ACTION-BASED DYNAMICAL MODELLING FOR THE MILKY WAY DISK

Wilma H. Trick[1,2], Jo Bovy[3], and Hans-Walter Rix[1]

Draft version October 23, 2015

ABSTRACT

We present *RoadMapping*, a full-likelihood dynamical modelling machinery that aims to recover the Milky Way's (MW) gravitational potential from ~~stellar sub-populations~~ in the Galactic disk. *RoadMapping* models the observed positions and velocities of stars with ~~an equilibrium,~~ three-integral distribution function (DF) in an axisymmetric potential ~~and accounts also~~ for survey selection effects. In preparation for the application to large data sets of ~~modern~~ surveys like Gaia, we create and analyze a large suite of mock data sets. Based on this we develop qualitative "rules of thumb" for which characteristics and limitations of data, model and machinery affect constraints on the potential and DF most. Overall we find that the potential can be reliably recovered if the model assumptions are fulfilled or even slightly wrong. *RoadMapping* gives constraints of high precision (i) for large sample sizes, (ii) for survey volumes of large radial and vertical coverage, and (iii) as long as measurement uncertainties are ~~perfectly~~ known (even for proper motion uncertainties up to $\delta\mu \sim 5$ mas yr^{-1}). Unbiased potential estimates are ensured, (i) for small to moderate misjudgements of the spatial selection function, (ii) if distances are known to within 10% (at least for distances smaller 3 kpc and $\delta\mu \lesssim 2$ mas yr^{-1}), and (iii) if proper motion uncertainties are known within 10% (at least for $\delta\mu \lesssim 3$ mas yr^{-1}). ~~Minor~~ differences between the true and ~~assumed~~ DF are ~~acceptable~~. When defining sub-populations by binning stars according to their chemical abundances, finite bin sizes and abundance errors should not affect the modelling as long as the DF parameters of neighbouring bins do not vary more than 20%. While hotter populations are less affected by pollution and misjudgements of $\delta\mu$, cooler populations recover the Galactic rotation curve more reliably. If the MW's true gravitational potential is not included in the assumed family of parametrized model potentials, we can—at least in the axisymmetric case—still find a potential that ~~is a reliable fit~~ within the limitations of the model. Challenges of the future are the rapidly increasing computational costs for high precision likelihood evaluations required for large sample sizes.

Keywords: Galaxy: disk — Galaxy: fundamental parameters — Galaxy: kinematics and dynamics — Galaxy: structure

1. INTRODUCTION

Dynamical modelling can be employed to infer the Milky Way's (MW) gravitational potential from stellar motions (Binney & Tremaine 2008; Binney 2011; Rix & Bovy 2013). Observational information on the 6D phase-space coordinates of stars is currently growing at a rapid pace, and will be taken to a whole new level in ~~number~~ and precision by the upcoming data from the Gaia mission (Perryman et al. 2001). Yet, rigorous and practical modelling tools that turn position-velocity data of individual stars into constraints both on the gravitational potential and on the distribution function (DF) of stellar orbits are scarce (Rix & Bovy 2013).

The Galactic gravitational potential is fundamental for understanding the MW's dark matter and baryonic structure (Rix & Bovy 2013; McMillan 2012; Strigari 2013; Read 2014) and the stellar-population dependent orbit DF is a basic constraint on the Galaxy's formation history (Binney 2013; ~~Rix & Bovy 2013;~~ Sanders & Binney 2015).

There is a variety of practical approaches to dynamical modelling of discrete collisionless tracers, such as the stars in the MW, e.g., Jeans modelling (Kuijken & Gilmore 1989; Bovy & Tremaine 2012; Garbari et al. 2012; Zhang et al. 2013; Büdenbender et al. 2015), action-based DF modelling (Bovy & Rix 2013; Piffl et al. 2014; Sanders & Binney 2015), torus modelling (McMillan & Binney 2008; McMillan & Binney 2012; McMillan & Binney 2013), made-to-measure modelling (Syer & Tremaine 1996; de Lorenzi et al. 2007; or Hunt & Kawata 2014). Most of them–explicitly or implicitly–describe the stellar distribution through a DF.

Recently, Binney (2012b) and Bovy & Rix (2013) proposed to constrain the MW's gravitational potential by combining parametrized axisymmetric potential models with DFs that are simple analytic functions of the three orbital actions (Binney & Tremaine 2008, §3.5 & §4.6; Binney 2011) to model discrete data.

