Draft version September 21, 2015
Preprint typeset using LaTeX style emulateapj v. 12/16/11

ACTION-BASED DYNAMICAL MODELLING OF THE MILKY WAY DISK WITH *ROADMAPPING* AND OUR IMPERFECT KNOWLEDGE OF THE "REAL WORLD"

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

Draft version September 21, 2015

ABSTRACT

We present *RoadMapping*, a dynamical modelling machinery that aims to recover the Milky Way's (MW) gravitational potential and the orbit distribution of stellar populations in the Galactic disk. *RoadMapping* is a full likelihood analysis that models the observed positions and velocities of stars with an equilibrium, three-integral distribution function (DF) in an axisymmetric potential. In preparation for the application to the large data sets of modern surveys like Gaia, we create and analyze a large suite of mock data sets and develop qualitative "rules of thumb" for which characteristics and limitations of data, model and machinery affect constraints on the potential and DF most. We find that, while the precision of the recovery increases with the number of stars, the numerical accuracy of the likelihood normalisation becomes increasingly important and dominates the computational efforts. The modelling has to account for the survey's selection function, but *RoadMapping* seems to be very robust against small misjudgments of the data completeness. Large radial and vertical coverage of the survey volume gives in general the tightest constraints. But no observation volume of special shape or position and stellar population should be clearly preferred, as there seem to be no stars that are on manifestly more diagnostic orbits. We propose a simple approximation to include measurement errors at comparably low computational cost that works well if the distance error is $\lesssim 10\%$. The model parameter recovery is also still possible, if the proper motion errors are known to within 10% and are $\lesssim 2$ mas yr^{-1}. We also investigate how small deviations of the stars' distribution from the assumed DF influence the modelling: An over-abundance of high velocity stars affects the potential recovery more strongly than an under-estimation of the DF's low-velocity domain. Selecting stellar populations according to mono-abundance bins of finite size can give reliable modelling results, as long as the DF parameters of two neighbouring bins do not vary more than 20% [TO DO: CKECK]. As the modelling has to assume a parametric form for the gravitational potential, deviations from the true potential have to be expected. We find, that in the axisymmetric case we can still hope to find a potential that is indeed a reliable best fit within the limitations of the assumed potential. Overall *RoadMapping* works as a reliable and unbiased estimator, and is robust against small deviations between model and the "real world".

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

1. DYNAMICAL MODELLING

[TO DO: HW: In this section you have to indicate somehow, where you recapitulate BR13 and what is added new. "as in BR13", "beyond BR13"] TO DO: Comment from Hans-Walter: Basics of the method (binney, BR13) can be dramatically shortened. No need to sing the praises of DF's and actions.

In this section we summarize the basic elements of *RoadMapping*, the dynamical modelling machinery presented in this work, which in many respects follows Bovy & Rix (2013).

1.1. *Coordinate System*

Our modelling takes place in the Galactocentric rest-frame with cylindrical coordinates $\boldsymbol{x} \equiv (R, \phi, z)$ and corresponding velocity components $\boldsymbol{v} \equiv (v_R, v_\phi, v_z)$. If the stellar phase-space data is given in observed 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 $\tilde{\boldsymbol{v}} \equiv (\mu_{\mathrm{RA}}, \mu_{\mathrm{DEC}}, v_{\mathrm{los}})$ as proper motions $\boldsymbol{\mu} = (\mu_{\mathrm{RA}}, \mu_{\mathrm{DEC}})$ [TO DO: cos somwhere???] in both RA and DEC direction and line-of-sight velocity v_{los}, the data $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{v}})$ has to be converted first into the Galactocentric rest-frame coordinates $(\boldsymbol{x}, \boldsymbol{v})$ using the sun's position and velocity. We assume for the sun

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

1.2. *Actions and Potential Models*

Orbits in axisymmetric potentials are best described and fully specified by the three actions $\boldsymbol{J} \equiv (J_R, J_z, J_\phi = L_z)$, defined as

$$J_i = \frac{1}{2\pi} \oint_{\mathrm{orbit}} p_i \, \mathrm{d}x_i, \tag{1}$$

and which depend on the potential via the connection between position x_i and momentum p_i along the

[1] Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany
[2] Correspondence should be addressed to trick@mpia.de.
[3] University of Toronto [TO DO: What is Jo's current address???]

Table 1
Gravitational potentials of the reference galaxies used troughout this work and the respective ways to calculate actions in these potentials. All four potentials are axisymmetric. The potential parameters are fixed for the mock data creation at the values given in this table. In the subsequent analyses we aim to recover these potential parameters again. The parameters of `MW13-Pot` and `KKS-Pot` were chosen to resemble the `MW14-Pot` (see Figure 1). We use $v_{\rm circ}(R_\odot) = 230$ km s^{-1} for all potentials in this work.

name	potential type	potential parameters p_Φ		action calculation
`Iso-Pot`	isochrone potential[a] (Henon 1959)	b	0.9 kpc	***analytical and exact*** (Binney & Tremaine 2008, §3.5.2)
`KKS-Pot`	2-component Kuzmin-Kutuzov-Stäckel potential[b] (disk + halo) (Batsleer & Dejonghe 1994)	Δ	0.3	***exact*** using *Stäckel Fudge* (Binney 2012) and interpolation on action grid[e] (Bovy 2015)
		$\left(\frac{a}{c}\right)_{\rm Disk}$	20	
		$\left(\frac{a}{c}\right)_{\rm Halo}$	1.07	
		k	0.28	
`MW13-Pot`	MW-like potential[c] with Hernquist bulge, spherical power-law halo, 2 exponential disks (stars + gas) (Bovy & Rix 2013)	R_d	3 kpc	***approximate*** (same as `KKS-Pot`)
		z_h	0.4 kpc	
		f_h	0.5	
		$\frac{{\rm d}\ln(v_{\rm circ}(R_\odot))}{{\rm d}\ln(R)}$	0	
`MW14-Pot`	MW-like potential[d] with cut-off power-law bulge, Miyamoto-Nagai stellar disk, NFW halo (Bovy 2015)			***approximate*** (same as `KKS-Pot`)

[a] The isochrone potential `Iso-Pot` has one free parameter, the scale length b.
[b] The coordinate system of each of the two Stäckel-potential components of the `KKS-Pot` is $\frac{R^2}{\tau_{i,p}+\alpha_p} + \frac{z^2}{\tau_{i,p}+\gamma_p} = 1$ with $p \in \{\text{Disk}, /\text{Halo}\}$ and $\tau_{i,p} \in \{\lambda_p, \nu_p\}$. Both components have the same focal distance $\Delta \equiv \sqrt{\gamma_p - \alpha_p}$, to make sure that the superposition of the two components itself is still a Stäckel potential. The axis ratio of the coordinate surfaces $\left(\frac{a}{c}\right)_p := \sqrt{\frac{\alpha_p}{\gamma_p}}$ describes the flatness of the corresponding Stäckel component. The parameter k describes the relative contribution of the disk mass to the total mass.
[c] The free parameters of the `MW13-Pot` are stellar disk scale length R_d and height z_d, as well as the relative halo contribution to $v^2_{\rm circ}(R_\odot)$, f_h, and the slope of the rotation curve, $\frac{{\rm d}\ln(v_{\rm circ}(R_\odot))}{{\rm d}\ln(R)}$.
[d] The `MWPotential2014` by Bovy (2015) (see their Table 1) has a circular velocity at the Sun of $v_{\rm circ}(R_\odot) = 220$ km s^{-1}. In this work we use however $v_{\rm circ}(R_\odot) = 230$ km s^{-1} for all potentials.
[e] We use a finely spaced action interpolation grid with $R_{\rm max} = 10$ [TO DO: What's that??? units???] and 50 grid points in E and ψ [TO DO: Find out what's that???], and 60 grid points in L_z.

