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; Binney & Tremaine 2008) [TO DO: other / better references???]. ~~The o~~bservational 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???]

~~Accurately determining~~ the Galactic gravitational potential is fundamental for understanding ~~its~~ dark matter and baryonic structure [TO DO: REF]. ~~Accurately determining~~ the stellar-population dependent orbit distribution function is a ~~fundamental~~ contraint on the

Summary of Comments on DynamicsPaper1_draft_v2.02.pdf

Page: 3

Author: rix Subject: Polygon Date: 30.08.15 13:37:18
This reads like the intro to a book, or thesis. The basic elements are there, and should be there. We should just perhaps be a bit more sparing with "far-reaching", "fundamenta;" etc..

Author: rix Subject: Polygonal Line Date: 30.08.15 13:33:10

Author: rix Subject: Inserted Text Date: 30.08.15 13:34:30
individual

Author: rix Subject: Inserted Text Date: 30.08.15 13:33:25
O

Author: rix Subject: Inserted Text Date: 30.08.15 13:33:40
is

Author: rix Subject: Cross-Out Date: 30.08.15 13:33:42

Author: rix Subject: Inserted Text Date: 30.08.15 13:33:59
from the Gaia mission [REF]

Author: rix Subject: Inserted Text Date: 30.08.15 13:35:00
position-velocity data of individual stars

Author: rix Subject: Inserted Text Date: 30.08.15 13:34:15
Sanders and Binney

Author: rix Subject: Cross-Out Date: 30.08.15 13:35:52

Author: rix Subject: Inserted Text Date: 30.08.15 13:36:29
Th Milky Way's

Author: rix Subject: Inserted Text Date: 30.08.15 13:37:29
And

Author: rix Subject: Inserted Text Date: 30.08.15 13:38:10
basic

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 ([TO DO: 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) [TO DO: 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 of stars measured by Gaia (Perryman et al. 2001) to unprecedented precision by the end of 2016. 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.~~

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 (e.g. Jeans modelling: Büdenbender et al. (2015);

Page: 4

Author: rix Subject: Polygonal Line Date: 30.08.15 13:38:10

Author: rix Subject: Cross-Out Date: 30.08.15 13:40:58

Author: rix Subject: Line Date: 30.08.15 13:42:30

This paragraph is good, but "too soon"

Author: rix Subject: Sticky Note Date: 30.08.15 13:41:51

There is a very strong emphasis (so far) on "local" references.

Loebman et al. (2012); action-based DF modelling: Bovy & Rix (2013); Piffl et al. (2014); Sanders & Binney (2015); torus modelling: Made-to-measure modelling: McMillan & Binney (2012, 2013), De Lorenzi et al. (2007); Syer & Tremaine (1996); Bissantz et al. (2004) or Hunt & Kawata (2014); ~~[TO DO: What kind of modelling is Xiangxiang doing?]; Xue et al. (2015))~~. 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 & Tremaine 2008; McMillan & Binney 2008; Binney 2010; Binney & McMillan 2011; Binney 2011b).

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). Ting et al. (2013) 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,c,d).

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 *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 ($\sim$ 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

Page: 5

Author: rix Subject: Cross-Out Date: 30.08.15 13:42:39

Author: rix Subject: Polygonal Line Date: 31.08.15 08:14:31

Author: rix Subject: Polygonal Line Date: 31.08.15 08:14:21

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 ~~explicitely~~ 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~~ (§2.5). (iii) ~~P~~arameter estimates become so precise~~, that we start to be able to distinguish between similar models. We therefore~~ want more flexibility and more free fit 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 recovery. (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 (§3.2). (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.

Page: 6

Author: rix Subject: Inserted Text Date: 31.08.15 08:15:20
explores and

Author: rix Subject: Cross-Out Date: 31.08.15 08:15:24

Author: rix Subject: Highlight Date: 31.08.15 08:16:08
This sentence, just stands by itself, and seems a bit out of context

Author: rix Subject: Cross-Out Date: 31.08.15 08:16:52

Author: rix Subject: Highlight Date: 31.08.15 08:17:25
if you introduce the term RoadMapping, use it, whenever applicable..

