

## ACTION-BASED DYNAMICAL MODELLING FOR THE MILKY WAY DISK: THE INFLUENCE OF SPIRAL ARMS

WILMA H. TRICK<sup>1,2</sup>, JO BOVY<sup>3</sup>, ELENA D'ONGHIA<sup>4,5</sup>, AND HANS-WALTER RIX<sup>1</sup>

*Draft version August 12, 2016*

### ABSTRACT

*RoadMapping* is a well-tested dynamical modelling machinery developed to constrain the Milky Way's (MW) gravitational potential by simultaneously fitting an axisymmetric parametrized potential model and an axisymmetric action-based orbit distribution function (DF) to discrete 6D phase-space measurements of stars in the Galactic disk. In this work we demonstrate *RoadMapping*'s robustness in the presence of spiral arms by modelling data drawn from a N-body simulation snapshot of a MW-like spiral galaxy in different survey volumes with radii  $500 \text{ pc} \leq r_{\max} \leq 5 \text{ kpc}$ . The potential constraints are surprisingly successful [TO DO: Jo writes: "Remove surprising, maybe robust."], even when using the most simple action-based DF, the quasi-isothermal DF (qDF). The best-fit *RoadMapping* model always recovers the correct gravitational forces where most of the stars that entered the analysis are located. For data from large survey volumes *RoadMapping* finds axisymmetric models that average well over the spiral arms. As expected, the models are slightly biased by the excess of stars in the spiral arms. Survey volumes with  $r_{\max} \leq 3 \text{ kpc}$  give predictions for an overall axisymmetric potential as good as those from larger volumes. However, only for  $r_{\max} = 5 \text{ kpc}$  could the correct halo scale length be recovered. Potential models derived from small survey volumes ( $r_{\max} < 2 \text{ kpc}$ ) centered on inter-arm regions can be more reliably used for extrapolation than those from small volumes affected by spiral arms. We discuss two possible applications of *RoadMapping* to the stellar phase-space measurements from the upcoming Gaia data releases: (i) finding an overall axisymmetric potential model for the MW from data covering a large volume, (ii) measuring non-axisymmetric structures in the MW's disk by modelling data in small volumes independently. [TO DO: Jo writes: "Need some sort of conclusion for these two."] [TO DO: Currently 257 words. Might have to cut a few more.]

*Keywords:* Galaxy: disk — Galaxy: fundamental parameters — Galaxy: kinematics and dynamics — Galaxy: structure — [TO DO]

### 1. INTRODUCTION

The first step in learning more about the Milky Way's (MW) overall gravitational potential and orbit distribution function (DF)—which are crucial for understanding galaxy structure and formation (?)—is to find the “best possible” axisymmetric model for the Galaxy. Given such a model the identification and characterization of non-axisymmetries like the bar, spiral arms or stellar streams in stellar phase-space (and chemical abundance) data would then become more straightforward.

Several approaches to constrain an axisymmetric potential and/or orbit DF have recently been put forward: Bovy & Rix (2013) and Piffl et al. (2014) fitted potential and DF simultaneously to stellar kinematics in the disk and got precise constraints on the overall potential; Sanders & Binney (2015) and Das & Binney (2016) investigated extended DFs for the disk and halo respectively (given a fiducial potential), that included in addition to the distribution in orbit space also the metallicity of each star.

In this work we will continue our investigation of the *RoadMapping* approach (“Recovery of the Orbit Action Distribution of Mono-Abundance Populations and Potential INference for our Galaxy”). The first applica-

tion of *RoadMapping* was done by Bovy & Rix (2013) and Trick et al. (2016), hereafter Paper I, performed a detailed analysis of the strengths and weaknesses of the approach. The idea behind *RoadMapping* is, that simple stellar populations in the MW disk—be it mono-abundance populations (Bovy et al. 2012a,c,b, 2016) (i.e., stars with the same [Fe/H] and [α/Fe]) or maybe also mono-age populations (Bird et al. 2013; Martig et al. 2014, 2016; Minchev et al. 2014; Ness et al. 2016)—follow simple orbit DFs, like, e.g., the quasi-isothermal DF (qDF) by Binney & McMillan (2011) (Ting et al. 2013). The qDF is expressed in terms of the orbital actions  $\mathbf{J} = (J_R, J_\phi = L_z, J_z)$ , which are integrals of motion, quantify the amount of the orbit's oscillation in each of the coordinate directions ( $R, \phi, z$ ) and are therefore excellent orbit labels. Given an assumed gravitational potential one can calculate the orbital actions from the stars' current phase-space positions ( $\mathbf{x}, \mathbf{v}$ ) (Binney 2012; Sanders & Binney 2016) [TO DO: Jo writes: "Bovy (2014) presents general method  $(\mathbf{x}, \mathbf{v}) \rightarrow (\mathbf{J}, \boldsymbol{\theta})$ ."]. Only if the assumed gravitational potential is realistic, this stellar action distribution will follow a realistic orbit DF like the qDF [TO DO: Jo thinks the last half sentence is not very clear.]. This allows one to simultaneously fit potential and orbit DF to observations.