Bovy & Rix (2013) (BR13 hereafter) put this in practice by implementing a rigorous modelling approach for so-called mono-abundance populations (MAPs), i.e, sub-sets of stars with similar [Fe/H] and [α/Fe] within the Galactic disk, which seem to allow simple DFs (Bovy et al. 2012b,c,d). Given an assumed (axisymmetric) model for the Galactic potential and action-based DF (Binney 2010; Binney & McMillan 2011; Ting et al. 2013) they calculated the likelihood of the observed $(\vec{x}, \vec{v})$ for each MAP among SEGUE G-dwarf stars

[1] Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany
[2] Correspondence should be addressed to trick@mpia.de.
[3] Department of Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON, M5S 3H4, Canada

Summary of Comments on DynamicsPaper1_draft_v3.01_HWRcomments.pdf

Page: 1

Author: rix Subject: Inserted Text Date: 23.10.15 19:23:43
large samples of stars

Author: rix Subject: Inserted Text Date: 23.10.15 19:28:59
a parameterized

Author: rix Subject: Inserted Text Date: 23.10.15 19:29:11
parameterized

Author: rix Subject: Inserted Text Date: 23.10.15 19:29:40
, accounting

Author: rix Subject: Cross-Out Date: 23.10.15 19:24:07

Author: rix Subject: Inserted Text Date: 23.10.15 19:26:03
and accuracy

Author: rix Subject: Inserted Text Date: 23.10.15 19:26:18
well

Author: rix Subject: Inserted Text Date: 23.10.15 19:26:38
only

Author: rix Subject: Cross-Out Date: 23.10.15 19:27:17

Author: rix Subject: Inserted Text Date: 23.10.15 19:27:14
Modest systematic

Author: rix Subject: Inserted Text Date: 23.10.15 19:27:38
the best-fit model

Author: rix Subject: Inserted Text Date: 23.10.15 19:27:50
inconsequential

Author: rix Subject: Inserted Text Date: 23.10.15 19:30:26
robustly approximates the potential

Author: rix Subject: Inserted Text Date: 23.10.15 19:31:31
quantity

Author: rix Subject: Inserted Text Date: 23.10.15 19:33:08
Not all of them avoid binning to exploit the full discrete information content of the data.

Author: rix Subject: Polygonal Line Date: 23.10.15 19:31:57

Author: rix Subject: Line Date: 23.10.15 19:32:16

Author: rix Subject: Inserted Text Date: 23.10.15 19:33:58
fo

(Yanny et al. 2009). They also accounted for the complex, but known selection function of the kinematic tracers (Bovy et al. 2012d). For each MAP the modelling resulted in an independent estimate on the same gravitational potential. Taken as an ensemble, they constrained the disk surface mass density over a wide range of radii ($\sim 4-9$ kpc), and proved to be a powerful constraint on the disk mass scale length and on the disk-to-dark-matter ratio at the Solar radius.

BR13 made however a number of quite severe and idealizing assumptions about potential, DF and the knowledge of observational effects. These idealizations ~~are likely to~~ translate into systematic errors on the inferred potential, well above the formal error bars of the upcoming surveys with their wealth and quality of data.

In this work we present *RoadMapping* ("Recovery of the Orbit Action Distribution of Mono-Abundance Populations and Potential INference for our Galaxy")—an improved, refined, flexible, robust and well-tested version of the original dynamical modelling machinery by BR13, explicitly developed to deal with large data sets. Our goal is to explore which of the assumptions BR13 made and which other aspects of data, model and machinery limit *RoadMapping*'s recovery of the true gravitational potential.

We investigate the following aspects of the *RoadMapping* machinery that become especially important for a large number of stars: (i) Numerical inaccuracies must not be an important source of systematics (Section 2.6). (ii) As parameter estimates become much more precise, we need more flexibility in the potential and DF model and effective strategies to find the best fit parameters. The improvements made in *RoadMapping* as compared to the machinery used in BR13 are presented in Section 2.7. (iii) ~~We have to make sure that~~ *RoadMapping* ~~is an~~ unbiased estimator (Section 3.1).

We also explore how different aspects of the observational experiment design impact the parameter recovery: (i) ~~It might be worth to explore~~ the importance of the survey volume geometry, size, shape and position within the MW to constrain the potential (Section 3.2). (ii) What if our knowledge of the sample selection function is imperfect, and potentially biased (Section 3.3)? (iii) How to best account for individual and possibly misjudged measurement uncertainties (Section 3.4)? (iv) Given several stellar sub-populations of different kinematic temperature—what is the best choice (Section 3.7)?

One of the strongest assumptions is to restrict the dynamical modelling to a certain family of parametrized ~~models.~~ We investigate how well we can hope to recover the true potential, when our models do not encompass the true DF (Section 3.5) and potential (Section 3.6).

~~The most severe idealization that goes into this kind of dynamical modelling might be that of the Galaxy being axisymmetric and 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.~~

For all of the above 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 Sections 2.1-2.4 following the procedure in Section 2.5, analysed according to the description of the *RoadMapping* machinery in Sections 2.6-2.7. The results on the investigated modelling aspects are presented in Section 3 and summarized and discussed in Section 4.

2. DYNAMICAL MODELLING

In this section we summarize the basic elements of *RoadMapping*, the dynamical modelling machinery presented in this work, which in many respects follows BR13 and makes extensive use of the `galpy` Python package[4] (Bovy 2015).