Author: rix Subject: Inserted Text Date: 31.08.15 08:16:36
the imminent

Author: rix Subject: Inserted Text Date: 31.08.15 08:16:49
data,

Author: rix Subject: Cross-Out Date: 31.08.15 08:17:45

Author: rix Subject: Inserted Text Date: 31.08.15 08:18:02
may be

Author: rix Subject: Cross-Out Date: 31.08.15 08:18:40
Unbiased, yes, but why does it have to be asymptotically normal?

Author: rix Subject: Inserted Text Date: 31.08.15 08:19:23
must not be an important source of systematics

Author: rix Subject: Inserted Text Date: 31.08.15 08:19:36
As p

Author: rix Subject: Inserted Text Date: 31.08.15 08:20:17
we need

Author: rix Subject: Cross-Out Date: 31.08.15 08:22:07

Author: rix Subject: Inserted Text Date: 31.08.15 08:22:05
We also explore how different aspects of the observational experiment design impact:

Author: rix Subject: Inserted Text Date: 31.08.15 08:22:45
systematically imperfect

Author: rix Subject: Inserted Text Date: 31.08.15 08:22:50
best

Author: rix Subject: Inserted Text Date: 31.08.15 08:36:25
individual

Author: rix Subject: Inserted Text Date: 31.08.15 08:38:09
do not encompass

(2013); Bovy & Rix (2013). ~~Second, and assuming *MAPs* do indeed follow the qDF, what would be the effect of pollution of *MAPs* through stars from neighbouring *MAPs* in the ([Fe/H],[α/Fe]) plane due to too big abundance errors or bin sizes.~~ ~~And last but not least~~ we test in §3.6 how well the modelling works, if our assumed potential family deviaties from the true potential.

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.

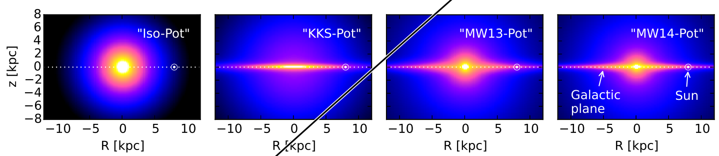

Fig. 1.— Density distribution of the four reference galaxy potentials in Table 1, for illustration purposes. These potentials are used throughout this work for mock data creation and potential recovery. [TO DO: Halo sichtbarer machen, evtl. mit isodensity contours]

Page: 7

Author: rix Subject: Cross-Out Date: 31.08.15 08:38:56
that can come in the main part..

Author: rix Subject: Cross-Out Date: 31.08.15 08:38:58

Author: rix Subject: Inserted Text Date: 31.08.15 08:39:03
Second,

Author: rix Subject: Sticky Note Date: 31.08.15 08:54:56
You might want to look at other paper's introduction: a standard part is the roadmap for the paper for the paper as a whole: In Section 2 we .., while in Section 3 we .... Section 4 finally ...

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

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

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 Equation (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 Equation (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.]

Page: 8

Author: rix Subject: Sticky Note Date: 02.09.15 09:11:03
Give a 1-2 sentence intro to the Section. E.g.

In this Section, we summarize the basic elements of the dynamical modelling presented here, which in many respects follows BR14.

Author: rix Subject: Cross-Out Date: 02.09.15 09:13:42

Author: rix Subject: Inserted Text Date: 02.09.15 09:11:30
,

Author: rix Subject: Inserted Text Date: 02.09.15 09:51:26
{vector J}\equiv

Author: rix Subject: Inserted Text Date: 02.09.15 09:11:43
,

Author: rix Subject: Inserted Text Date: 02.09.15 09:11:48
which

Author: rix Subject: Polygonal Line Date: 02.09.15 09:14:55

Author: rix Subject: Polygonal Line Date: 02.09.15 09:14:09

Author: rix Subject: Inserted Text Date: 02.09.15 09:34:52
excellent

Author: rix Subject: Cross-Out Date: 02.09.15 09:48:54

Author: rix Subject: Inserted Text Date: 02.09.15 09:36:21
and

Author: rix Subject: Inserted Text Date: 02.09.15 09:36:35
.

Author: rix Subject: Inserted Text Date: 02.09.15 09:36:43
potential

Author: rix Subject: Inserted Text Date: 02.09.15 09:37:10
page ....