Bovy & Rix (2013) employed this approach to measure the Milky Way's surface density profile within 1.1 kpc using 43 MAPs in the Galactic disk from the SDSS/SEGUE survey (Yanny et al. 2009). Their potential model had only two free parameters (disk scale length and relative halo-to-disk contribution to the radial force at the so-

<sup>1</sup> Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

<sup>2</sup> Correspondence should be addressed to trick@mpia.de.

<sup>3</sup> Department of Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON, M5S 3H4, Canada

<sup>4</sup> Department of Astronomy, University of Wisconsin, 2535 Sterling Hall, 475 N. Charter Street, Madison, WI 53076, USA

<sup>5</sup> Alfred P. Sloan Fellow

lar radius). To account for missing model flexibility they constrained the surface density for each MAP only at one best radius. The profile they derived in this fashion had a scale length of  $R_s = 2.5$  kpc and was—in the regime  $R > 6.6$  kpc—later confirmed by Piffl et al. (2014) using a different action-based procedure.

Given the success of this first application and in anticipation of the upcoming data releases from Gaia in 2016–2022 (Eyer et al. 2013), Paper I improved the *RoadMapping* machinery and studied its strengths and breakdowns in detail, by investigating a large suite of mock data sets. Under the prerequisite of axisymmetric data and model, they found that *RoadMapping*'s modelling success is stable against minor misjudgements of DF or selection function, and that—if the true potential is not contained in the proposed family of model potentials—one can still find a good fit [TO DO: Jo wants me to add “that returns the correct forces” here.], given the limitations of the model. Paper I also found that measurement uncertainties of the order of those by the final Gaia data release should be good enough (within 3 kpc from the Sun) to allow for precise and unbiased modelling results.

The MW is, however, not axisymmetric. The bulge contains a strong bar (Liszt & Burton 1980; Blitz & Spergel 1991; Hammersley et al. 2000; Wegg & Gerhard 2013) and the disk itself is threaded by spiral arms (?Georgelin & Georgelin 1976; Churchwell et al. 2009; Reid et al. 2014) [TO DO: More references] and (ring-like) overdensities (Newberg et al. 2002; Jurić et al. 2008; Xu et al. 2015), which induce non-circular motions and asymmetries in stellar number counts and correspond consequently to a non-axisymmetric gravitational potential and stellar DF. [TO DO: Jo writes: “Could also cite kinematic evidence: moving groups (Eggen (????), ????)) and streaming motions (in 21 cm or velocities) (??), Siebert et al. 2013 or 2014” → Ich finde das Eggen und Siebert paper nicht ...]

We are however planning to ultimately use an axisymmetric model to describe and model our non-axisymmetric Galaxy with *RoadMapping*. This is an important breakdown of modelling assumptions which was not investigated in Paper I. But will *RoadMapping* still be able to give reliable constraints on the MW's gravitational potential in the presence of spiral arms? Answering this question will be this work's objective.

Our investigation makes use of a N-body simulation snapshot of a spiral galaxy with strong spiral arms by D'Onghia et al. (2013). We draw mock data from this simulation in different regions with different spiral arm strengths and run the *RoadMapping* machinery on these data sets and test how well we recover the local and overall gravitational potential.

In Paper I we tested and confirmed separately the robustness of *RoadMapping* in the case that data came from a different model family of either potential or DF than assumed in the dynamical modelling. What would happen if both potential and DF model families were slightly wrong at the same time? The set-up of this study will automatically cover this important test case: The disk potential and orbit DF models that we are using are partly chosen because of their simplicity and computational advantages and—in case of the qDF—because that's what we are planning to use in the MW. We did not have any indications that they might be especially

well-suited to describe an axisymmetric version of this simulated galaxy—quite the contrary.

A third modelling aspect that comes with the breakdown of axisymmetry in the data, is that the assumption that actions are integrals of motion is not given anymore (Vera-Ciro & D'Onghia 2016) [TO DO: more references] [TO DO: Jo writes: “obvious?”]. This somewhat dilutes the advantages of using actions in the first place and it is not clear how well the orbit DF describes the distribution of instantaneous actions estimated in a slightly wrong and axisymmetric potential.