2.1. *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 heliocentric coordinates, position $\tilde{\boldsymbol{x}} \equiv (\mathrm{RA}, \mathrm{Dec}, m-M)$ in right ascension RA, declination Dec and distance modulus $(m-M)$ as proxy for the distance from the Sun, and velocity $\tilde{\boldsymbol{v}} \equiv (\mu_{\mathrm{RA}} \cdot \cos(\mathrm{Dec}), \mu_{\mathrm{Dec}}, v_{\mathrm{los}})$ as proper motions $\boldsymbol{\mu} = (\mu_{\mathrm{RA}} \cdot \cos(\mathrm{Dec}), \mu_{\mathrm{Dec}})$ and line-of-sight velocity v_{los}, the data $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{v}})$ has to be converted into the Galactocentric rest-frame coordinates $(\boldsymbol{x}, \boldsymbol{v})$ using the Sun's position and velocity. We assume for the Sun

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

2.2. *Actions and potential models*

Orbits in an axisymmetric gravitational potential Φ are best described and fully specified by the three actions $\boldsymbol{J} \equiv (J_R, J_z, J_\phi \equiv L_z)$ (Binney & Tremaine 2008, §3.5). ~~Their~~ computation from a star's phase-space coordinates, $(\boldsymbol{x}, \boldsymbol{v}) \longrightarrow \boldsymbol{J}$, is typically very expensive. The spherical isochrone potential (Henon 1959) and axisymmetric Stäckel potential (de Zeeuw 1985) are the most general (Galactic) potentials, that allow exact action calculations (Binney & Tremaine 2008, §3.5.2 and §3.5.3). In all other potentials actions have to be numerically estimated. We use the *Stäckel fudge* by Binney (2012a) for axisymmetric potentials and action interpolation grids (Bovy 2015; Binney 2012a) to speed up the calculation. The latter is one of the improvements employed by *RoadMapping*, which was not used in BR13.

For the gravitational potential in our modelling we assume a family of parametrized models. We use: The MW-like potential from BR13 (`MW13-Pot`) with bulge, disks and halo; the spherical isochrone potential (`Iso-Pot`); and the 2-component Kuzmin-Kutuzov Stäckel potential (Batsleer & Dejonghe 1994; `KKS-Pot`), which also displays a disk and halo structure. Table 1 summarizes all reference potentials used in this work together with their free parameters p_Φ. The density distribution of these potentials is illustrated in Figure 1.

2.3. *Stellar distribution functions*

The action-based quasi-isothermal distribution function (qDF) by Binney (2010) and Binney & McMillan (2011) is a simple DF which we will employ as a specific

[4] `galpy` is an open-source code that is being developed on `http://github.com/jobovy/galpy`. The latest documentation can be found at `http://galpy.readthedocs.org/en/latest/`.

Page: 2

Author: rix Subject: Inserted Text Date: 23.10.15 19:34:36
could plausibly

Author: rix Subject: Inserted Text Date: 23.10.15 20:26:44
Action

Author: rix Subject: Cross-Out Date: 23.10.15 19:36:53

Author: rix Subject: Inserted Text Date: 23.10.15 19:37:09
should be an

Author: rix Subject: Inserted Text Date: 23.10.15 19:41:11
What is

Author: rix Subject: Inserted Text Date: 23.10.15 19:41:21
,

Author: rix Subject: Inserted Text Date: 23.10.15 19:41:27
,

Author: rix Subject: Inserted Text Date: 23.10.15 19:42:03
functions for the gravitational potential and the DF.

Author: rix Subject: Cross-Out Date: 23.10.15 19:43:07
The place to say what you DON'T do is in the discussion/conclusion section not the intro

Table 2
Reference parameters for the qDF in Equations 1-6, used to create 6D phase-space mock data sets for stellar populations of different kinematic temperature. The parameters of the `cooler` & `colder` (`warmer`) qDFs were chosen to have the same anisotropy $\sigma_{R,0}/\sigma_{z,0}$ as the `hot` (`cool`) qDF, with X being a free parameter describing the temperature difference. Hotter populations have shorter tracer scale lengths (Bovy et al. 2012d) and the velocity dispersion scale lengths were fixed according to Bovy et al. (2012c).

name	qDF parameters $p_{\rm DF}$				
	h_R [kpc]	$\sigma_{R,0}$ [km s^{-1}]	$\sigma_{z,0}$ [km s^{-1}]	$h_{\sigma,R}$ [kpc]	$h_{\sigma,z}$ [kpc]
`hot`	2	55	66	8	7
`cool`	3.5	42	32	8	7
`cooler`	3	27.5	33	8	7
`colder`	$2+X\%$	$55-X\%$	$66-X\%$	8	7
`warmer`	$3.5-X\%$	$42+X\%$	$32+X\%$	8	7

The term $[1+\tanh(L_z/L_0)]$ suppresses counter-rotation for orbits in the disk with $L \gg L_0$ (with $L_0 = 10 \times R_\odot/8 \times v_{\rm circ}(R_\odot)/220$).