Author: rix Subject: Inserted Text Date: 02.09.15 09:38:27
For the gravitational

Author: rix Subject: Inserted Text Date: 02.09.15 09:39:22
in

Author: rix Subject: Cross-Out Date: 02.09.15 09:41:29

Author: rix Subject: Inserted Text Date: 02.09.15 09:41:43
;

Author: rix Subject: Inserted Text Date: 02.09.15 09:42:20
; and

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

2.2. Distribution Function

Distribution Function. ~~M~~otivated by the findings of Bovy et al. (2012b,c,d) 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

$$
\begin{aligned}
&\mathrm{qDF}(\boldsymbol{J} \mid p_{\mathrm{DF}}) \\
&= f_{\sigma_R}(J_R, L_z \mid p_{\mathrm{DF}}) \times f_{\sigma_z}(J_z, L_z \mid p_{\mathrm{DF}}) \qquad (2)
\end{aligned}
$$

with

$$
\begin{aligned}
f_{\sigma_R}(J_R, L_z \mid p_{\mathrm{DF}}) &= n \times \frac{\Omega}{\pi \sigma_R^2(R_g)\kappa} \exp\left(-\frac{\kappa J_R}{\sigma_R^2(R_g)}\right) \\
&\quad \times [1 + \tanh(L_z/L_0)] \qquad (3) \\
f_{\sigma_z}(J_z, L_z \mid p_{\mathrm{DF}}) &= \frac{\nu}{2\pi\sigma_z^2(R_g)} \exp\left(-\frac{\nu J_z}{\sigma_z^2(R_g)}\right) \qquad (4) \\
& \qquad (5)
\end{aligned}
$$

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. (2012b,b,c) ~~about the 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~~,

$$
\begin{aligned}
n(R_g \mid p_{\mathrm{DF}}) &\propto \exp\left(-\frac{R_g}{h_R}\right) \qquad (6) \\
\sigma_R(R_g \mid p_{\mathrm{DF}}) &= \sigma_{R,0} \times \exp\left(-\frac{R_g - R_\odot}{h_{\sigma,R}}\right) \qquad (7) \\
\sigma_z(R_g \mid p_{\mathrm{DF}}) &= \sigma_{z,0} \times \exp\left(-\frac{R_g - R_\odot}{h_{\sigma,z}}\right). \qquad (8)
\end{aligned}
$$

Page: 9

Author: rix Subject: Inserted Text Date: 02.09.15 09:45:33
Stellar

Author: rix Subject: Inserted Text Date: 02.09.15 09:45:34
s

Author: rix Subject: Inserted Text Date: 02.09.15 09:50:35
This is m

Author: rix Subject: Polygonal Line Date: 02.09.15 09:50:25

Author: rix Subject: Inserted Text Date: 02.09.15 09:49:28
Throughout, we assume

Author: rix Subject: Inserted Text Date: 02.09.15 09:49:51
the orbits of

Author: rix Subject: Inserted Text Date: 02.09.15 09:50:02
can be described

Author: rix Subject: Cross-Out Date: 02.09.15 10:00:51

Author: rix Subject: Sticky Note Date: 02.09.15 09:53:27
. [end of sentence]

Author: rix Subject: Sticky Note Date: 02.09.15 09:59:26
in all BT08 reference, mention subsection.

Author: rix Subject: Cross-Out Date: 02.09.15 10:02:24

Author: rix Subject: Inserted Text Date: 02.09.15 10:00:46
To match the observed properties of MAPs

Author: rix Subject: Cross-Out Date: 02.09.15 10:03:18

Author: rix Subject: Inserted Text Date: 02.09.15 10:02:34
(see

Author: rix Subject: Inserted Text Date: 02.09.15 10:02:59
chose the functional forms

The qDF for each *MAP* has therefore a set of five free parameters p_{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 $h_{\sigma,R}$ and $h_{\sigma,z}$, that describe the radial decrease of the velocity dispersion. The *MAPs* we use for illustration through out this work are summarized in Table 2.