Overall, there are several reasons why the *RoadMapping* modelling could result in severely biased potential constraints when applying it to the spiral galaxy simulation snapshot at hand—or directly to the MW, for that matter. Similar studies, like the ones by Bovy & Rix (2013); Piffl et al. (2014); Sanders & Binney (2015) [TO DO: more references], implicitly assumed that the effect of the spiral arms should not matter too much. In this study we explicitly demonstrate that the *RoadMapping* potential estimates are still surprisingly accurate, which makes us optimistic that they will be for the MW as well and were already in previous similar action-based modelling attempts. [TO DO: Jo writes: “BR13 did not use in-plane motions to avoid spiral arm effects...”]

This paper is organized as follows. Section 2 describes the N-body simulation snapshot of a MW-like spiral galaxy, that we are going to model in this study, explains how we extract 6D stellar phase-space data from it, and how we quantify the spiral arm strength. Section 3 summarizes the *RoadMapping* dynamical modelling framework and introduces the DF and potential model that will fit to the data. Section 4 is dedicated to presenting the results: In Section 4.1 we discuss in detail the *RoadMapping* modelling results derived from a data set within a survey volume with radius  $r_{\max} = 4$  kpc around the Sun. Section 4.2 investigates then a whole suite of *RoadMapping* analyses, corresponding to survey volumes of different sizes and different positions within the galaxy and with respect to the spiral arms. In Section 5 we discuss the results and conclude in Section 6.

[TO DO: Jo writes: “investigated here?” → no idea what he means?]

## 2. DATA FROM A GALAXY SIMULATION

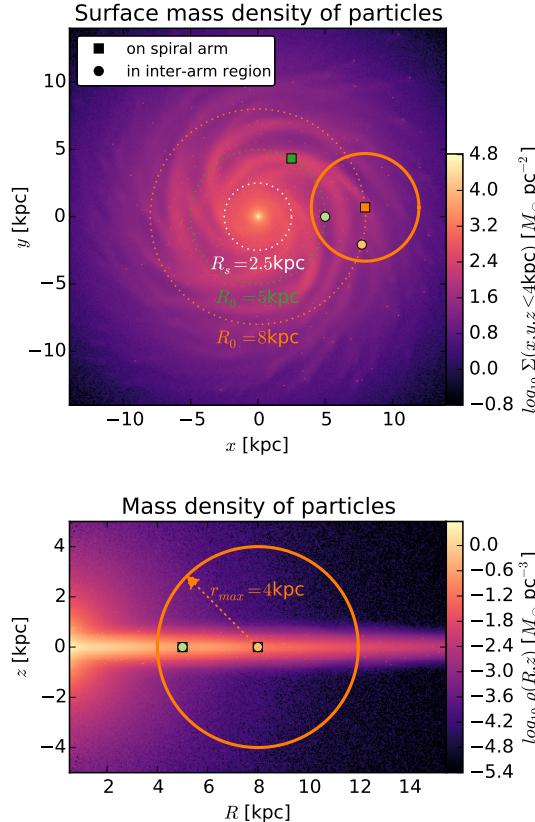
### 2.1. Description of the galaxy simulation snapshot

In this work we use stellar phase-space data drawn from a high-resolution N-body simulation snapshot of a disk galaxy by D'Onghia et al. (2013) carried out with the GADGET-3 code and described in Springel et al. (2005) [TO DO: Jo writes: “not actually described there (Gadget ? is)” → I can't read that.], in which overdensities with properties similar to giant molecular clouds induced prominent spiral arms—and therefore non-axisymmetric sub-structure—via the swing amplification mechanism. For details see D'Onghia et al. (2013), here we summarize the essential characteristics.

The simulation has a gravitationally evolving stellar disk within a static/rigid analytic dark matter halo.

The analytic halo follows a Hernquist (1990) profile

$$\rho_{\text{dm}}(r) = \frac{M_{\text{dm}}}{2\pi} \frac{a_{\text{dm}}}{r(r + a_{\text{dm}})^3} \quad (1)$$



**Figure 1.** Simulation snapshot by D’Onghia et al. (2013). Shown are the surface mass density (in  $(x, y)$  plane, upper panel) and mass density (in  $(R, z)$  plane, lower panel) of the “star” particles belonging to disk, bulge and giant molecular clouds. (The dark matter halo in this simulation is static and analytic and not shown here.) Overplotted are the disk’s scale length  $R_s = 2.5 \text{ kpc}$  (see Section 2.1) and the radii at which we center our test survey volumes in this investigation,  $R = 8$  and  $5 \text{ kpc}$ . The centers of the different survey volumes are marked with a square, if the volume is centered on a spiral arm, or with a circle, if the volume is centered on an inter-arm region. The yellow circle with radius  $r_{\max} = 4 \text{ kpc}$  marks the survey volume in which we conduct the analysis discussed in detail in Section 4.1. [TO DO: overplot names of positions.] [TO DO: Jo writes: “I don’t like boxes around legends”] [TO DO: Jo writes: “Why does this not span the entire y-axis?”]

with total halo mass  $M_{\text{dm}} = 9.5 \cdot 10^{11} M_\odot$  and scale length  $a_{\text{dm}} = 29 \text{ kpc}$  (Elena D’Onghia, private communication).