Following BR13, we choose the functional forms

$$n(R_g \mid p_{\rm DF}) \propto \exp\left(-\frac{R_g}{h_R}\right) \tag{4}$$

$$\sigma_R(R_g \mid p_{\rm DF}) = \sigma_{R,0} \times \exp\left(-\frac{R_g - R_\odot}{h_{\sigma,R}}\right) \tag{5}$$

$$\sigma_z(R_g \mid p_{\rm DF}) = \sigma_{z,0} \times \exp\left(-\frac{R_g - R_\odot}{h_{\sigma,z}}\right), \tag{6}$$

which indirectly set the stellar number density and radial and vertical velocity dispersion profiles. The qDF has therefore a set of five free parameters $p_{\rm DF}$: the density scale length of the tracers h_R, the radial and vertical velocity dispersion at the Solar position $R_\odot$, $\sigma_{R,0}$ and $\sigma_{z,0}$, and the scale lengths $h_{\sigma,R}$ and $h_{\sigma,z}$, that describe the radial decrease of the velocity dispersion. *RoadMapping* allows to fit any number of DF parameters simultaneously, while BR13 kept $\{\sigma_{R,0}, h_{\sigma,R}\}$ fixed. Throughout this work we make use of a few example stellar populations whose qDF parameters are given in in Table 2: Most tests use the `hot` and `cool` qDFs, which correspond to kinematically hot and cool populations, respectively.

One ~~crucial point~~ in our dynamical modelling technique (Section 2.6), as well as in creating mock data (Section 2.5), is to calculate the (axisymmetric) spatial tracer density $\rho_{\rm DF}(\boldsymbol{x} \mid p_\Phi, p_{\rm DF})$ for a given DF and potential. Analogously to BR13,

$$\begin{aligned} &\rho_{\rm DF}(R,|z| \mid p_\Phi, p_{\rm DF}) \\ &= \int_{-\infty}^{\infty} {\rm DF}(\boldsymbol{J}[R,z,\boldsymbol{v} \mid p_\Phi] \mid p_{\rm DF})\,{\rm d}^3\boldsymbol{v} \\ &\approx \int_{-n_\sigma\sigma_R(R|p_{\rm DF})}^{n_\sigma\sigma_R(R|p_{\rm DF})} \int_{-n_\sigma\sigma_z(R|p_{\rm DF})}^{n_\sigma\sigma_z(R|p_{\rm DF})} \int_0^{1.5 v_{\rm circ}(R_\odot)} \\ &\qquad {\rm DF}(\boldsymbol{J}[R,z,\boldsymbol{v} \mid p_\Phi] \mid p_{\rm DF})\,{\rm d}v_T\,{\rm d}v_z\,{\rm d}v_R, \end{aligned} \tag{7}$$

where $\sigma_R(R \mid p_{\rm DF})$ and $\sigma_z(R \mid p_{\rm DF})$ are given by Equations 5 and 6.[5] Each integral is evaluated using a N_v-th order Gauss-Legendre quadrature. For a given p_Φ and $p_{\rm DF}$ we explicitly calculate the density on $N_x \times N_x$ regular grid points in the (R,z) plane and interpolate in between using bivariate spline interpolation. The grid is chosen to cover the extent of the observations (for $|z| \geq 0$, because the model is symmetric in z by construction). The total number of actions to be calculated to set up the density interpolation grid is $N_x^2 \times N_v^3$, which is one of the computation speed limiting factors. To complement the work by BR13, we will specifically work out in Section 2.6 and Figure 4 how large N_x, N_v and n_σ have to be chosen to get the density with a sufficiently high numerical accuracy.

[5] The integration ranges over the velocities are motivated by Figure 3 and n_σ should be chosen as $n_\sigma \sim 5$. The integration range $[0, 1.5 v_{\rm circ}(R_\odot)]$ over v_T is in general sufficient, only for observation volumes with larger mean stellar v_T this upper limit needs to be increased.

2.4. *Selection functions*

Any survey's selection function (SF) can be understood as defining an effective sample sub-volume in the space of observables, e.g., position on the sky (limited by the pointing of the survey), distance from the Sun (limited by brightness and detector sensitivity), colors and metallicity of the stars (limited by survey mode and targeting). In our modelling we use simple spatial SFs, which describe the probability to observe a star at position $\boldsymbol{x}$,

$${\rm SF}(\boldsymbol{x}) \equiv \begin{cases} {\rm completeness}(\boldsymbol{x}) & \text{if } \boldsymbol{x} \text{ within obs. volume,} \\ 0 & \text{if } \boldsymbol{x} \text{ outside.} \end{cases}$$