Tracer Density. One crucial point in our dynamical modelling technique (§2.5), as well as in creating mock data (§2.4), is to calculate the (axisymmetric) spatial tracer density $\rho_{\mathrm{DF}}(\boldsymbol{x} \mid p_\Phi, p_{\mathrm{DF}})$ 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_{velocity}-th order Gauss-Legendre quadrature for each integral:

$$\begin{aligned} &\rho_{\mathrm{DF}}(R, |z| \mid p_\Phi, p_{\mathrm{DF}}) \\ &= \int_{-\infty}^{\infty} \mathrm{qDF}(\boldsymbol{J}[R, z, \boldsymbol{v} \mid p_\Phi] \mid p_{\mathrm{DF}})\, \mathrm{d}^3\boldsymbol{v} \qquad (9) \\ &\approx \int_{-N_{\mathrm{sigma}}\sigma_R(R|p_{\mathrm{DF}})}^{N_{\mathrm{sigma}}\sigma_R(R|p_{\mathrm{DF}})} \int_{-N_{\mathrm{sigma}}\sigma_z(R|p_{\mathrm{DF}})}^{N_{\mathrm{sigma}}\sigma_z(R|p_{\mathrm{DF}})} \int_0^{1.5 v_{\mathrm{circ}}(R_\odot)} \\ &\quad \boxed{\mathrm{qDF}(J[R, z, \boldsymbol{v} \mid p_\Phi] \mid p_{\mathrm{DF}})\, \mathrm{d}v_T\, \mathrm{d}v_z\, \mathrm{d}v_R}, \qquad (10) \end{aligned}$$

where $\sigma_R(R \mid p_{\mathrm{DF}})$ and $\sigma_z(R \mid p_{\mathrm{DF}})$ are given by Equations (7) and (8) and the integration ranges are motivated by Figure 2. The integration range $[0, 1.5 v_{\mathrm{circ}}(R_\odot)]$ over v_T is in general sufficient (only for observation volumes at smaller Galactocentric radii with larger velocities this upper limit needs to be increased). For a given p_Φ and p_{DF} we explicitly calculate the density on $N_{\mathrm{spatial}} \times N_{\mathrm{spatial}}$ 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 $N_{\mathrm{spatial}}^2 \cdot N_{\mathrm{velocity}}^3$. §2.5 and Figure 3 show the importance of choosing N_{spatial}, N_{velocity} and N_{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 restframe 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 $\tilde{\boldsymbol{v}} \equiv (\mu_\alpha, \mu_\delta, v_{\mathrm{los}})$ as proper motions $\boldsymbol{\mu} = (\mu_\alpha, \mu_\delta)$ [TO DO: cos somwhere???] and

Page: 10

Author: rix Subject: Sticky Note Date: 02.09.15 10:04:13
In this section, you have so indicate somehow, where you recapitulate BR13, and what is added new..

Author: rix Subject: Line Date: 02.09.15 10:05:55

Author: rix Subject: Line Date: 02.09.15 10:06:00

Author: rix Subject: Line Date: 02.09.15 10:05:49

Author: rix Subject: Polygonal Line Date: 02.09.15 10:05:30

Author: rix Subject: Highlight Date: 02.09.15 10:09:04
Why z>0? Is it not symmetric in z, by construction?

Author: rix Subject: Highlight Date: 02.09.15 10:09:32
This seems out of place here; move to the beginning of the paragraph..

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

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

~~**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 $r_{\max}$ 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 $R \in [R_{\min}, R_{\max}], \phi \in [\phi_{\min}, \phi_{\max}], z \in [z_{\min}, z_{\max}]$ 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~~

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

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

Page: 11

Author: rix Subject: Highlight Date: 02.09.15 10:09:09

Author: rix Subject: Line Date: 02.09.15 10:10:10

Author: rix Subject: Inserted Text Date: 02.09.15 10:11:29
Any

Author: rix Subject: Inserted Text Date: 02.09.15 10:16:45
defining an effective sample

Author: rix Subject: Cross-Out Date: 02.09.15 10:20:47

Author: rix Subject: Inserted Text Date: 02.09.15 10:20:53
area

Author: rix Subject: Cross-Out Date: 02.09.15 10:25:12

Author: rix Subject: Inserted Text Date: 02.09.15 10:22:36

Author: rix Subject: Polygonal Line Date: 02.09.15 10:26:03

Author: rix Subject: Cross-Out Date: 02.09.15 10:26:15

Author: rix Subject: Cross-Out Date: 02.09.15 10:25:18

Author: rix Subject: Inserted Text Date: 02.09.15 11:30:47
.