Table 2
Reference distribution-function parameters for the qDF in Equations (2)-(7). These qDFs describe the phase-space distribution of stellar MAPs for which mock data is created and analysed throughout this work for testing purposes. The parameters of the `cooler` & `colder` (`hotter` & `warmer`) MAPs were chosen to have the same $\sigma_{R,0}/\sigma_{z,0}$ ratio as the `hot` (`cool`) MAP. The `colder` and `warmer` MAPs have a free parameter X that governs how much colder/warmer they are then the reference `hot` and `cool` qDFs. Hotter populations have shorter tracer scale lengths (Bovy et al. 2012d) and the velocity dispersion scale lengths were fixed according to Bovy et al. (2012c).

name	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`	2 +50%	55-50%	66-50%	8	7
`hotter`	3.5-50%	42+50%	32+50%	8	7
`colder`	2 +X%	55-X%	66-X%	8	7
`warmer`	3.5-X%	42+X%	32+X%	8	7

orbit. Actions have a clear physical meaning: They quantify the amount of oscillation in each coordinate direction of the full orbit [TO DO: REF: HW suggested Binney & Tremaine (2008), but I can't find a corresponding statement in the book]. 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, §3.5.1).

Even though actions are excellent orbit labels and arguments for stellar distribution functions, their computation is typically very expensive and depends on the choice of potential in which the star moves. The spherical isochrone potential (Henon 1959) is the only [TO DO: Jo suggested "most general Galactic" instead of "only", but the isochrone is actually not Galactic... Ask him.] potential for which Equation 1 takes an analytic form (Binney & Tremaine 2008, §3.5.2). For 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 along the whole (often not periodic) orbit. 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).

For the gravitational potential in our modelling we assume a family of parametrized potential models with a fixed number of free parameters. We use different kinds of potentials: The Milky Way like potential from Bovy & Rix (2013) (`MW13-Pot`) with bulge, disk and halo; 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; and the 2-component Kuzmin-Kutuzov Stäckel potential (Batsleer & Dejonghe 1994; `KKS-Pot`), which displays a disk and halo structure and also provides exact actions. Table 1 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.

1.3. *Stellar Distribution Functions*

Throughout, we assume that the orbits of each MAP can be described by a single qDF of the form given by Binney & McMillan (2011). This is motivated by the findings of Bovy et al. (2012b,c,d) and Ting et al. (2013) about the simple phase-space structure of MAPs, and following Bovy & Rix (2013) and their successful application. This qDF has the form

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

with

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

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

Here $R_g \equiv R_g(L_z)$ and $\Omega \equiv \Omega(L_z)$ are the (guiding-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, §3.2.3) [TO DO: ask someone, if I'm messing up different definitions of κ]. 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 small value ($L_0 = 10 \times R_\odot/8 \times v_\mathrm{circ}(R_\odot)/220$ [TO DO: Jo said, galpy default is 10 km/s kpc. But I got the value actually from the code...]).
To match the observed properties of MAPs (see Bovy et al. 2012b,c,d), we chose the functional forms

$$n(R_g \mid p_\mathrm{DF}) \propto \exp\left(-\frac{R_g}{h_R}\right) \tag{5}$$

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

$$\sigma_z(R_g \mid p_\mathrm{DF}) = \sigma_{z,0} \times \exp\left(-\frac{R_g - R_\odot}{h_{\sigma,z}}\right), \tag{7}$$

which indirectly set the stellar number density and radial and vertical velocity dispersion profiles. 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. Throughout this work we use for illustration purposes a few example stellar populations, each following a single qDF, whose parameters are given in in Table 2. Most tests use the `hot` and `cool` qDFs from Table 2, which correspond to kinematically hot and cool populations, respectively.

One crucial point in our dynamical modelling technique (§1.5), as well as in creating mock data (§**??**), 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_v-th order Gauss-Legendre quadrature for each integral:

$$\begin{aligned}&\rho_\mathrm{DF}(R, |z| \mid p_\Phi, p_\mathrm{DF})\\ &= \int_{-\infty}^{\infty} \mathrm{qDF}(\boldsymbol{J}[R, z, \boldsymbol{v} \mid p_\Phi] \mid p_\mathrm{DF})\, \mathrm{d}^3\boldsymbol{v}\end{aligned} \tag{8}$$

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

where $\sigma_R(R \mid p_\mathrm{DF})$ and $\sigma_z(R \mid p_\mathrm{DF})$ are given by Equations 6 and 7 and the integration ranges are motivated

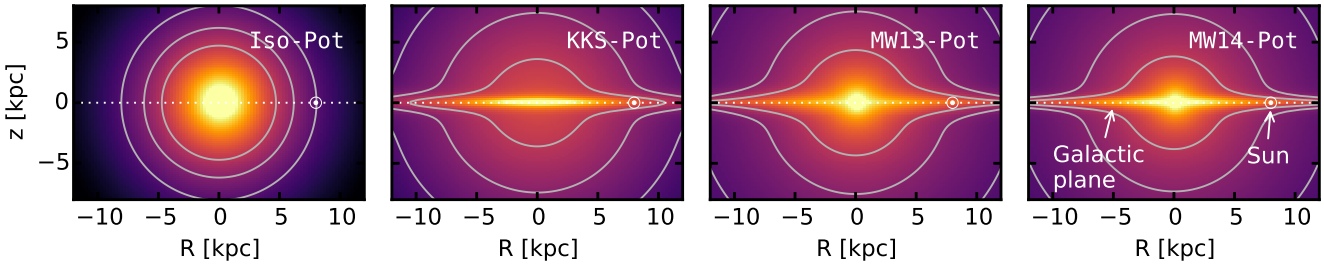


Figure 1. Density distribution of the four reference galaxy potentials in Table 1, for illustration purposes. These potentials are used throughout this work for mock data creation and potential recovery. [TO DO: Potential and/or population names in typewriter]

by Figure **??**. The integration range $[0, 1.5v_{\rm circ}(R_\odot)]$ 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 $p_{\rm DF}$ we explicitly calculate the density on $N_x \times N_x$ 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 $|z| \leq 0$, because the model is symmetric in z by construction). The total number of actions that need to be calculated to set up the density interpolation grid is $N_x^2 \cdot N_v^3$. §1.5 and Figure 2 show the importance of choosing N_x, N_v and n_σ sufficiently large in order to get the density with an acceptable numerical accuracy [TO DO: Jo thinks that this statement is difficult to understand here, because you have not yet talked about the normalization].

1.4. *Selection Functions*

Any survey's selection function can be understood as defining an effective sample subvolume in the space of observables: e.g. position on the plane of the sky (the survey area), 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).
We simply use spatial selection functions, which describe the probability to observe a star at $\boldsymbol{x}$,

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

For the observed volume we use simple geometrical shapes. Either 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 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. Here $0 \leq \text{completeness}(\boldsymbol{x}) \leq 1$ everywhere inside the observed volume, so it can be understood as a position-dependent detection probability. Unless explicitly stated otherwise, we simplify to $\text{completeness}(\boldsymbol{x}) = 1$.

1.5. *Data Likelihood*

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

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

The model that we fit is specified by a number of fixed and free parameters,

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

For the qDF parameters (see Section 1.3) we assume a prior that is flat in

$$p_{\rm DF} := \{\ln h_R, \ln \sigma_{R,0}, \ln \sigma_{z,0}, \ln h_{\sigma,R}, \ln h_{\sigma,z}\}. \quad (10)$$

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

The likelihood of the data given the model is

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

where N is the number of stars in the data set D, and 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 treat it as unimportant proportionality factor in the likelihood calculation. We find the best set of model parameters by maximizing the posterior probability distribution $pdf(p_M \mid D)$, which is according to Bayes' theorem proportional the likelihood $\mathscr{L}(D \mid p_M)$ times the prior. We assume flat priors in both p_Φ and $p_{\rm DF}$ (see Equation 10) through out this work, then *pdf* and likelihood can and will be used interchangeably for the remainder of the work.