The disk consists of  $10^8$  “disk star” particles, each having a mass of  $\sim 370 M_\odot$ , and  $1000$  “giant molecular cloud” particles with mass  $\sim 9.5 \cdot 10^5 M_\odot$ . Initially the particles are distributed following an exponential disk profile with density

$$\rho_*(R, z) = \frac{M_*}{4\pi z_0 R_s^2} \operatorname{sech}^2\left(\frac{z}{z_0}\right) \exp\left(-\frac{R}{R_s}\right), \quad (2)$$

with  $R_s = 2.5 \text{ kpc}$  and  $z_0 = 0.1R_s$  and total disk mass  $M_* = 0.04 \cdot M_{\text{dm}} = 3.8 \cdot 10^{10} M_\odot$ .

The bulge consists of  $10^7$  “bulge star” particles with mass  $\sim 950 M_\odot$  and they are distributed following a spherical Hernquist profile analogous to Equation (1), with total mass  $M_{\text{bulge}} = 0.01 \cdot M_{\text{dm}} = 9.5 \cdot 10^9 M_\odot$  and scale length  $a_{\text{bulge}} = 0.1 \cdot R_s = 0.25 \text{ kpc}$ .

The simulation snapshot which we are using in this work has evolved under its own gravity for  $\sim 250 \text{ Myr}$ , which corresponds to approximately one orbital period

**Table 1**  
Vantage points within the galaxy simulation snapshot around which we center survey volumes of radius  $r_{\max}$ .

name	position	$R_0$ [kpc]	$\phi_0$ [degrees]	legend
S8	on spiral arm	8	5	■
I8	in inter-arm region	8	-15	○
S5	on spiral arm	5	60	■
I5	in inter-arm region	5	0	○

Note. — All volumes are centered on  $z_0 = 0$  in the plane of the disk, and  $\phi_0$  is measured counter-clockwise from the positive  $x$ -coordinate axis.

at  $R \sim 8 \text{ kpc}$ . The mass density of simulation particles (without the DM halo) at this snapshot time is shown in Figure 1. Pronounced spiral arms have developed due to the “molecular cloud perturbers”, which can be seen in Figure 1 as small overdensities in the disk. The spherical bulge and very flattened disk are shown in the lower panel in Figure 1.

We have confirmed that the gravitational center of the particles corresponds to the coordinate origin.

## 2.2. Survey volume and data

The selection function of all-sky surveys like Gaia, that are only limited by the brightness of the tracers, are contiguous and—when ignoring anisotropic effects like dust obscuration—spherical in shape. For simplicity we will use spherical survey volumes centered on different vantage points and with sharp edges at a radius  $r_{\max}$  around it (see also Section 3.1), which corresponds to a magnitude cut for stellar tracers all having the same luminosity.

Figure 1 illustrates the different survey volume positions we use in this study: We use each a volume with  $r_{\max} = 1, 2, 3, 4$  or  $5 \text{ kpc}$  centered on a spiral arm and on an inter-arm region at both the equivalent of the solar radius in this simulation,  $R = 8 \text{ kpc}$ , and at  $R = 5 \text{ kpc}$ , where the spiral arms are more pronounced (see Figure 2) than at  $R = 8 \text{ kpc}$ . The exact positions of the vantage points are summarized in Table 1.

From each volume we draw  $N_* = 20,000$  random “disk star” particles from the simulation and use their phase-space positions  $(\mathbf{x}_i, \mathbf{v}_i)$  within the simulated galaxy’s restframe as data.

To make the data sample more realistic, one would actually have to add measurement uncertainties, especially to the distances from the survey volume’s central vantage point and the proper motions measured from there. We decided not to include measurement uncertainties: Firstly, their effect on *RoadMapping* modelling has been already investigated in Paper I and we found that the measurement uncertainties of the last data release of Gaia should be small enough to not disturb the modelling to much. Secondly, in this study we want to isolate and investigate the deviations of the data from the assumed potential and DF model and axisymmetry independently of other effects.