Author: rix Subject: Inserted Text Date: 02.09.15 11:31:22
us

Author: rix Subject: Cross-Out Date: 02.09.15 11:32:21

Author: rix Subject: Inserted Text Date: 02.09.15 11:34:36
Here

Author: rix Subject: Cross-Out Date: 02.09.15 11:35:01

Author: rix Subject: Inserted Text Date: 02.09.15 11:34:57
, so it can be

Author: rix Subject: Cross-Out Date: 02.09.15 11:35:26

Author: rix Subject: Polygonal Line Date: 02.09.15 11:35:44

Author: rix Subject: Inserted Text Date: 02.09.15 11:35:22
simplify to

Author: rix Subject: Inserted Text Date: 02.09.15 14:53:35
We will rely on mock data as input to explore the limitations of the modelling. We

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 $\rho(R, |z| \mid p_\Phi, p_{\rm DF})$ generated by the given qDF and according to §2.2 and Equation (10). For the creation of the mock data we use $N_{\rm spatial} = 20$, $N_{\rm velocity} = 40$ and $N_{\rm sigma} = 5$.

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 pre-calculated $\rho_{\rm DF}(R, |z| \mid p_\Phi, p_{\rm DF})$. ~~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

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

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 $\mathrm{qDF}(R_i, z_i, \boldsymbol{v} \mid p_{Phi}, p_{\rm DF})$ 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:~~

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

~~**Example:**~~ Figure 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 $\sim \sigma_{R,0}$

Page: 13

Author: rix Subject: Inserted Text Date: 02.09.15 14:54:06
Subsequently,

Author: rix Subject: Polygonal Line Date: 02.09.15 14:57:47

Author: rix Subject: Cross-Out Date: 02.09.15 14:55:44

Author: rix Subject: Inserted Text Date: 02.09.15 14:56:23
" " space

Author: rix Subject: Inserted Text Date: 02.09.15 14:58:41
within the entire observable

Author: rix Subject: Inserted Text Date: 02.09.15 14:59:06
To

Author: rix Subject: Polygonal Line Date: 02.09.15 15:00:11

Author: rix Subject: Inserted Text Date: 02.09.15 14:59:42
resulting

Author: rix Subject: Inserted Text Date: 02.09.15 15:00:34
within the

Author: rix Subject: Inserted Text Date: 02.09.15 15:01:44
satisfying

Author: rix Subject: Polygonal Line Date: 02.09.15 15:01:57

Author: rix Subject: Cross-Out Date: 02.09.15 15:02:25

Author: rix Subject: Inserted Text Date: 02.09.15 15:05:48
The mock data from the

Author: rix Subject: Inserted Text Date: 02.09.15 15:06:25
lead to the expected distributions in configuration space

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]~~ 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 $L_z \sim R v_{\text{circ}}$ 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 Figure 2. [TO DO: Part of this could also be mentioned in the figure 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_{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 $\tilde{\boldsymbol{x}} = (\alpha, \delta, (m - M)), \tilde{\boldsymbol{v}} = (\mu_\alpha, \mu_\delta, v_{\text{los}})$.
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 perturb them according to the measurement errors as well.

Page: 14

Author: rix Subject: Cross-Out Date: 02.09.15 15:07:31
no, we don't

Author: rix Subject: Inserted Text Date: 02.09.15 15:07:45
illustrates

Author: rix Subject: Highlight Date: 02.09.15 15:08:11
Did you ever talk about "cool" and "hot" before?

Author: rix Subject: Cross-Out Date: 02.09.15 15:09:40

Author: rix Subject: Inserted Text Date: 02.09.15 15:13:02
cannot

Author: rix Subject: Cross-Out Date: 02.09.15 15:16:30

Author: rix Subject: Cross-Out Date: 02.09.15 15:16:19
You've said that before.

Author: rix Subject: Inserted Text Date: 02.09.15 15:16:49
, where we have taken m-M as a proxy for distance.

Author: rix Subject: Highlight Date: 02.09.15 15:15:31
Looks good to me.