The SF of the SEGUE survey (Bovy et al. 2012d) used by BR13 consists of many pencil-beams. In anticipation of large contiguous volume surveys like Gaia, we use SFs that span large observed volumes of simple geometrical shapes: a sphere of radius $r_{\rm max}$ with the Sun at its center; or an angular segment of an cylindrical annulus (wedge), i.e., the volume with $R \in [R_{\rm min}, R_{\rm max}], \phi \in [\phi_{\rm min}, \phi_{\rm max}], z \in [z_{\rm min}, z_{\rm max}]$ within the model Galaxy. The sharp outer edge of the survey volume could be interpreted as a detection limit in apparent brightness in the case where all stars have the same luminosity. We set $0 \leq {\rm completeness}(\boldsymbol{x}) \leq 1$ everywhere inside the observed volume, so it can be understood as a position-dependent detection probability. Unless explicitly stated otherwise, we simplify to ${\rm completeness}(\boldsymbol{x}) = 1$.

2.5. *Mock data*

We will rely on mock data as input to explore the limitations of the modelling. We assume that the positions and velocities of our stellar mock sample are indeed drawn from our assumed family of potentials and DFs (with given parameters p_Φ and $p_{\rm DF}$). The DF is in

Page: 4

Author: rix Subject: Inserted Text Date: 23.10.15 20:29:21
(spatial)

Author: rix Subject: Inserted Text Date: 23.10.15 20:28:31
indispensable step

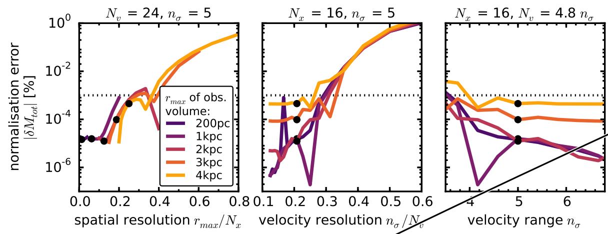

Figure 4. Relative error of the likelihood normalization, $\delta M_{\rm tot}$, in Equation 11 depending on the accuracy of the grid-based density calculation in Equation 7 (and surrounding text) in five spherical observation volumes with different radius $r_{\rm max}$. (Test 2 in Table 3 summarizes the model parameters.) The tracer density in Equation 7 is calculated on $N_x \times N_x$ spatial grid points in $R \in [R_\odot \pm r_{\rm max}]$ and $|z| \in [0, r_{\rm max}]$. The integration over the velocities is performed with Gauss-Legendre quadratures of order N_v within an integration range of $\pm n_\sigma$ times the dispersion $\sigma_R(R)$ and $\sigma_z(R)$ (and $[0, 1.5 v_{\rm circ}]$ in v_T). (We vary N_x, N_v and n_σ separately and keep the other two fixed at the values indicated above each panel.) We calculate the "true" normalization $M_{\rm tot}$ in Equation 11 with high accuracy as $M_{\rm tot} \equiv M_{\rm tot,approx}(N_x = 20, N_v = 56, n_\sigma = 7)$. The black dots indicate the accuracy used in our analyses: It is better than 0.001% (dotted line). We find that the accuracy depends on the *spatial* resolution of the grid and requires more accurate integrations over the *velocity* for larger volumes within which the density varies more strongly.

selection function $\mathrm{SF}(\boldsymbol{x})$,

$$D \equiv \{\boldsymbol{x}_i, \boldsymbol{v}_i \mid \quad (\text{star } i \text{ ~~being~~ in given sub-population}) \\ \wedge (\mathrm{SF}(\boldsymbol{x}_i) > 0)\}.$$

We fit a model potential and DF (here: the qDF) which are specified by a number of fixed and free model parameters,

$$p_M \equiv \{p_{\rm DF}, p_\Phi\}.$$

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 DF evaluated for the i-th star is then $\mathrm{DF}(\boldsymbol{J}_i \mid p_M) := \mathrm{DF}(\boldsymbol{J}[\boldsymbol{x}_i, \boldsymbol{v}_i \mid p_\Phi] \mid p_{\rm DF})$.

The likelihood of the data given the model is, following BR13,

$$\begin{aligned} &\mathscr{L}(D \mid p_M) \\ &\equiv \prod_i^{N_*} p(\boldsymbol{x}_i, \boldsymbol{v}_i \mid p_M) \\ &= \prod_i^{N_*} \frac{\mathrm{DF}(\boldsymbol{J}_i \mid p_M) \cdot \mathrm{SF}(\boldsymbol{x}_i)}{\int \mathrm{d}^3x\, \mathrm{d}^3v\ \mathrm{DF}(\boldsymbol{J} \mid p_M) \cdot \mathrm{SF}(\boldsymbol{x})} \\ &\propto \prod_i^{N_*} \frac{\mathrm{DF}(\boldsymbol{J}_i \mid p_M)}{\int \mathrm{d}^3x\ \rho_{\rm DF}(R, |z| \mid p_M) \cdot \mathrm{SF}(\boldsymbol{x})}, \end{aligned} \tag{8}$$