The normalisation in Equation 11 is a measure for the total number of tracers inside the survey volume,

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

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} in Equation 9 over the differential survey volume, i.e. $\frac{\partial M_{\mathrm{tot}}}{\partial \phi}(R,z) = \int \mathrm{d}R\ \mathrm{d}z\ \rho_{\mathrm{DF}} \times \frac{\partial V}{\partial \phi}$ [TO DO: missing factor of R???]. We perform this integral as a Gauss Legendre quadrature of order 40 in each R and z direction. The angular integral, i.e. $M_{\mathrm{tot}} = \int R\,\mathrm{d}\phi\ \frac{\partial M_{\mathrm{tot}}}{\partial \phi}$, can be solved analytically.
It turns out that the 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 the number of stars in the sample and the numerical accuracy of the integrals in Equation 9 needed for the normalisation, which requires $N_x^2 \times N_v^3$ action calculations. The accuracy has to be chosen high enough, such that a resulting numerical error

$$\delta_{M_{tot}} \equiv \frac{M_{\mathrm{tot,approx}}(N_x, N_v, N_\sigma) - M_{\mathrm{tot}}}{M_{\mathrm{tot}}} \tag{13}$$

[TO DO: make sure every Mtottrue is replaced by Mtot] does not dominate the likelihood, i.e.

$$\begin{aligned}&\log \mathscr{L}(p_M \mid D)\\ &= \sum_i^N \log \mathrm{qDF}(\boldsymbol{J_i} \mid p_M) - 3N \log(r_o v_o)\\ &-N \log(M_{\mathrm{tot}}) - N \log(1+\delta_{M_{tot}}),\end{aligned} \tag{14}$$

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 log-likelihoods differ, to be clearly distinguishable, by 1. Otherwise numerical inaccuracies could lead to systematic biases in the potential and DF fitting. For data sets as large as $N = 20{,}000$ stars, which in the age of Gaia could very well be the case [TO DO: Really???], one needs a numerical accuracy of 0.005% in the normalisation. Figure 2 demonstrates that the numerical accuracy we use in the analysis, $N_x = 16$, $N_v = 24$ and $N_{sigma} = 5$, does satisfy this requirement.

[TO DO: Comment by HW: What I am missing in this Section is any distinction of what aspects are "new" (not addressed in existing papers) and what is recapitulated to be coherent.]

If the data is affected by measurement errors, they have to be incorporated in the likelihood. We assume Gaussian errors in the observable space $\boldsymbol{y} \equiv (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{v}}) = (\mathrm{RA}, \mathrm{DEC}, (m-M), \mu_{\mathrm{RA}}, \mu_{\mathrm{DEC}}, v_{\mathrm{los}})$, i.e. the i-th star's observed $\boldsymbol{y}_i \sim N[\boldsymbol{y'}_i, \delta\boldsymbol{y}_i](\boldsymbol{y}) = N[\boldsymbol{y}, \delta\boldsymbol{y}_i](\boldsymbol{y'}_i)$ [TO DO: Talk to HW about best notation.], with $\boldsymbol{y_i}'$ being the true position and velocity of the star. Stars follow the (quasi-isothermal) distribution function ($\mathrm{DF}(\boldsymbol{y'}) \equiv \mathrm{qDF}(\boldsymbol{J}[\boldsymbol{y'} \mid p_\Phi] \mid p_{\mathrm{DF}})$ for short), convolved with the error distribution $N[0, \delta\boldsymbol{y}](\boldsymbol{y'})$ [TO DO: CHECK AGAIN]. The selection function $\mathrm{sf}(\boldsymbol{y})$ acts on the space of (error affected) observables. Then the probability of one star becomes

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

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,

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

with

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

In doing so, we ignore errors in the star's position $\boldsymbol{x}_i$ [TO DO: something is not clear to HW here] altogether. This simplifies the normalisation drastically and makes it independent of measurement errors, including the velocity errors. Distance errors however are included [TO DO: something is not clear to HW here], but only implicitly in the convolution over the stars' velocity errors in the Galactocentric rest frame. We calculate the convolution using Monte Carlo integration with N_{error} samples drawn from the full error Gaussian in observable space, $y'_{i,n}$. Figure 3 demonstrates that in the absence of distance errors the N_{error} needed for the convolution integral to converge depends as

$$N_{\mathrm{error}} \propto \delta v^2$$

on the uncertainties in the (1D) velocities.

1.6. *Measurement Errors and their Effect on the Parameter Recovery*

[TO DO: Comment from HW: This Section has three parts:
– convergence of the integral (ALREADY REMOVED)
– testing the approximation
– underestimating errors
It seems to me that the basic Section: What is the impact of the errors? Is missing. That should be the center piece, and the other three aspects should be quick summary notes, only 1-2 sentences long.] [I'll try to address this with a plot mean(SE) vs. proper motion error - also for cold population (currently running on wolf).]

In absence of distance uncertainties the error convolved likelihood given in Equation 15 is unbiased. When including distance (modulus) errors Equation 15 is just an approximation for the true likelihood. The systematic bias thus introduced in the parameter recovery gets larger with the size of the error. This is demonstrated in Figure 6.2. We find however that in case of $\delta(m-M) \lesssim 0.3$ mag (if also $\delta\mu \leq 2$ mas yr^{-1} and a maximum distance of $r_{\mathrm{max}} = 3$ kpc, see Test 6.2 in Table 3) the potential parameters can still be recovered within 2 sigma [TO DO: Make sure this is what I claim