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 $\mathrm{sf}(\boldsymbol{x})$,

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

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~~

$$\begin{aligned} p_{\mathrm{DF}} := \{ \quad & \ln\left(h_R/8\mathrm{kpc}\right), \\ & \ln\left(\sigma_{R,0}/220\mathrm{km\ s}^{-1}\right), \ln\left(\sigma_{z,0}/220\mathrm{km\ s}^{-1}\right), \\ & \ln\left(h_{\sigma,R}/8\mathrm{kpc}\right), \ln\left(h_{\sigma,z}/8\mathrm{kpc}\right)\}. \end{aligned}$$

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 Equation (9). The factor $\prod_i \mathrm{sf}(\boldsymbol{x}_i)$ is independent of the model parameters, so we simply evaluate Equation (11) in the likelihood calculation. We find the best set of model parameters by maximising the likelihood.

Page: 15

Author: rix Subject: Inserted Text Date: 02.09.15 15:21:36
Data

Author: rix Subject: Cross-Out Date: 02.09.15 15:21:39

Author: rix Subject: Inserted Text Date: 02.09.15 15:22:37
consider here

Author: rix Subject: Inserted Text Date: 02.09.15 15:25:25
is specified by a

Author: rix Subject: Inserted Text Date: 02.09.15 15:25:30
.

Author: rix Subject: Inserted Text Date: 02.09.15 15:25:50
For the

Author: rix Subject: Inserted Text Date: 02.09.15 15:26:26
we assume a prior that is flat in

Author: rix Subject: Sticky Note Date: 08.09.15 08:05:44
If the prior is flat, then you don't need the normalizations (8kpc etc..), and everything fits on one line.

Author: rix Subject: Cross-Out Date: 08.09.15 08:07:06
these quantities have been defined previously, no?

Author: rix Subject: Sticky Note Date: 08.09.15 08:11:42
what happened to sf(x_i) in the numerator of the last step? Correct or explain?

Author: rix Subject: Sticky Note Date: 08.09.15 08:12:17
B.t.w. Equation 9 and 10 are ONE equation.... shouldn't give it two numbers

Author: rix Subject: Inserted Text Date: 02.09.15 15:27:39
, and in

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, $v_{\mathrm{circ}}(R_\odot = 8.\mathrm{kpc}) \sim 230\mathrm{km\ s}^{-1}$, 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 $v_o := v_{\mathrm{circ}}(R_\odot)$. The prefactor $1/\left(r_o v_o\right)^3$ in Equation (11) makes sure that the likelihood has the correct units to satisfy:

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

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

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

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

In the case of an axisymmetric galaxy model and $\mathrm{sf}(\boldsymbol{x})=1$ 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 Equation (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 Equation (10) needed for the normalisation, which requires $N_{\mathrm{spatial}}^2 \times N_{\mathrm{velocity}}^3$ action calculations. The accuracy has to be chosen high enough, such that a resulting numerical error

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

does not dominate the likelihood, i.e.

$$\log \mathscr{L}(p_M \mid D) = \sum_i^N \log \mathrm{qDF}(\boldsymbol{J}_i \mid p_M) - 3N\log\left(r_o v_o\right) - N\log(M_{\mathrm{tot}\cancel{,\mathrm{true}}}) - N\log(1+\delta_{M_{tot}}), \tag{14}$$

Page: 16

Author: rix Subject: Polygon Date: 08.09.15 08:15:23
I think you need to use two different symbols, because else it is unclear what units v_circ has. In you first use here it has units of [1] in the second instant it has units of [km/s].
Use \tilde \hat or someting for the velocities and distances that have unit [1].

Author: rix Subject: Polygonal Line Date: 08.09.15 08:16:34

Author: rix Subject: Highlight Date: 08.09.15 08:18:56
Not sure I underdstand this: isn't it just the differential Volume. The "angular contribution" seems not well defined (to me)

Author: rix Subject: Polygonal Line Date: 08.09.15 08:20:57

Author: rix Subject: Inserted Text Date: 08.09.15 08:20:08
It turns out that the sufficiently

Author: rix Subject: Inserted Text Date: 08.09.15 08:21:11
even for

Author: rix Subject: Inserted Text Date: 08.09.15 08:21:47
This expense is dominated

Author: rix Subject: Inserted Text Date: 08.09.15 08:22:33
which in turn depends on

Author: rix Subject: Inserted Text Date: 08.09.15 08:22:41
in the sample

Author: rix Subject: Polygon Date: 08.09.15 08:24:08
There are many equations that seem needlessly split into two lines, for reasons that are not clear to me. Please try to put as many equations back into the most compact way to write them. Eithier within the text, or just one line.

