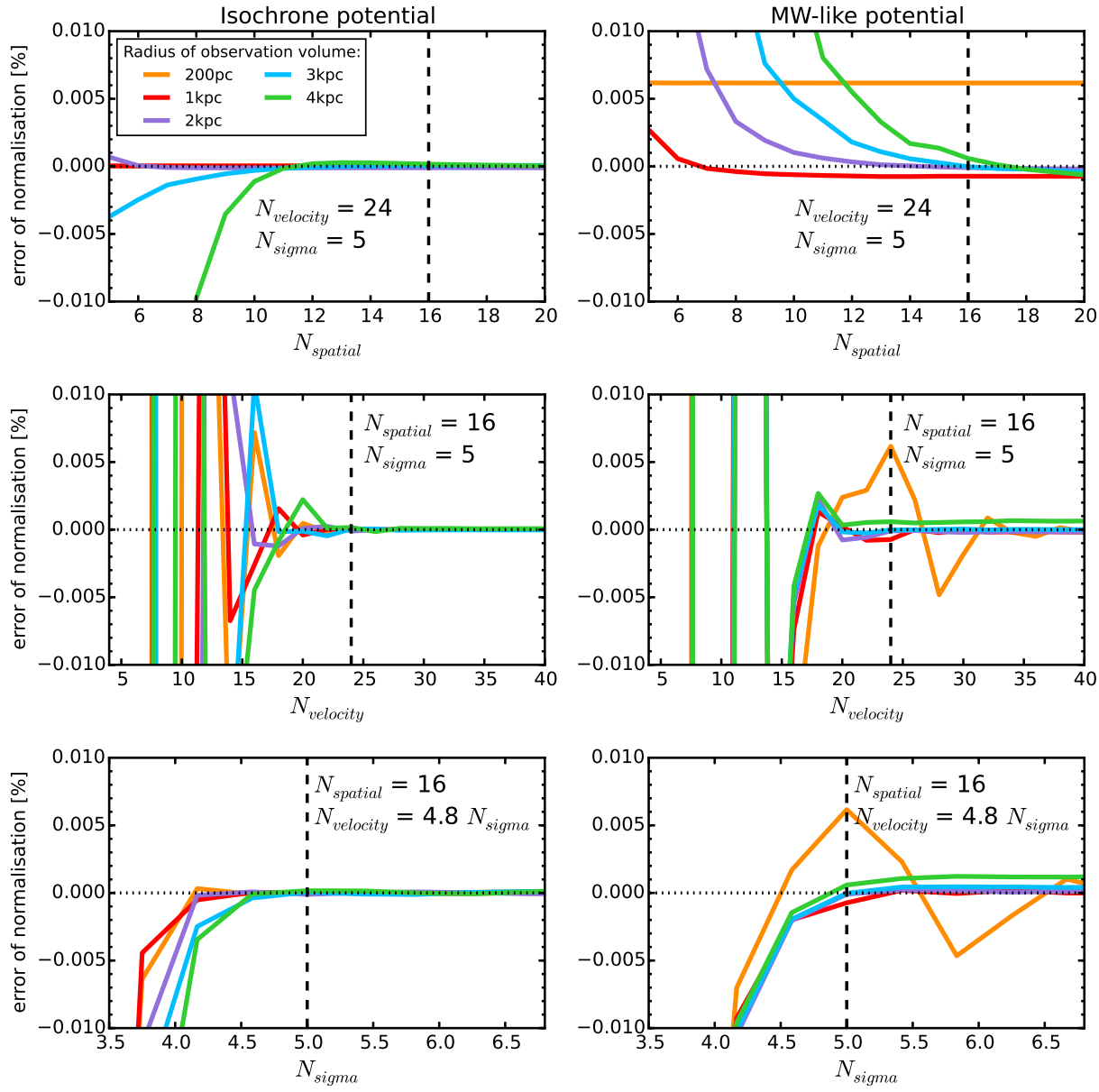

Fig. 4.— Relative error of the likelihood normalization in eq. (???) depending on the accuracy of the density calculation in §2.5.1. The different colors represent calculations for different radii of the spherical observation volume around the sun, as indicated in the legend. $N_{\rm spatial}$ is the 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. (2). At each (R, z) a Gauss-Legendre integration of order $N_{velocity}$ is performed over an integration range of $\pm N_{\rm spatial}$ times the dispersion in v_R and v_z and $[0, 1.5 v_{circ}(R_\odot)]$ in v_T. To integrate the interpolated density over the observed volume to arrive at the likelihood normalization in eq. (???), we perform a 40th-order Gauss-Legendre integration in each R and z direction. The distribution function that was evaluated for these plots has the parameters $p_{DF} = \{h_R, \sigma_R, \sigma_z, h_{\sigma_R}, h_{\sigma_z}\} = \{2 \text{ kpc }, 55 \text{ km/s }, 66 \text{ km/s }, 8 \text{ kpc }, 7 \text{ kpc }\}$. We show the results for two different potentials, an isochrone potential with parameters $p_\Phi = \{v_{circ}, b\} = \{230 \text{ km/s }, 0.9 \text{ kpc }\}$ and a MW-like potential (cf. ???) with parameters $p_\Phi = \{v_{circ}, R_d, z_h, f_h, \frac{\mathrm{d}\ln v_c}{\mathrm{d}\ln R}]\} = \{230 \text{ km/s }, 2.5 \text{ kpc }, 400 \text{ pc }, 0.8, 0\}$. (Caption continues on next page.)