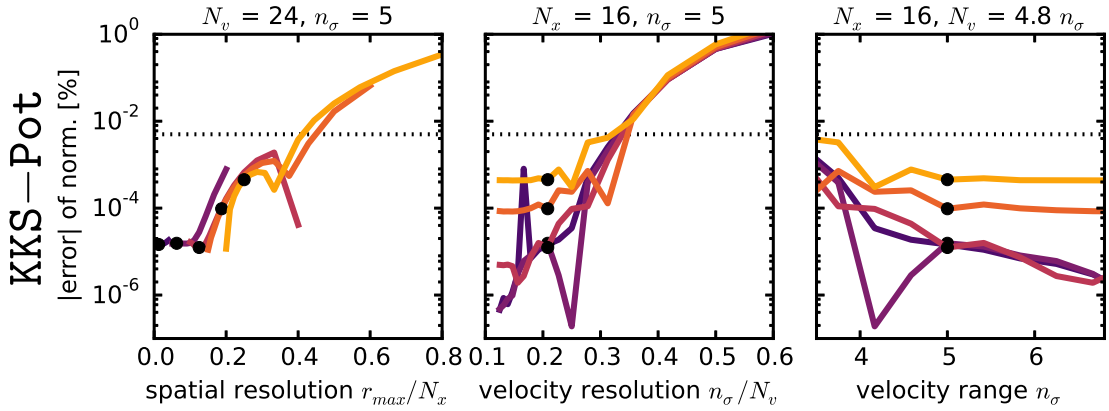


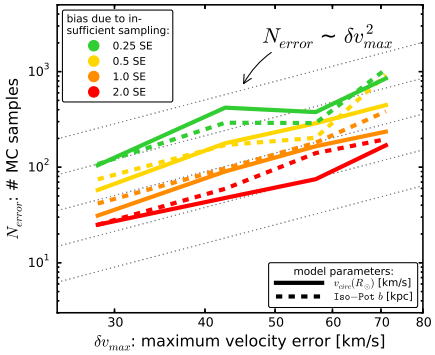
Figure 2. Relative error $\delta M_{\rm tot}$ of the likelihood normalization $M_{\rm tot}$ in Equation 13 depending on the accuracy of the grid-based density calculation in Equation 9 (and surrounding text). We show how $\delta M_{\rm tot}$ varies with the spatial resolution (first column), velocity resolution (second column) and velocity integration range (third column) for two different potentials (`KKS-Pot` in the first row and `MW13-Pot` in the second row) and five different spherical observation volumes with radius $r_{\rm max}$ (color coded according to the legend). (Test 2 in Table 3 summarizes all model parameters.) N_x is the number of spatial grid points in $R \in R_\odot{\rm kpc} \pm r_{\rm max}$ and $|z| \in [0, r_{\rm max}]$ on which the density is evaluated according to Equation 9. The spatial resolution in z is therefore $r_{\rm max}/N_x$ and $2r_{\rm rmax}/N_x$ in R. This choice is reasonable because the density is symmetric in z and varies less in R than in z, because the tracer scale length of the disk is much larger than its scale height. At each (R, z) of the grid a Gauss-Legendre integration of order N_v is performed over an integration range of $\pm n_\sigma$ times the velocity dispersion in v_R and v_z and $[0, 1.5 v_{\rm circ}(R_\odot)]$ in v_T. n_σ/N_v is therefore a proxy for the velocity resolution of the grid. (We vary N_x, N_v and n_σ separately and keep the other two fixed at the values indicated above the columns.) To arrive at the approximation $M_{\rm tot,approx}$ for $M_{\rm tot}$ in Equation 12, we perform a 40th-order Gauss-Legendre integration in each R and z direction of the interpolated density over the observed volume. We calculate the "true" normalization with high accuracy as $M_{\rm tot} \approx M_{\rm tot,approx}(N_x = 20, N_v = 56, N_\sigma = 7)$. The black dots indicate the accuracy used in our analyses: It is better than 0.002%. 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 Staeckel potential `KKS-Pot` exactly, we have to resort to an approximate action calculation for the MW-like potential `MW13-Pot` (see Section 1.2). [TO DO: Write $|\delta M_{\rm tot}|$ on y-axis] [TO DO: Remove MW13-Pot completely from this plot, caption and test table] [TO DO: Caption too long] [TO DO: Rewrite caption, text and table, I changed the plot]

in abstract and discussion.]. This corresponds to a relative distance error of $\sim 10\%$.

[TO DO: Introduce a test and plot that demonstrates how the SE depends on proper motion error. Then write this little section.] Overall the standard errors on the recovered parameters are quite small (a few percent at most for 10,000 stars), which demonstrates that, if we perfectly knew the measurement errors, we still could get very precise constraints on the potential. The constraints also get tighter the smaller the proper motion error becomes. We found that for $\delta\mu = 1$ mas yr^{-1} the precision of the recovered parameters reduce by $\sim$ half compared to $\delta\mu = 5$ mas yr^{-1}. [TO DO: Comment from HW: This seems to be a (sensible) statement about the impact of the errors. Why is it under the heading of testing an approximation?]

We found that in case we perfectly knew the measurement errors (and the distance error is negligible), the convolution of the model probability with the measurement uncertainties gives precise and accurate constraints on the model parameters - even if the error itself is quite large.

Underestimation of the proper motion error. — [TO DO: Remove Headline. Shorten this text. A lot.] Now we investigate what would happen if the quoted measurement errors, e.g. the proper motion errors, were actually smaller than the true errors. Figure 6 shows the case for two different stellar populations and an error underestimation of 10% and 50%.
Overall the parameter recovery gets worse the larger the proper motion error and the stronger the underestimation. The relation between the bias due to error misjudgment and the size of the proper motion error seems to be linear.
For the recovery of the isochrone potential scale length b the hotness of the population does not matter (see lower left panel in Figure 6). The circular velocity $v_{\rm circ}(R_\odot)$ is, as always, better measured by cooler than by hotter populations (see upper left panel in Figure 6).
We find that the recovery of the qDF parameters on the other hand is more strongly affected by the misjudgment of the velocity error for *cooler* stellar popluations. The measured velocity dispersion is the convolution of the intrinsic dispersion with the measurement errors. If the proper motion error is underestimated, the deconvolved velocity dispersion is larger than the intrinsic velocity dispersion and the relative difference is bigger for a cooler population (see upper right panel for $\sigma_{z,0}$ in Figure 6). The intrinsic velocity dispersion is also cooler at larger radii than at smaller radii, therefore the deconvolved dispersion is overestimated more strongly at large R and the velocity dispersion scale length will be overestimated as well (see lower left panel for $h_{\sigma,z}$ in Figure 6). We get analogous results for the qDF parameters $\sigma_{R,0}$ and $h_{\sigma,R}$. The recovery of the tracer density scale length h_R is not affected by the misjudgment of velocity errors.
The most important and encouraging result from Figure 6 is, that for an underestimation of 10% the bias is still $\lesssim 2$ sigma for 10,000 stars [TO DO: Check] - even for proper motion errors of almost 3 mas/yr.
[TO DO: Comment from Jo: Always use 'uncertainty' when describing how ou deal with the errors. 'Error' means the actual error (difference between observed and true).]



(a) $N_* = 10,000$

(b) $N_* = 5,000$

Figure 3. Number of Monte Carlo (MC) samples $N_{\rm error}$ needed for the numerical convolution of the model probability with the measurement uncertainties in Equation 15, given the maximum velocity error $\delta v_{\rm max}$ within the stellar sample. Unsufficent sampling introduces systematic biases in the parameter recovery as indicated in the legend. The relation found here, $N_{\rm error} \propto \delta v_{\rm max}^2$, was distilled from a set of analyses of mock data sets with different proper motion uncertainties $\delta\mu \in [2,5]$ mas yr^{-1} in the absence of distance errors (see Test 6.1 in Table 3). The proper motion error $\delta\mu$ translates to heteroscedastic [TO DO: make sure that this word is written correctly everywhere.] velocity errors according to $\delta v[\mathrm{km\ s^{-1}}] \equiv 4.74047 \cdot r[\mathrm{kpc}] \cdot \delta\mu[\mathrm{mas\ yr^{-1}}]$, with r being the distance of the star from the Sun. Stars with larger δv require more $N_{\rm error}$ for the integral over its measurement uncertainties to converge. We therefore show how the $N_{\rm error}$ needed for the potential *pdf* of the whole data set to be converged, depends on the largest velocity error $\delta v_{\rm max} \equiv \delta v(r_{\rm max})$ within the data set. We used $N_{\rm error} = 800$ and 1200 for $\delta\mu \leq 3$mas yr^{-1} and $\delta\mu > 3$mas yr^{-1}, respectively, as the reference for the converged convolution integral (see also left panels in Figure 4). [TO DO: some of the 25 MC sample analyses have to be re-done.] [TO DO: Replace right plot with new plot with $N_* = 5,000$] [TO DO: Use N_* everywhere where applicable] [TO DO: Intriduce N_* somewhere.] [TO DO: Comment from Jo: I think it is important to test, if the MC vs error plot depends on number pf stars. Maybe test it with less stars (5000), to test this quickly. Naively, I would expect a large depedence on Ndata.]

2. DISCUSSION AND SUMMARY

[TO DO: Introduce DF somewhere - use DF wherever we don't need qDF.]
[TO DO: Compare these sections with the results. Points should be made detailed in the results section and short here in the discussion. Says Hans-Walter.]