## 2.3. Symmetrized potential model

For a galaxy with pronounced spiral arms an axisymmetric model matter distribution can per se not reproduce the true matter distribution globally. We derived an *overall best fit symmetrized* potential model from the distribution of particles to be able (i) to quantify the non-axisymmetries in the simulation snapshot better and (ii)

**Table 2**  
Best fit parameters of the DEHH-Pot (see Section 2.3), which we use as the global best fit symmetrized potential model.

circular velocity	$v_{\text{circ}}(R_\odot)$	222 km s <sup>-1</sup>
disk scale length	$h_r$	2.5 kpc
disk scale height	$h_z$	0.17 kpc
halo fraction	$f_{\text{halo}}$	0.54
halo scale length	$a_{\text{halo}}$	29 kpc
bulge mass	$M_{\text{bulge}}$	$0.95 \cdot 10^{10} M_\odot$
bulge scale length	$a_{\text{bulge}}$	0.25 kpc

**Note.** — The halo fraction,  $f_{\text{halo}}$ , and circular velocity at the Sun,  $v_{\text{circ}}(R_\odot)$ , which scales the total mass of the model, are defined in Equations (14) and (15).

to compare how close our axisymmetric *RoadMapping* results can get to it.

We derive this model by fitting axisymmetric analytical functions to the density distribution of each of the galaxy component's particles: The bulge and halo follow each a Hernquist profile by construction (see Section 2.1). The disk with its spiral arms however does deviate strongly from its initial conditions in Equation 2. We chose a double exponential disk model to fit the particle distribution in the disk. The best fit parameters for this reference potential, to which we will refer as the DEHH-Potential (Double-Exponential disk + Hernquist halo + Hernquist bulge) in the remainder of this work, are given in Table 2. As can be seen in Figures 4 and ?? in Section 4.1 below, the DEHH-Pot fits the overall density distribution very well. Its density profile might be a little stepper around  $z \sim 0$  than the actual particle distribution, but this should not affect the overall discussion, as the radial density, surface density and disk-to-halo ratio profiles are so well reproduced.

#### 2.4. Quantifying influence of spiral arm

Depending on size and position of the survey volume the spiral arms and inter-arm regions dominate the stellar distribution within the survey volume to different degrees. To quantify the strength of the spiral arm, we introduce the quantity

$$\kappa(x_j, y_j) \equiv \frac{\Sigma_{\text{disk},T}(x_j, y_j \mid |z| \leq 1.5 \text{ kpc})}{\Sigma_{\text{disk},S}(x_j, y_j \mid |z| \leq 1.5 \text{ kpc})} - 1 \quad (3)$$

[TO DO: Jo writes: "When I see  $\kappa$  I think "epicycle frequency". Why not  $\delta$ ?"] where  $\Sigma_{\text{disk},i}$  is the surface density of the disk component of the simulation snapshot ( $i = T$ ) or of the symmetrized snapshot model DEHH-Pot in Section 2.3 ( $i = S$ ) within the survey volume,

$$\Sigma_{\text{disk},i} \equiv \int_{-1.5 \text{ kpc}}^{1.5 \text{ kpc}} \rho_{\text{disk},i}(x_j, y_j, z) \, dz. \quad (4)$$

( $x_j, y_j$ ) is the centroid of a surface area element ( $x_j \pm \delta, y_j \pm \delta$ ) with  $\delta = 0.25 \text{ kpc}$ . ( $x_c = R_{0,c} \times \cos \phi_{0,c}, y_c = R_{0,c} \times \sin \phi_{0,c}, z_c = 0$ ) is the position of the survey volume's center within the simulation, with  $c \in \{\text{S8, I8, S5, I5}\}$  and  $(R_{c,0}, \phi_{c,0})$  given in Table 1. We consider all  $n \simeq \pi r_{\max}^2 / 4\delta^2$  values of  $\kappa(x_j, y_j)$  inside a given survey volume of radius  $r_{\max}$  around position  $c$  and cal-

culate the mean and standard deviation,

$$\langle \kappa \rangle \equiv \langle \kappa(r \leq r_{\max} \mid c) \rangle \equiv \frac{1}{n} \sum_{j=1}^n \kappa(x_j, y_j) \quad (5)$$

$$\sigma_\kappa \equiv \sigma[\kappa(r \leq r_{\max} \mid c)] \equiv \sqrt{\frac{1}{n} \sum_{j=1}^n n [\kappa(x_j, y_j) - \langle \kappa \rangle]^2} \quad (6)$$

with  $(x_j - x_c)^2 + (y_j - y_c)^2 \leq r_{\max}$ .

This gives us information if a spiral arm or an inter-arm region dominates the survey volume ( $\langle \kappa \rangle > 0$  for spiral arms,  $\langle \kappa \rangle < 0$  for inter-arm regions) and how large the relative contrast between spiral arms and inter-arm regions is ( $\sigma_\kappa$ ). For example, small volumes sitting completely within an inter-arm region, or large volumes that contain—in addition to some spiral arms and depleted inter-arm regions—also large areas of unperturbed disk, will have an overall small relative spiral contrast  $\sigma_\kappa$ .