Fig. 4.— (Continued.) We calculate the true normalization as $M_{\rm tot,true} \approx M_{\rm tot}(N_{\rm spatial} = 20, N_{\rm velocity} = 56, N_{\rm sigma} = 7)$, which has high enough accuracy. The relative error of the normalization is then calculated as $(M_{\rm tot}[N_{\rm spatial}, N_{\rm velocity}, N_{\rm sigma}] - M_{\rm tot,true})/M_{\rm tot,true}$. The dashed lines indicate the accuracy used in our analyses: it is better than 0.001% for both potential types (except the smallest volume in MW potential [TO DO: Why???]). Choosing $N_{\rm velocity}$ too small can have much more severe effects on the accuracy of the normalization than $N_{\rm spatial}$. A relative error of 0.3% can already introduce biases in the outcome of the analysis, that are, for a data set of $10,000$ stars already as large as a few sigma. When using a fiducial qDF to set the integration range over the velocity in the analysis (cf. §???) of an isochrone potential for 20,000 stars, already $N_{\rm sigma} = 4$ and $N_{\rm velocity} = 20$ behaves well: a precision of 0.05% in the normalisation leads to an error in the log-likelihood of ~ 1 [TO DO: Is this correct? $N\log(n + np) = N\log(n) + N\log(1 + p)$ with $N\log(1 + p) \sim 1$ for N=20000, p=0.00005. But why do we know that 1 is small enough? Argument is also different to Jo's argument with N_{stars}*precision=1, but I don't get that.???] [TO DO: Does this higher accuracy make the biases in MW potential analyses smaller????]

This preprint was prepared with the AAS LaTeX macros v5.2.

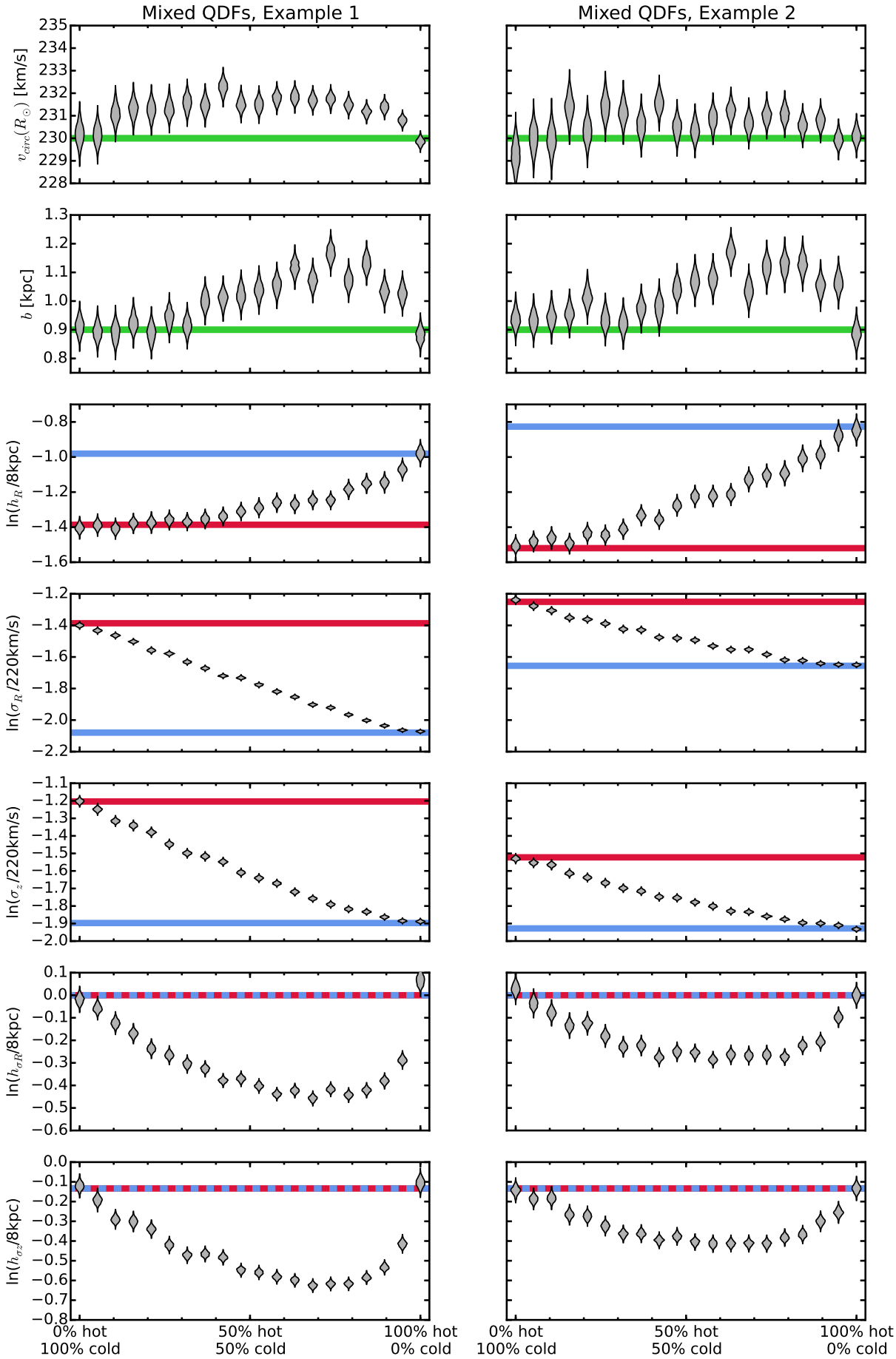

Fig. 5.— Caption [TO DO]