where N_* is the number of stars in D, and in the last step we used Equation 7. $\prod_i \mathrm{SF}(\boldsymbol{x}_i)$ is independent of p_M, so we treat it as unimportant proportionality factor. We find the best fitting p_M by maximizing the posterior probability distribution $pdf(p_M \mid D)$, which is, according to Bayes' theorem, proportional to the likelihood $\mathscr{L}(D \mid p_M)$ times a prior $p(p_M)$. We assume flat priors in both p_Φ and

$$p_{\rm DF} := \{\ln h_R, \ln \sigma_{R,0}, \ln \sigma_{z,0}, \ln h_{\sigma,R}, \ln h_{\sigma,z}\} \tag{9}$$

(see Section 2.3) throughout this work. Then *pdf* and likelihood are proportional to each other and differ only in units.

The normalisation in Equation 8 is a measure for the total number of tracers inside the survey volume,

$$M_{\rm tot} \equiv \int \mathrm{d}^3x\ \rho_{\rm DF}(R, |z| \mid p_M) \cdot \mathrm{SF}(\boldsymbol{x}). \tag{10}$$

In the case of an axisymmetric Galaxy model and $\mathrm{SF}(\boldsymbol{x}) = 1$ within the observation volume (as in most tests in this work), the normalisation is essentially a two-dimensional integral in the R-z plane over ρ_{DF} with finite integration limits. We evaluate the integrals using Gauss-Legendre quadratures of order 40. The integral over the azimuthal direction can be solved analytically.

It turns out that a sufficiently accurate evaluation of the likelihood is computationally expensive, even for only one set of model parameters. This expense is dominated by the number of action calculations required, which in turn depends on N_* and the numerical accuracy of the tracer density interpolation grid with $N_x^2 + N_v^3$ grid points in Equation 7 needed for the likelihood normalization in Equation 10. The accuracy of the normalization has to be chosen high enough, such that the resulting numerical error

$$\delta_{M_{\rm tot}} \equiv \frac{M_{\rm tot,approx}(N_x, N_v, n_\sigma) - M_{\rm tot}}{M_{\rm tot}} \tag{11}$$

does not dominate the numerically calculated log-likelihood, i.e.,

$$\begin{aligned} &\log \mathscr{L}_{\rm approx}(D \mid p_M) \\ &= \sum_i^{N_*} \log \mathrm{DF}(\boldsymbol{J_i} \mid p_M) - N_* \log(M_{\rm tot}) \\ &-N_* \log(1 + \delta_{M_{\rm tot}}), \end{aligned} \tag{12}$$

with

$$\log(1 + \delta_{M_{\rm tot}}) \leq \frac{1}{N_*}. \tag{13}$$

Otherwise numerical inaccuracies could lead to systematic biases in the potential and DF recovery. For data

Page: 6

Author: rix Subject: Cross-Out Date: 23.10.15 20:31:57

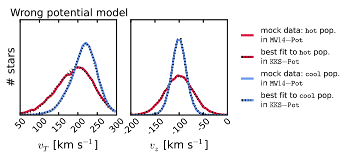

Figure 20. Comparison of the distribution of mock data v_T and v_z created in the `MW14-Pot` potential and with two different stellar populations (see Test 9 in Table 3 for all mock data model parameters), and the best fit distribution recovered by fitting the family of `KKS-Pot` potentials to the data. The best fit potentials are shown in Figure 18 and the corresponding best fit qDF parameters in Figure 19. The data is very well recovered, even though the fitted potential family did not incorporate the true potential.

The results for the potential are shown in Figure 18. We find that the potential recovered by *RoadMapping* is in good agreement with the true potential. Especially the force contours, to which the orbits are sensitive, and the rotation curve are very tightly constrained and reproduce the true potential even outside of the observed volume of the mock tracers.

Overplotted in Figure 18 is also the `KKS-Pot` with the parameters from Table 1, which were fixed based on a (by-eye) fit *directly* to the force field (within $r_{\max} = 4$ kpc from the Sun) and rotation curve of the `MW14-Pot`. The potential found with the *RoadMapping* analysis is an equally good or even slightly better fit. This demonstrates that *RoadMapping* fitting infers a potential that in its actual properties resembles the input potential for the mock data as closely as possible, given the differences in functional forms.

The density contours are less tightly constrained than the forces, but we still capture the essentials. Overall the best fit disk is less dense in the midplane than the true disk, because the generation of very flattened components like exponential disks with Stäckel potentials is very difficult.

Figure 19 compares the true qDF parameters with the best fit qDF parameters belonging to the best fit potentials from Figure 18. While we recover h_R, $\sigma_{R,0}$ and $h_{\sigma,R}$ within the errors, we misjudge the parameters of the vertical velocity dispersion ($\sigma_{0,z}$ and $h_{\sigma,z}$), even though the actual mock data distribution is well reproduced. This discrepancy could be connected to the `KKS-Pot` not being able to reproduce the flattness of the disk. Also, σ_z and σ_R in Equations 5-6 are scaling profiles for the qDF (cf. BR13) and how close they are to the actual velocity profile depends on the choice of potential.