### 3. ROADMAPPING MODELLING

[TO DO: Jo writes: "Refer to paper I for full details!" → Also, I should write an intro sentence here.]

#### 3.1. Likelihood

The data that goes into the modelling are the 6D position and velocity coordinates  $(\mathbf{x}_i, \mathbf{v}_i)$  of  $N_*$  stars within the survey volume. For simplicity we use a purely spatial selection function  $\text{sf}(\mathbf{x})$  of spherical shape,

$$\text{sf}(\mathbf{x}) \equiv \begin{cases} 1 & \text{if } |\mathbf{x} - \mathbf{x}_0| \leq r_{\max}, \\ 0 & \text{otherwise} \end{cases},$$

whose maximum radius  $r_{\max}$  defines the boundary of the survey volume, and which is centered on  $\mathbf{x}_0 \equiv (R_0, \phi_0, z_0 = 0)$ . Given a parametrized potential model  $\Phi(R, z)$  with parameters  $p_\Phi$ , the  $i$ -th star is on an orbit characterized by the orbital actions

$$\mathbf{J}_i \equiv \mathbf{J}[\mathbf{x}_i, \mathbf{v}_i \mid p_\Phi].$$

The probability of stars to be on the orbit  $\mathbf{J}_i$  is proportional to a given orbit distribution function  $\text{df}(\mathbf{J})$  with parameters  $p_{\text{DF}}$ , [TO DO: Jo writes: "DF?" → Maybe I should use DF and SF everywhere instead of df and sf.]

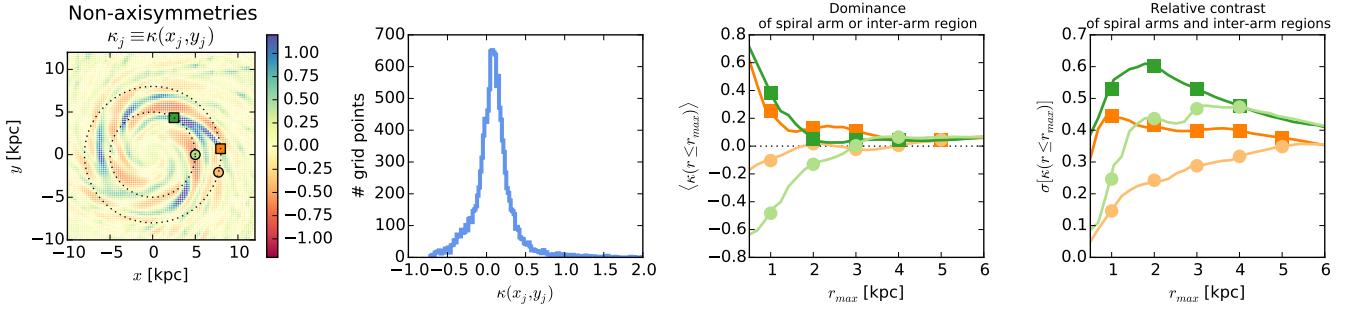
$$\text{df}(\mathbf{J}_i \mid p_{\text{DF}}) \equiv \text{df}(\mathbf{J}[\mathbf{x}_i, \mathbf{v}_i \mid p_\Phi] \mid p_{\text{DF}}) \equiv \text{df}(\mathbf{x}_i, \mathbf{v}_i \mid p_\Phi, p_{\text{DF}}),$$

where the latter equivalence arises from the Jacobian determinant between the angle-action coordinates  $(\boldsymbol{\theta}, \mathbf{J})$  and cartesian phase-space coordinates  $(\mathbf{x}, \mathbf{v})$ , which is  $|\partial(\mathbf{x}, \mathbf{v})/\partial(\boldsymbol{\theta}, \mathbf{J})| = 1$  and therefore allows us to treat the df equivalently as a distribution of current phase-space coordinates or a distribution of orbital actions only, with uniform distribution in the angles  $\boldsymbol{\theta}$ . In some sense, the  $\text{df}(\mathbf{J})$  describes what we expect a realistic phase-mixed stellar population in the MW disk to look like.