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 (Bovy & Rix 2013; McMillan & Binney 2013; Piffl et al. 2014; Sanders & Binney 2015).

We present *RoadMapping*, an improved implementation

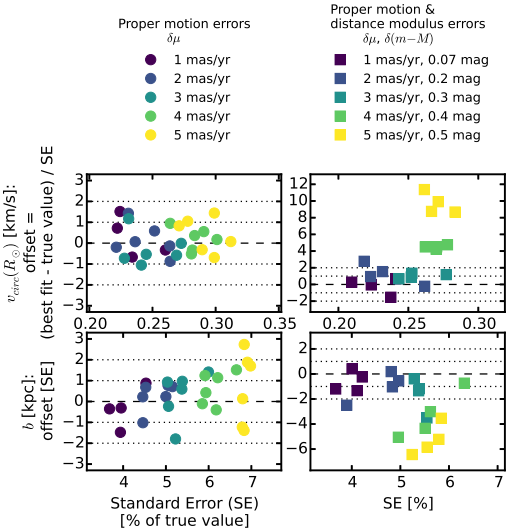


Figure 4. Potential parameter recovery using the approximation for the model probability convolved with measurement uncertainties in Equation 15. We show *pdf* offset and relative width (i.e., standard error SE) for the potential parameters derived from mock data sets, which were created according to Test 6.2 in Table 3). The data sets in the left panels have only uncertainties in line-of-sight velocity and proper motions, while the data sets in the right panels also have distance (modulus) uncertainties, as indicated in the legends in the first row. For data sets with proper motion error errors $\delta(m-M) \leq 3$ mas yr^{-1} Equation 15 was evaluated with $N_{\rm error} = 800$, for $\delta(m-M) > 3$ mas yr^{-1} we used $N_{\rm error} = 1200$. In absence of distance uncertainties Equation 15 gives unbiased results. For $\delta(m-M) \geq 3$mas yr^{-1} (which corresponds in this test to $\delta v_{\rm max} \gtrsim 43$ km s^{-1}, see Equation **??**) however biases of several sigma are introduced as Equation 15 is only an approximation for the true likelihood in this case.

Figure 5. [TO DO: This should be a figure that plots precision (SE) vs. proper motion error for a hot and a cool population (for no distance error). This is to demonstrate the effect of measurement errors in general. Currently running on cluster....]

of the dynamical modelling machinery of Bovy & Rix (2013), to recover the potential and orbit distribution function of stellar MAPs 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

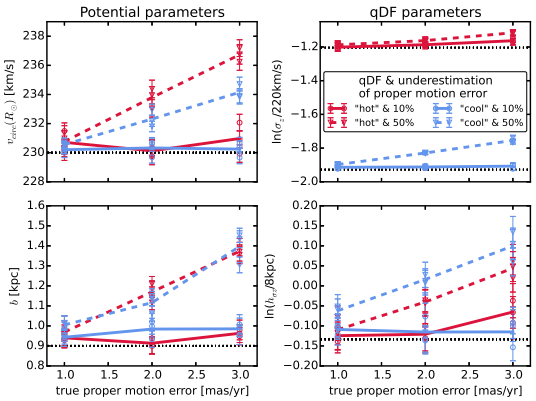


Figure 6. Effect of a systematic underestimation of proper motion errors in the recovery of the model parameters. The true model parameters used to create the mock data are summarized as Test 6.3 in Table 3, four of them are given on the y-axes and the true values are indicated as black dashed lines. The velocities of the mock data were perturbed according to Gaussian errors in the RA and DEC proper motions as indicated on the x-axis. The circles and triangles are the best fit parameters of several mock data sets assuming the proper motion uncertainty, with which the model probability was convolved, was underestimated in the analysis by 10% or 50%, respectively. The error bars correspond to 1 sigma confidence. The lines connect the mean of each two data realisations and are just to guide the eye. [TO DO: rename $h_{\sigma z}$ to $h_{\sigma,z}$, σ_z to $\sigma_{z,0}$] [TO DO: Potential and/or population names in typewriter font]

statistically well-behaved. It allows for a straightforward implementation of different potential model families and a flexible number of free fit parameters in potential and DF. It also accounts for selection effects by using full 3D selection functions (given some symmetries). *RoadMapping* is an asymptotically normal, unbiased estimator and the precision of parameter recovery increases by $1/\sqrt{N}$ with the number of stars.

Computational speed: Large data sets in the age of Gaia require more, and more accurate, likelihood evaluations for more flexible models. To be able to deal with these increased computational demands and explore larger parameter spaces, we sped up the code by combining a nested grid approach with MCMC and by faster action calculation using the Stäckel (Binney 2012) interpolation grid by Bovy (2015). However, application of *RoadMapping* to millions of stars simultaneously with acceptable accuracy 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 the limit of large datasets, as long as the modelling assumptions are fulfilled. We also found that the location of the survey volume matters little. At given sample size, the volume over which the data are sampled also matters little, if the modelling assumptions are fulfilled. are fulfilled. Concerning the *shape* of the survey volume, a large radial *and* vertical coverage is best, because in the axisymmetric regime the azimuthal coverage does not matter.
Stellar populations of different scale length and temperature probe different regions of the Galaxy (Bovy & Rix 2013). But there is no easy rule of thumb for which survey volume and stellar population which potential and DF parameter is constrained best.
Surprisingly, (cf. Rix & Bovy 2013) *RoadMapping* seems to be very robust against misjudgments in the selection function of the data. We speculate that this is because missing stars in the data set do not affect the connection between a star's velocity and position, which is given by the potential. Much of the information about the potential profile is stored in the rotation curve, but we find that even when when we do not include measurements of tangential velocities in the analysis, small misjudgments of the incompleteness do not affect the potential recovery.

[TO DO: Comment from HW: Author: rix Subject: This paragraph shouldbe 1 or 2 sentences, following the first paragraph on "Sample/Data Properties". This – at the moments –reads to bequite confusing. I don't quite get whatthe "upshot" is; thereistechnical detai on N_{error} [enought to say it's expensive]; and, as noted earlier; I don't understand why theerror convolution for a nearbydata point needs to know about $\delta v_{\rm max}$]
Properly convolving the likelihood with measurement errors is computational very expensive. By ignoring positional errors and only including distance errors as part of the velocity error, we can drastically reduce the computational costs. For stars within 3 kpc from the sun this approximation works well for distance errors of $\sim 10\%$ or smaller. The number of MC samples needed for the error convolution using MC integration scales by $N_{\rm error} \propto (\delta v_{\rm max})^2$ with the maximum velocity error at the edge of the sample. If we did not know the true size of the proper motion measurement errors perfectly, we can only reproduce the true model parameters to within $\lesssim 2$ sigma [TO DO: Check???] as long as we do not underestimate it by more than 10% and for proper motion errors $\lesssim 2$ mas yr^{-1}.