Author: rix Subject: Cross-Out Date: 08.09.15 08:25:50

Author: rix Subject: Inserted Text Date: 08.09.15 08:25:31
,estimated

Author: rix Subject: Polygon Date: 08.09.15 08:27:15
This, too, can fit on one line

Author: rix Subject: Cross-Out Date: 08.09.15 08:26:06

with

$$N \log(1 + \delta_{M_{tot}}) \lesssim 1.$$

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. Figure 3 demonstrates that the numerical accuracy we use in the analysis, $N_{\text{spatial}} = 16$, $N_{\text{velocity}} = 24$ and $N_{sigma} = 5$, does satisfy this requirement.

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

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

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

$$\begin{aligned} &\tilde{P}(\boldsymbol{y}_i \mid p_\Phi, p_{\text{DF}}, \delta\boldsymbol{y}_i) \\ \equiv\ &\frac{\text{sf}(\boldsymbol{y}_i) \cdot \int \text{d}^6y' \ \text{DF}(\boldsymbol{y}') \cdot N[\boldsymbol{y}_i, \delta\boldsymbol{y}_i](\boldsymbol{y}')}{\int \text{d}^6y \ \text{DF}(\boldsymbol{y}) \cdot \int \text{d}^6y' \ \text{sf}(\boldsymbol{y}') \cdot N[\boldsymbol{y}, \delta\boldsymbol{y}_i](\boldsymbol{y}')}. \end{aligned}$$

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

$$\begin{aligned} &\tilde{P}(\boldsymbol{y}_i \mid p_\Phi, p_{\text{DF}}, \delta\boldsymbol{y}_i) \\ \approx\ &\frac{\text{sf}(\boldsymbol{x}_i)}{\int \text{d}^6y \ \text{DF}(\boldsymbol{y}) \cdot \text{sf}(\boldsymbol{x})} \cdot \frac{1}{N_{\text{error}}} \sum_n^{N_{\text{error}}} \text{DF}(\boldsymbol{x}_i, \boldsymbol{v}[\boldsymbol{y}'_{i,n}]) \end{aligned} \quad (15)$$

with

$$\boldsymbol{y}'_{i,n} \sim N[\boldsymbol{y}_i, \delta\boldsymbol{y}_i](\boldsymbol{y}')$$

Page: 17

Author: rix Subject: Polygon Date: 08.09.15 08:29:13
I would strongly advise against using linear and log likelihoods mixed in "talking" about them. Do you mean DeltalogLikelihood=1?

Author: rix Subject: Cross-Out Date: 08.09.15 08:27:28

Author: rix Subject: Sticky Note Date: 09.09.15 17:59:48
remember: Gaia will mostly give us motions (and distances), NOT detailed abundances; probably not even a/Fe

Author: rix Subject: Inserted Text Date: 09.09.15 17:58:35
one needs

Author: rix Subject: Sticky Note Date: 09.09.15 18:00:55
What I am missing in this Section is any distinction of what aspects are "new"
(not addressed in existing papers) and what is repaitulated to be coherent..

Author: rix Subject: Inserted Text Date: 09.09.15 18:00:01
Incorporating

Author: rix Subject: Sticky Note Date: 09.09.15 18:03:31
I am not sure I understand this notation [..]; also Y' seems undefined in the text. Why not call it y, as opposed to y_i,k

Author: rix Subject: Highlight Date: 09.09.15 18:04:04
something is screwy with the parentheses here...