3.7. *The influence of the stellar population's kinematic temperature*

Overall, we found that it does not make a big and generic difference if we use hot or cool stellar populations in our modelling. Only to a certain extent the kinematic temperature plays a role for how precise and reliable model parameters can be recovered.

While different populations constrain different parameters in different survey volumes with different precision, there is no easy rule of thumb, what combination would give the best results (see Figure 8).

There are two exceptions:

First, the circular velocity at the Sun, $v_{\rm circ}(R_\odot)$, is always best recovered with cooler populations (see Figures 12, 14, 16, 17 and 18), because more stars are on near-circular orbits (see Figure 2). As cooler populations probe the rotation curve better, which in turn probes the gravitational potential, the potential recovery using cool stellar populations is less sensitive to misjudgements of (spatial) selection functions (see Figures 10 and 11).

Second, hotter populations seem to be less sensitive to misjudgements of proper motion measurement uncertainties (see Figure 14) and pollution with stars from a cooler population (see Figures 16 and 17), because of their higher intrinsic velocity dispersion (see Figure 3).

In addition we found indications in Figure 18, that different regions within the Galaxy are probed best by populations of different kinematic temperature: The `hot` stellar population, with more stars reaching to high $|z|$ and a shorter tracer scale length, constrained force and density contours in the halo better—especially at smaller radii; the `cool` population, with more stars in the plane and longer tracer scale length, gave tighter force and density constraints in the outer regions of the halo and recovered the disk more reliably.

4. SUMMARY AND DISCUSSION

Recently, implementations of action DF-based modelling of 6D data in the Galactic disk have been put forth, in part to lay the ground-work fo Gaia (BR13; McMillan & Binney 2013; Piffl et al. 2014; Sanders & Binney 2015).

We present *RoadMapping*, an improved implementation of the dynamical modelling machinery of BR13, to recover the MW's gravitational potential by fitting an orbit DF to stellar populations within the Galactic disk. In this work we investigated the capabilities, strengths and weaknesses of *RoadMapping* by testing its robustness against the breakdown of some of its assumptions—for well-defined, isolated test cases using mock data. Overall the method works very well and is ~~reliable~~ even when there are small deviations of the model assumptions from the "real world" Galaxy.

RoadMapping applies a full likelihood analysis and is statistically well-behaved. It goes beyond BR13 by allowing for a straightforward and flexible implementation of different model families for potential and DF. It also accounts for selection effects by using full 3D selection functions (given some symmetries).

Computational speed: Large data sets in the age of Gaia require increasingly accurate likelihood evaluations and flexible models. To be able to deal with these computational demands, we sped up the *RoadMapping* code by combining a nested-grid approach with MCMC and by faster action calculation using the Stäckel (Binney 2012a) interpolation grid by Bovy (2015). However, application of *RoadMapping* to millions of stars will still be a task for supercomputers and calls for even more improvements and speed-up in the fitting machinery.

Properties of the data set: We could show that *RoadMapping* can provide potential and DF parameter estimates that are very accurate (i.e., unbiased) and precise in the limit of large datasets, as long as the modelling assumptions are fulfilled.

Page: 15

Author: rix Subject: Inserted Text Date: 23.10.15 20:34:01
r

Author: rix Subject: Inserted Text Date: 23.10.15 20:34:38
robust

In case the data set is affected by measurement uncertainties, the potential can still be recovered to high precision, as long as these uncertainties are perfectly known and distance uncertainties are negligible. For large proper motion uncertainties, e.g., $\delta\mu \sim 5$ mas yr^{-1}, the formal errors on the parameters are only twice as large as in the case of no measurement uncertainties. However, properly accounting for measurement uncertainties is computationally expensive.

For the results to be accurate within 2 ~~sigma~~ (for 10,000 stars), we need to know to within 10% both the true stellar distances (at $r_{\max} \leq 3$ kpc and $\delta\mu \lesssim 2$ mas yr^{-1}) and the true proper motion uncertainties (with $\delta\mu \lesssim 3$ mas yr^{-1}).

We also found that the location of the survey volume within the Galaxy matters little. At given sample size a larger survey volume with large coverage in both radial and vertical direction will give the tightest constraints on the model parameters.

Surprisingly (cf. Rix & Bovy 2013), the potential recovery with *RoadMapping* seems to be very robust against misjudgements of the spatial data SF. We speculate that this is because missing stars in the data set do not affect the measured rotation curve, which contains information about the potential.

We found indications that populations of different scale lengths and temperature probe different regions of the Galaxy best. This supports the approach by BR13, who constrained for each MAP the surface mass density only at one single best radius to account for missing flexibility in their potential model. While cooler populations probe the Galaxy rotation curve better and hotter populations are less sensitive to pollution, overall stellar populations of different kinematic temperature seem to be equally well-suited for dynamical modelling.