Deviations from the qDF Assumption: Our modelling is founded on the assumption, that we can identify *a priori* sub-components of the Galactic disk that follow a qDF (e.g., by considering MAPs). There are two reasons why any chosen sub-sample of stars (here a MAP) may not be well described by any qDF. Either, because nature is more complex, or because even if perfect MAPs would be well described by qDFs finite abundance errors would mix MAPs. We have considered both cases. [TO DO: Comment from HW: it feels to me that this is the 3rd time you said this.It's OK to say, but in1 line at most.]
In Example 1 in §**??** we investigated how well we can recover the potential, if this assumption was not perfectly satisfied, i.e., the MAP's true DF does not perfectly follow a qDF. We considered two cases: a) a hot DF, that has less stars at small radii and more stars with low velocities than predicted by the qDF (reddish data sets in Figure **??**), or b) a cool DF that has broader velocity dispersion wings and less stars at large radii than predicted by the qDF (bluish data sets). We find that case a) would give more reliable results for the potential parameter recovery, but in both cases biases are small if the contamination is less than [TO DO: CHECK].
If we assumed that the distribution of stars from one

MAP is caused by radial migration away from the initial location of star formation, it would more likely that the qDF overestimates the true number of stars at smaller radii than underestimating it at larger radii. [TO DO: Is this actually a sensible argument??? Jo is not convinced that this is right.]
Following this, focusing the analysis especially on hotter MAPs could be an advisable way to go in the future, if there is doubt that the stars truly follow the qDF.
Another critical point is the binning of stars into MAPs depending on their metallicity and α abundances. Example 2 in §**??** could be understood as a model scenario for decreasing bin sizes in the metallicity-α plane when sorting stars in different MAPs, assuming that there is a smooth variation of qDF within the metallicity-α plane and each MAP indeed follows a qDF. We find that, in the case of 20,000 stars in each bin, differences of 20% in the qDF parameters of two neighbouring bins can still give quite good constraints on the potential parameters. We compare this with the relative difference in the qDF parameters in the bins in Figure 6 of Bovy & Rix (2013), which have sizes of $[\mathrm{Fe/H}] = 0.1$ dex and $\Delta[\alpha/\mathrm{Fe}] = 0.05$ dex. It seems that these bin sizes are large enough to make sure that $\sigma_{R,0}$ and $\sigma_{z,0}$ of neighbouring MAPs do not differ more than 20%. Figure **??** and **??** suggest that especially the tracer scale length h_R needs to be recovered to get the potential right. For this parameter however the bin sizes in Figure 6 of Bovy & Rix (2013) might not yet be small enough to ensure no more than 20% of difference in neighbouring h_R. This is especially the case in the low-α ($[\alpha/\mathrm{Fe}] \lesssim 0.2$), intermediate-metallicity ($[\mathrm{Fe/H}] \sim -0.5$) MAPs, which where however not used in the dynamical modelling by Bovy & Rix (2013). The above is valid for 20,000 stars per MAP. In case there are less than 20,000 stars in each bin the constraints are less tight and due to Poisson noise one could also allow larger differences in neighbouring MAPs while still getting reliable results.
[TO DO: Include the following comments by Jo somewhere: This is a general approach of fitting action-based disk DFs for getting the potential. The qDF is a specific example that we use in this paper. In futures studies different forms of DFs might be fitted to data. That the results are quite robust to the form of the DF not entirely correct motivates this further.]
Gravitational Potential beyond the Parameterized Functions Considered: [TO DO: Comment from HW: In style and contentthis seems verysimilar tothe RESULTSsection. We should eitherdrastically shorten the text in the results section (probably not), or here.] In the long run *RoadMapping* should incorporate a family of gravitational potential models that can reproduce the essential features of the MW's true mass distribution. While our fundamental assumption of the Galaxy's axisymmetry is at odds with the obvious existence of non-axisymmetries in the MW, we will not dive into investigating this implications in this paper. Instead we test how a misjudgment of the parametric potential form affects the recovery by fitting a potential of Stäckel form (Batsleer & Dejonghe 1994) to mock data from a different potential family with halo, bulge and exponential disk. The recovery is quite successful and we get the best fit within the limits of the model. However, even a strongly flattened Stäckel potential component has difficulties to recover the very flattened mass distribution of an exponential disk. This leads to misjudgment of the qDF parameters describing the vertical action distribution, $\sigma_{z,0}$ and h_{σ_z}. As the qDF parameter $\sigma_{z,0}$ corresponds to the physical vertical velocity dispersion at the sun, a comparison with direct measurements could be a valuable cross-checking reference. [TO DO: This might not be true. For isochrone and Staeckel potential I get this behaviour, but not for the MW14-Pot!!! Might be, because it's not separable??? Check!!!] [TO DO: best to simply remove it...] In case of as many as 20,000 stars we should therefore already be able to distinguish between different potential models.
The advantage of using a Stäckel potential with *RoadMapping* is firstly the exact and fast action calculation via the numerical evaluation of a single integral, and secondly that the potential has a simple analytic form, which greatly speeds up calculations of forces and frequencies (as compared to potentials in which only the density has an easy description like the exponential disk). A superposition of several simple Kuzmin-Kutuzov Stäckel components can successfully produce MW-like rotation curves (see Batsleer & Dejonghe (1994), Famaey & Dejonghe (2003) and Figure **??**) and one could think of adding even more components for more flexibility, for example a small roundish component for the bulge. [TO DO: Comment by Jo: In a sense the two approaches (a) using the Staeckel action approximation with a MW like potential and (b) using a Staeckel potential directly are dong the same thing (approximating the true potential as a Staeckel potential). The question is which is best.]
The potential model used by Bovy & Rix (2013) had only two free parameters (disk scale lentgh and halo contribution to $v_{\mathrm{circ}}(R_\odot)$. To circumvent the obvious disadvantage of this being at all not flexible enough, they fitted the potential separately for each MAP and recovered the mass distribution for each MAP only at that radius for which it was best constrained - assuming that MAPs of different scale length would probe different regions of the Galaxy best. Based on our results in Figure **??** this seems to be indeed a sensible approach [TO DO: Check that this is indeed the case - it is not clear to me from the plot. ???].
We suggest that combining the flexibility and computational advantages of a superposition of several Stäckel potential components with probing the potential in different regions with different MAPs as done by Bovy & Rix (2013), could be a promising approach to get the best possible constraints on the MW's potential.

Different Modelling Approaches using Action-based Distribution Functions: We have focussed for the time being 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, e.g., those from different MAPs, must orbit in the same potential. Therefore each MAP will and can yield quite different DF parameters; but each MAP will also provide a (statistically) independent estimate of the potential parameters. At the same time—if all is well—those potential parameters, derived from different MAPs, should be mutually consistent. In some sense,

this approach focusses on constraining the potential, treating the DF parameters as nuisance parameters.
The main drawback is that we have many astrophysical reasons that the DF properties (for reasons of galaxy evolution and chemical evolution) are astrophysically linked between different MAPs. Ultimately, the goal is to do a fully consistent chemodynamical model that simultaneously fits the potential and DF($\boldsymbol{J}$, [X/H]) simultaneously (where [X/Fe] denotes the full abundance space) with a full likelihood analysis. This has not yet been attempted here, because the behaviour is quite complex.
Since the first application of *RoadMapping* by Bovy & Rix (2013) there have been two similar efforts to constrain the Galactic potential and/or orbit distribution using action-based distribution functions:
Piffl et al. (2014) fitted both potential and a $f(\boldsymbol{J})$ to giant stars from the RAVE survey (Steinmetz et al. 2006) and the vertical stellar number density profiles in the disk by Jurić et al. (2008). They did not include any chemical abundances in the modelling. Instead, they used a superposition of action-based DFs to describe the overall stellar distribution at once: a superposition of qDFs for cohorts in the thin disk, a single qDF [TO DO: CHECK] for the thick disk stars and an additional DF for the halo stars. Taking proper care of the selection function requires a full likelihood analysis and the calculation of the likelihood normalisation is computational expensive. Piffl et al. (2014) choose to circumvent this problem by directly fitting a) histograms of the three velocity components in eight spatial bins to the velocity distribution predicted by the DF and b) the vertical density profile predicted by the DF to the profiles by Jurić et al. (2008). The vertical force profile of their best fit mass model nicely agrees with the results from Bovy & Rix (2013) for $R > 6.6$ kpc. The disadvantage of their approach is, that by binning the stars spatially, a lot of information is not used.
Sanders & Binney (2015) have focussed on understanding the abundance-dependence of the DF, relying on a fiducial potential. They developed extended distribution functions, i.e., functions of both actions and metallicity for a superposition of thin and thisk disk, each consisting of several cohorts described by qDFs, a DF for the halo, a functional form of the metallicity of the interstellar medium at the time of birth, and a simple prescription for radial migration. They applied a full likelihood analysis accounting for selection effects and found a best fit for the eDF in a fixed fiducial potential by Dehnen & Binney (1998) to the stellar phase-space and metallicity [TO DO: CHECK] data of the Geneva-Copenhagen Survey (Nordström et al. 2004; Holmberg et al. 2009) and the stellar density curves by Gilmore & Reid (1983). Their best fit predicted the velocity distribution of SEGUE G dwarfs quite well, but had biases in the metallicity distribution, which they accounted to being a problem with the SEGUE metallcities.