Author: rix Subject: Cross-Out Date: 09.09.15 18:04:52

Fig. 3.— (Continued.) We calculate the true normalization with high accuracy as $M_{\rm tot,true} \approx M_{\rm tot}(N_{\rm spatial} = 20, N_{\rm velocity} = 56, N_{\rm sigma} = 7)$. The dashed lines indicate the accuracy used in our analyses: it is better than 0.002% for all three potential types. Only for the smallest volume in the "MW13-Pot" (yellow line) the error is only $\sim 0.005\%$. This could be due to the fact, that, while we have analytical formulas to calculate the actions for the isochrone and the Staeckel potential exactly, we have to resort to an approximate action calculation for the MW-like potential (see §2.1). [TO DO: Try to redo yellow curve in MW. Weird, that it does not depend on $N_{spatial}$.???] [TO DO: fancybox Legend]

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, $y'_{i,n}$.

2.6. Fitting Procedure

We search the $(p_\Phi, p_{\rm DF})$ parameter space for the maximum of the likelihood in Equation (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.
- *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 Equation (11) for all qDF parameters $p_{\rm DF}$ with the actions in the given potential and (analogously to Figure 9 in Bovy & Rix (2013)).

Page: 19

Author: rix Subject: Highlight Date: 09.09.15 18:06:26

Author: rix Subject: Highlight Date: 09.09.15 18:06:33

- *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 $N_j \times N_{\text{error}}$ stellar data actions $\boldsymbol{J}_i$. To calculate the normalisation in Equation (11), $N^2_{\text{spatial}} \times N^3_{\text{velocity}}$ actions $\boldsymbol{J}_n$ are needed. Formally the spatial coordinates at which the $\boldsymbol{J}_n$ are calculated depend on the p_{DF} parameters via the integration ranges in Equation (10). To relax this dependence we instead use the same velocity integration limits in the likelihood calculations for all p_{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 Equation (7) and (8), 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_{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_{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_{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 $N_j \times N_{\text{error}} + N^2_{\text{spatial}} \times N^3_{\text{velocity}}$ 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.

Page: 20

Author: rix Subject: Inserted Text Date: 09.09.15 18:12:49
Computational expense

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 Foreman-Mackey et al. (2013) 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 Equation (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 within the MCMC as well. As fiducial qDF we use the qDF parameters of the likelihood peak, found by the nested-grid search.

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?

Page: 21

Author: rix Subject: Inserted Text Date: 09.09.15 18:20:26
s

Author: rix Subject: Inserted Text Date: 09.09.15 18:21:01
as a power

Author: rix Subject: Inserted Text Date: 09.09.15 18:21:05
the

Author: rix Subject: Sticky Note Date: 09.09.15 18:23:10
The likelihood (units: per data) and the posterior (units: per parameters) have different units. One can use them almost interchangably, if one assumed flat priors. We should state the priors to be clear). Here you call it first a likelihood, and then -- as apology -- say, well, it's kind of a poterior...

Author: rix Subject: Cross-Out Date: 09.09.15 18:23:33

Author: rix Subject: Polygon Date: 09.09.15 18:26:50
This has all been said before (intro) and will be addressed inturn in teh subsequent section; so, no need to spend half a page on it.

I would just say:
the limitations of action based modelling posed in the introduction (unbiased estimates; survey volume; imperfect selection function; measurement errors; actual DF or Phi not spanned by the space of models).

Author: rix Subject: Cross-Out Date: 09.09.15 18:23:50

Author: rix Subject: Inserted Text Date: 09.09.15 18:24:00
even

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.

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 Equation 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 3.2, 3.3 and 3.1, we verified that *RoadMapping* satisfies all these conditions and expectations. Figure 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. Figure 5 then illustrates that the *pdf* width, SE, indeed scales as $1/\sqrt{N_{\rm sample}}$. Figure 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.

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 two suites of mock data sets:

Page: 22

Author: rix Subject: Inserted Text Date: 09.09.15 18:33:24
s

Author: rix Subject: Polygonal Line Date: 09.09.15 18:27:18

Author: rix Subject: Cross-Out Date: 09.09.15 18:33:37

Author: rix Subject: Highlight Date: 09.09.15 18:34:44
did that child not have a name in Sectin 2.5? \deltaM_{tot} ?

Author: rix Subject: Inserted Text Date: 09.09.15 18:49:12
considerably

Author: rix Subject: Inserted Text Date: 09.09.15 18:50:06
the

Author: rix Subject: Highlight Date: 09.09.15 19:55:18
..I don't understand this here: both gramatically and