Deviations from the DF assumption: *RoadMapping* assumes that stellar sub-populations can be described by simple DFs. We investigated how much the modelling would be affected if the assumed family of DFs would differ from the stars' true DF.

In Example 1 in Section 3.5 we considered true stellar DFs being (i) hot with more stars with low velocities and less stars at small radii than assumed (reddish data sets in Figure 15 and 16), or (ii) cool with broader velocity dispersion wings and less stars at large radii than assumed (bluish data sets). We find that case (i) would give more reliable results for the potential parameter recovery.

Binning of stars into MAPs in [α/Fe] and [Fe/H], as done by BR13, could introduce systematic errors due to abundance uncertainties or too large bin sizes—always assuming MAPs follow simple DF families (e.g., the qDF). In Example 2 in Section 3.5 we found that, in the case of 20,000 stars per bin, differences of $\lesssim 20\%$ in the qDF parameters of two neighbouring bins can still give quite good constraints on the potential parameters.

The relative differences in the qDF parameters $\sigma_{R,0}$ and $\sigma_{z,0}$ of neighbouring MAPs in Figure 6 of BR13 (which have bin sizes of [Fe/H] = 0.1 dex and Δ[α/Fe] = 0.05 dex) are indeed smaller than 20%. Figure 16 and 17 suggest that especially the tracer scale length h_R needs to be recovered to get the potential scale length right. For this parameter however the bin sizes in Figure 6 of BR13 might not yet be small enough to ensure no more than 20% of difference in neighbouring h_R.

The qDF is a specific example for a simple DF for stellar sub-populations which we used in this paper. But it is not essential for the *RoadMapping* approach. Future studies might apply slight alternatives or completely different DFs to data.

Gravitational potential beyond the parametrized functions considered: In addition to the DF, *RoadMapping* also assumes a parametric model for the gravitational potential. We test how using a potential of Stäckel form (`KKS-Pot`, Batsleer & Dejonghe 1994) affects the *RoadMapping* analysis of mock data from a different potential family with halo, bulge and exponential disk. The potential recovery is quite successful: We properly reproduce the mock data distribution in configuration space; and the best fit potential is—within the limits of the model—as close as it gets to the true potential, even outside of the observation volume of the stellar tracers.

For as many as 20,000 stars constraints become already so tight that it should presumably be possible to distinguish between different parametric MW potential models (e.g., the `MW13-Pot` used by BR13 and the `KKS-Pot`).

BR13 fitted a MW-like model potential and calculated actions using the Stäckel approximation (Binney 2012a); in this case study we directly fitted a Stäckel potential to the data, with exact actions in the model potential. The latter is computationally much less expensive due to the simple analytic form of the potential. It would also allow flexibility by expressing the MW potential as a superposition of many more simple Kuzmin-Kutuzov Stäckel components (Famaey & Dejonghe (2003) used for example 3 components). The former approach by BR13 however allows to parametrize the potential with intuitive and physically motivated building blocks (exponential disks, power-law dark matter halo etc.). While both approaches are formally similar, it remains to decide which is better.

Different modelling approaches using action-based DFs: BR13 have focussed on MAPs for a number of reasons: First, they seem to permit simple DFs (Bovy et al. 2012b,c,d), i.e., approximately qDFs (Ting et al. 2013). Second, all stars must orbit in the same potential. While each MAP can yield different DF parameters, it will also provide a (statistically) independent estimate of the potential. This allows for a valuable cross-checking reference. In some sense, the *RoadMapping* approach focusses on constraining the potential, treating the DF parameters as nuisance parameters. That we were able to show in this work that *RoadMapping* results are quite robust to the form of the DF not being entirely correct motivates this approach further.

~~The main drawback is that—~~for reasons of galaxy and chemical evolution—the DF properties are astrophysically linked between different MAPs. Ultimately, the goal is to do a consistent chemodynamical model that simultaneously fits the potential and DF($\boldsymbol{J}$, [X/H]) (where [X/Fe] denotes the whole abundance space) with a full likelihood analysis. This has not yet been attempted

Page: 16

Author: rix Subject: Inserted Text Date: 23.10.15 20:36:39
substantive

Author: rix Subject: Inserted Text Date: 23.10.15 20:36:52
\sigma

Author: rix Subject: Inserted Text Date: 23.10.15 20:37:51

Author: rix Subject: Inserted Text Date: 23.10.15 20:37:46

Author: rix Subject: Cross-Out Date: 23.10.15 20:38:27

Author: rix Subject: Inserted Text Date: 23.10.15 20:38:05
(Sanders et al 2015)

Author: rix Subject: Inserted Text Date: 23.10.15 20:39:07
In its current implementation, RoadMAP treats all MAPs as independent and dose not exploit such correlations.