We know that real galaxies, including the Milky Way, are not axisymmetric. Using N-body models, we will explore in a subsequent paper what when data from a non-axisymmetric system get interpreted through axisymmetric modelling.

[TO DO: Comment from Jo: Maybe we also want a conclusion with a simple bullet-point list of the main conclusions discussed in detail in the Discussion section.]

REFERENCES

Batsleer, P., & Dejonghe, H. 1994, A&A, 287, 43
Binney, J. 2010, MNRAS, 401, 2318
Binney, J., & McMillan, P. 2011, MNRAS, 413, 1889
Binney, J. 2011, Pramana, 77, 39
Binney, J. 2012, MNRAS, 426, 1324
Binney, J. 2012, MNRAS, 426, 1328
Binney, J. 2013, NAR [TO DO: emulateapj doesn't know NAR], 57, 29
Binney, J., & Tremaine, S. 2008, Galactic Dynamics: Second Edition, by James Binney and Scott Tremaine. ISBN 978-0-691-13026-2 (HB). Published by Princeton University Press, Princeton, NJ USA, 2008.
Bovy, J., & Tremaine, S. 2012, ApJ, 756, 89
Bovy, J., Rix, H.-W., & Hogg, D. W. 2012b, ApJ, 751, 131
Bovy, J., Rix, H.-W., Hogg, D. W. et al., 2012c, ApJ, 755,115
Bovy, J., Rix, H.-W., Liu, C., et al. 2012, ApJ, 753, 148
Bovy, J., & Rix, H.-W. 2013, ApJ, 779, 115
Bovy, J. 2015, ApJS, 216, 29
Büdenbender, A., van de Ven, G., & Watkins, L. L. 2015, MNRAS, 452, 956
Dehnen, W., & Binney, J. 1998, MNRAS, 294, 429
de Lorenzi, F., Debattista, V. P., Gerhard, O., & Sambhus, N. 2007, MNRAS, 376, 71
Famaey, B., & Dejonghe, H. 2003, MNRAS, 340, 752
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Garbari, S., Liu, C., Read, J. I., & Lake, G. 2012, MNRAS, 425, 1445
Gilmore, G., & Reid, N. 1983, MNRAS, 202, 1025
Henon, M. 1959, Annales d'Astrophysique, 22, 126
Holmberg, J., Nordström, B., & Andersen, J. 2009, A&A, 501, 941
Hunt, J. A. S., & Kawata, D. 2014, MNRAS, 443, 2112
Hunt, J. A. S., & Kawata, D. 2014, MNRAS, 443, 2112
Jurić, M., Ivezić, Ž., Brooks, A., et al. 2008, ApJ, 673, 864
Kawata, D., Hunt, J. A. S., Grand, R. J. J., Pasetto, S., & Cropper, M. 2014, MNRAS, 443, 2757
Klement, R., Fuchs, B., & Rix, H.-W. 2008, ApJ, 685, 261
Kuijken, K., & Gilmore, G. 1989, MNRAS, 239, 605
McMillan, P. J. 2011, MNRAS, 414, 2446
McMillan, P. J. 2012, European Physical Journal Web of Conferences, 19, 10002
McMillan, P. J., & Binney, J. J. 2008, MNRAS, 390, 429
McMillan, P. J., & Binney, J. 2012, MNRAS, 419, 2251
McMillan, P. J., & Binney, J. J. 2013, MNRAS, 433, 1411
Nordström, B., Mayor, M., Andersen, J., et al. 2004, A&A, 418, 989
Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, A&A, 369, 339
Piffl, T., Binney, J., McMillan, P. J., et al. 2014, MNRAS, 445, 3133
Read, J. I. 2014, Journal of Physics G Nuclear Physics, 41, 063101
Rix, H.-W., & Bovy, J. 2013, A&A Rev., 21, 61
Sanders, J. L., & Binney, J. 2015, MNRAS, 449, 3479
Sellwood, J. A. 2010, MNRAS, 409, 145
Steinmetz, M., Zwitter, T., Siebert, A., et al. 2006, AJ, 132, 1645
Strigari, L. E. 2013, Phys. Rep., 531, 1
Syer D., Tremaine S. 1996, MNRAS, 282, 223
Ting, Y.-S., Rix, H.-W., Bovy, J., & van de Ven, G. 2013, MNRAS, 434, 652
Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, AJ, 137, 4377
Zhang, L., Rix, H.-W., van de Ven, G., et al. 2013, ApJ, 772, 108

[TO DO: In which order should I give the references????]
[TO DO: replace the references which I typed myself with the ones from ADS.]

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

Table 3
Summary of test suites in this work: The first column indicates the test suite, the second column the potential, DF and selection function model etc. used for the mock data creation, the third model the corresponding model assumed in the analysis, and the last column lists the figures belonging to the test suite. Parameters that are not left free in the analyis, are always fixed to their true value. Unless otherwise stated we calculate the likelihood by the nested-grid and MCMC approach outlined in §?? and use $N_x = 16$, $N_v = 24$, $n_\sigma = 5$ as numerical accuracy for the likelihood normalisation in Equations (11) and (9).