The joint likelihood of the  $i$ -th star being on an orbit  $\mathbf{J}$  in the potential  $\Phi$  and being within the survey volume is therefore

$$\mathcal{L}_i \equiv \mathcal{L}(\mathbf{x}_i, \mathbf{v}_i) = \frac{\text{df}(\mathbf{x}_i, \mathbf{v}_i \mid p_\Phi, p_{\text{DF}}) \cdot \text{sf}(\mathbf{x}_i)}{\int \text{df}(\mathbf{x}, \mathbf{v} \mid p_\Phi, p_{\text{DF}}) \cdot \text{sf}(\mathbf{x}) \, d^3x \, d^3v}.$$

The details how we evaluate the likelihood normalisation numerically to sufficiently high enough precision are



**Figure 2.** Contrast and dominance of the spiral arms. The left panel shows the values of  $\kappa_i$  calculated according to Equation 3 as described in Section 2.4 at regular grid points  $(x_i, y_i)$  with bin width 0.25 kpc. Marked are the centroids of the four test survey volumes of this study. The second panel shows a histogram over all  $\kappa(x_j, y_j)$  values in the region  $x, y \in [-14, 14]$  kpc. The two right panels show the dominance (with the mean  $\langle \kappa(r \leq r_{max}) \rangle$  of all grid points within the survey volume as measure) and the relative contrast (with the standard deviation  $\sigma[\kappa(r \leq r_{max})]$  as measure) of spiral arms and inter-arm regions inside a spherical survey volume with given centroid (colour-coded) and size (radius  $r_{max}$  shown on the  $x$ -axis). We chose two volumes in which the spiral arms dominate, and two in which an inter-arm region dominates. The dominance and contrast of spiral arms and inter-arm regions is stronger at  $R = 5$  kpc than at  $R = 8$  kpc. Also, inter-arm regions appear larger and smoother than spiral arms, as already inside a small volume centred on a spiral arm the contrast is quite large. The larger the volume the more does the overall effect of spiral arms and inter-arm regions average out. [TO DO: Add dots for the  $r_{max} = 500$  kpc volumes.] [TO DO: Jo suggested to look into implementing a  $\text{sech}^2$  disk for galpy to not have the slight bias in  $\text{mean}(\kappa)$  at higher  $r_{max}$ . Maybe it's easy to implement. Maybe it isn't.]

discussed in Paper I.<sup>6</sup>

In the scenario considered in this paper it can happen that there are a few ( $\sim 1$  in 20,000) stars entering the catalogue that are for some reason on rather extreme orbits, e.g., moving radially directly towards the center. These kinds of orbits do not belong to the set of orbits that we classically expect to make up a overall smooth galactic disk. To avoid that such single stars with very low likelihood have a strong impact on the modelling we introduce here a simple outlier model,

$$\mathcal{L}_i \rightarrow \max(\mathcal{L}_i, \epsilon \cdot \text{median}(\mathcal{L})),$$

where  $\epsilon = 0.001$  for  $N_* = 20,000$  stars and  $\text{median}(\mathcal{L})$  is the median of all the  $N_*$  stellar likelihoods  $\mathcal{L}_i$  with the given  $p_\Phi$  and  $p_{\text{DF}}$ . This outlier model was not used in Paper I.

[TO DO: Jo writes: "This is a strange outlier model! Realistically a broad Gaussian in v and flat in x might be better." — Yes, might be. But HW suggested this, and now it's to late anyway...]

Following Paper I, we assume for now uninformative flat priors on the model parameters  $p_\Phi$  and  $p_{\text{DF}}$  and find the maximum and width of the posterior probability function

$$\text{pdf}(p_\Phi, p_{\text{DF}} \mid \text{data}) \propto \prod_{i=1}^{N_*} \mathcal{L}_i \cdot \text{prior}(p_\Phi, p_{\text{DF}})$$

using a nested-grid approach and then explore the full shape of the  $\text{pdf}$  using a Monte Carlo Markov Chain (MCMC)<sup>7</sup>. Full details on this procedure are given in Paper I. [TO DO: Jo writes: "Also, don't think you've mentioned galpy yet."]

<sup>6</sup> In the terminology of Paper I we use the high numerical accuracy  $N_x = 20$ ,  $N_v = 28$ ,  $n_\sigma = 5.5$  to calculate the likelihood normalisation, or in other words, to evaluate the spatial and velocity integrals over the qDF within the survey volume. For the action interpolation grid following Bovy (2015), we use  $R_{\text{max}} = 40$  kpc,  $n_E = 70$ ,  $n_\psi = 40$ ,  $n_{L_z} = 50$  in their terminology.

<sup>7</sup> We use the MCMC software *emcee* by Foreman-Mackey et al. (2013).