Test		Model for Mock Data	Model in Analysis	Figures
Test 1 : Influence of survey volume on mock data distribution, also in action space	*Potential:*	`KKS-Pot`	-	Mock data: Figure ??
	DF:	`hot` or `cool` qDF		
	Survey volume:	a) $R \in [4, 12]$ kpc,$z \in [-4, 4]$ kpc, $\phi \in [-20°, 20°]$. b) $R \in [6, 10]$ kpc,$z \in [1, 5]$ kpc, $\phi \in [-20°, 20°]$.		
	# stars per data set:	20,000		
	# data sets:	4 (= 2 × 2 models)		
Test 2 : Numerical accuracy in calculation of the likelihood normalisation	*Potential:*	`Iso-Pot`, `MW13-Pot` & `KKS-Pot`	-	Convergence of normalisation: Figure 2
	DF:	`hot` qDF		
	Survey volume:	sphere around sun, $r_{\max} = 0.2, 1, 2, 3$ or 4 kpc		
	Numerical accuracy:	$N_x \in [5, 20]$, $N_v \in [6, 40]$, $n_\sigma \in [3.5, 7]$		
Test 3.1 : *pdf* is a multivariate Gaussian for large data sets.	*Potential:*	`Iso-Pot`	`Iso-Pot`, all parameters free	Figure ??
	DF:	`hot` qDF	qDF, all parameters free	
	Survey Volume:	sphere around sun, $r_{\max} = 2$ kpc	(fixed & known)	
	# stars per data set:	20,000		
	# data sets:	5 (only one is shown)		
	Numerical accuracy:		$N_v = 20$ and $n_\sigma = 4$	
Test 3.2 : Width of the likelihood scales with number of stars by $\propto 1/\sqrt{N}$.	*Potential:*	`Iso-Pot`	`Iso-Pot`, free parameter: b	Figure ??
	DF:	`hot` qDF	`hot` qDF, free parameters: $\ln\left(\frac{h_R}{8\text{kpc}}\right), \ln\left(\frac{\sigma_{R,0}}{230\text{km s}^{-1}}\right), \ln\left(\frac{h_{\sigma,R}}{8\text{kpc}}\right)$	
	Survey volume:	sphere around sun, $r_{\max} = 3$ kpc	(fixed & known)	
	# stars per data set:	between 100 and 40,000		
	# data sets:	132		
	Analysis method:		likelihood on grid	
	Numerical accuracy:		$N_v = 20$ and $n_\sigma = 4$ (for speed)	
Test 3.3 : Parameter estimates are unbiased.	*Potential:*	2 `Iso-Pot` with $b = 0.8$ kpc or $b = 1.5$ kpc	`Iso-Pot`, free parameter: b	Figure ??
	DF:	`hot` or `cool` qDF	`hot`/`cool` qDF, free parameters: $\ln\left(\frac{h_R}{8\text{kpc}}\right), \ln\left(\frac{\sigma_{R,0}}{230\text{km s}^{-1}}\right), \ln\left(\frac{h_{\sigma,R}}{8\text{kpc}}\right)$	
	Survey volume:	5 spheres around sun, $r_{\max} = 0.2, 1, 2, 3$ or 4 kpc	(fixed & known)	
	# stars per data set:	20,000		
	# data sets:	640 (= 2 × 2 × 5 models ×32 realisations)		
	Analysis method:		likelihood on grid	
	Numerical accuracy:		$N_v = 20$ and $N_\sigma = 4$ (for speed)	
Test 4 : Influence of position & shape of survey volume on parameter recovery	*Potential:*	i) `Iso-Pot` or ii) `MW13-Pot`	i) `Iso-Pot`, all parameters free ii) `MW13-Pot`, R_d and f_h free	Figure ??
	DF:	`hot` qDF	i) qDF, all parameters free ii) qDF, only h_R, $\sigma_{z,0}$ and $h_{\sigma,R}$ free	
	Survey volume:	4 different wedges, see Figure ??, upper right panel	(fixed & known)	
	# of stars per data set:	20,000		
	Analysis method:		i) MCMC, ii) likelihood on grid	
Test 5 : Influence of wrong assumptions about the data set (in-)completeness	*Potential:*	`Iso-Pot`	`Iso-Pot`, all parameters free	Illustration & mock data: Figures ?? & ?? Analysis results: Figures ?? & ?? Analysis results:
	DF:	a) `hot` or b) `cool` qDF	qDF, all parameters free	
	Survey volume:	sphere around sun, $r_{\max} = 3$ kpc	(fixed & known)	
	Completeness:	*Example 1:* radial incompleteness, completeness$(r) = 1 - \epsilon_r \frac{r}{r_{\max}}$, twenty $\epsilon_r \in [0, 0.7]$	data set complete, completeness$(r) = 1$, $\epsilon_r = 0$	

Table 3 — *Continued*

Test		Model for Mock Data	Model in Analysis	Figures
on parameter recovery		$r \equiv$ distance from sun, *Example 2:* planar incompleteness, completeness$(z) = 1 - \epsilon_z \frac{\lvert z \rvert}{r_{\max}}$, $\epsilon_r \in [0, 0.7]$, $z \equiv$ distance from Gal. plane.	data set complete, completeness$(r) = 1$, twenty $\epsilon_z = 0$	when not using v_T data: Figure **??**
	# stars per data set:	20,000		
	# data sets:	40 $(= 2 \times 2 \times 20)$		
Test 6.1 : Numerical convergence of convolution with measurement uncertainties	*Potential:*	`Iso-Pot`	"Iso-Pot, all parameters free"	Figure 3
	DF:	`hot` qDF	qDF, all parameters free	
	Survey Volume:	sphere around sun, $r_{\max} = 3$ kpc	(fixed & known)	
	Uncertainties:	$\delta\text{RA} = \delta\text{DEC} = \delta(m - M) = 0$ $\delta v_{\text{los}} = 2$ km/s $\delta\mu_{\text{RA}} = \delta\mu_{\text{DEC}} =$ 2,3,4 or 5 mas/yr	Convolution with perfectly known measurement uncertainties	
	Numerical Accuracy:		$N_{\text{error}} \in [25, 1200]$	
	# stars per data set:	10,000		
Test 6.2 : Testing the convolution with measurement & without uncertainties with distance errors	*Potential:*	`Iso-Pot`	`Iso-Pot`, all parameters free	Figure 4
	DF:	`hot` qDF	qDF, all parameters free	
	Survey Volume:	sphere around sun, $r_{\max} = 3$ kpc	(fixed & known)	
	Uncertainties:	$\delta\text{RA} = \delta\text{DEC} = 0$, $\delta v_{\text{los}} = 2$ km s^{-1}, $\delta\mu_{\text{RA}} = \delta\mu_{\text{DEC}} = 1$, 2,3,4 or 5 mas/yr, a) $\delta(m - M) = 0$ or b) $\delta(m - M) \neq 0$. (see Figure 4)	Convolution with measurement uncertainties, ignoring distance errors in position (see Section 1.5)	
	Numerical Accuracy:		$N_{\text{error}} = 800$ for $\delta\mu \leq 3$mas yr^{-1}, $N_{\text{error}} = 1200$ for $\delta\mu > 3$mas yr^{-1}	
	# stars per data set:	10,000		
Test 6.3 : Underestimation of proper motion errors	*Potential:*	`Iso-Pot`	`Iso-Pot`, all parameters free	Figure 6
	DF:	`hot` or `cool` qDF	qDF, all parameters free	
	Survey volume:	sphere around sun, $r_{\max} = 3$ kpc [TO DO: CHECK]	(fixed & known)	
	Uncertainties:	only proper motion errors 1, 2 or 3 mas/yr	Convolution with proper motion errors 10% or 50% underestimated	
	# stars per data set:	10,000		
	# data sets:	24 $(= 2 \times 2 \times 3 \times 3$ realisations)		
Test 7 : Deviations in the assumed DF from the star's true DF	*Potential:*	`Iso-Pot`	`Iso-Pot`, all parameters free	mock data: Figure **??** Analysis results: Figures **??** & **??**
	DF:	mix of two qDFs *Example 1:* with fixed qDF parameters, but 20 different mixing rates: a) `hot` & `cooler` qDF or b) `cool` & `hotter` qDF *Example 2:* 20 fixed 50/50 mixtures, with varying qDF parameters (by X%): a) `hot` & `colder` qDF or b) `cool` & `warmer` qDF	single qDF, all parameters free	
	Survey volume:	sphere around sun, $r_{\max} = 2$ kpc	(fixed & known)	
	# stars per data set:	20,000		
	# data sets:	40 $(= 2 \times 2 \times 20)$		
Test 8 : Deviations of the assumed potential model from the star's true potential	*Potential:*	`MW14-Pot`	`KKS-Pot`, all parameters free, only $v_{\text{circ}}(R_\odot) = 230$km s^{-1} fixed	potential contours: Figure **??**
	DF:	`hot` or `cool` qDF	qDF, all parameters free	qDF recovery: Figure **??**
	Survey volume:	sphere around sun, $r_{\max} = 4$ kpc	(fixed & known)	
	# stars per data set:	20,000		
	# data sets:	2		

[TO DO: Remove # data sets, where it actually is not important.] [TO DO: Jo suggested to make many tables from this. But I actually like one big table at the end of the paper. Otherwise we had 6 additional tables interrupting the flow of text and figures all the time. And the parameters in the table are really just for reference.] [TO DO: Overall, this table could do with a little less information.]