### 3.2. DF model

The most simple action-based orbit distribution function exhibiting a disk-like structure may be the quasi-isothermal distribution function (qDF) introduced by Binney (2010) and Binney & McMillan (2011). It has proven to be a successful model to describe the orbit distribution of individual mono-abundance populations (MAPs) in the Galactic disk (Bovy & Rix 2013; Ting et al. 2013), that seem to be isothermal in  $z$ -direction (i.e., "quasi-isothermal"). Modelling approaches trying to capture the overall disk distribution (Piffl et al. 2014; Sanders & Binney 2015) were describing the Galactic disk as a superposition of many qDFs. The qDF, which we already used in Paper I, has the functional form

$$\begin{aligned} \text{qDF}(\mathbf{J} \mid p_{\text{DF}}) \\ = f_{\sigma_R}(J_R, L_z \mid p_{\text{DF}}) \times f_{\sigma_z}(J_z, L_z \mid p_{\text{DF}}) \end{aligned} \quad (7)$$

where

$$\begin{aligned} f_{\sigma_R}(J_R, L_z \mid p_{\text{DF}}) &= n \times \frac{\Omega}{\pi \sigma_R^2(R_g) \kappa} \exp\left(-\frac{\kappa J_R}{\sigma_R^2(R_g)}\right) \\ &\times [1 + \tanh(L_z/L_0)] \end{aligned} \quad (8)$$

$$f_{\sigma_z}(J_z, L_z \mid p_{\text{DF}}) = \frac{\nu}{2\pi \sigma_z^2(R_g)} \exp\left(-\frac{\nu J_z}{\sigma_z^2(R_g)}\right) \quad (9)$$

(Binney & McMillan 2011). The guiding-center radius  $R_g$ , circular frequency  $\Omega$ , radial/epicycle frequency  $\kappa$  and vertical frequency  $\nu$  describe the near-circular orbit with given angular momentum  $L_z$  in a given potential. Counter-rotating orbits with  $L_z < L_0$  are suppressed by the term  $[1 + \tanh(L_z/L_0)]$  (with  $L_0 \sim 10 \text{ km s}^{-1} \text{ kpc}$ ). We set the radial stellar tracer density  $n(R_g)$  and velocity dispersion profiles  $\sigma_z(R_g)$  and  $\sigma_R(R_g)$  to

$$n(R_g \mid p_{\text{DF}}) \propto \exp\left(-\frac{R_g}{h_R}\right) \quad (10)$$

$$\sigma_R(R_g \mid p_{\text{DF}}) = \sigma_{R,0} \times \exp\left(-\frac{R_g - R_\odot}{h_{\sigma,R}}\right) \quad (11)$$

$$\sigma_z(R_g \mid p_{\text{DF}}) = \sigma_{z,0} \times \exp\left(-\frac{R_g - R_\odot}{h_{\sigma,z}}\right). \quad (12)$$

The free model parameters of the qDF are therefore

$$p_{\text{DF}} \equiv \{\ln h_R, \ln \sigma_{R,0}, \ln \sigma_{z,0}, h_{\sigma,R}, h_{\sigma,z}\}.$$

Even though we do not have any stellar abundance or age information in the simulation snapshot we are going to investigate (see Section 2.1) and we therefore cannot define stellar sub-populations for which the assumption of such a simple model might be reasonable, we will still try to model the whole disk with a single qDF—to see how far we can get with the simplest possible model. [TO DO: Jo writes: "Plausible, given how the disk is setup, no?"]

### 3.3. Potential model

[TO DO: each equation should have a number]

In all *RoadMapping* analyses in this work we will fit an axisymmetric gravitational potential model to the data consisting of a (fixed and known) Hernquist bulge, a free Hernquist halo and a free Miyamoto-Nagai disk (Miyamoto & Nagai 1975),

$$\Phi_{\text{disk}}(R, z) = -\frac{GM}{\sqrt{R^2 + (a_{\text{disk}} + \sqrt{z^2 + b_{\text{disk}}^2})^2}}, \quad (13)$$

where  $a_{\text{disk}}$  and  $b_{\text{disk}}$  are the equivalents of a disk scale length and scale height. To use Hernquist profiles for halo and bulge is motivated by our knowledge of the snapshot galaxy, and we fix the bulge's total mass and scale length to the true values (see Section 2.1 and Table 3). As the bulge contribution to the total radial force at  $R_{\odot} = 8$  kpc is only  $\sim 9 - 10\%$ , this will not give the modelling an "unfair advantage". The free model parameters of the halo are the halo scale length  $a_{\text{halo}}$  and the relative halo-to-disk contribution to the radial force at the solar radius, i.e.,

$$f_{\text{halo}} \equiv \left. \frac{F_{R,\text{halo}}}{F_{R,\text{disk}} + F_{R,\text{halo}}} \right|_{\substack{R=R_{\odot} \\ z=0}}. \quad (14)$$

As parameter that scales the total mass of the galaxy model we use the circular velocity at the "solar" radius,

$$v_{\text{circ}}(R_{\odot} = 8 \text{ kpc}) \equiv \sqrt{R \left. \frac{\partial \Phi}{\partial R} \right|_{\substack{R=R_{\odot} \\ z=0}}}. \quad (15)$$

The total set of free potential model parameters is therefore

$$p_{\Phi} \equiv \{v_{\text{circ}}(R_{\odot}), a_{\text{disk}}, b_{\text{disk}}, a_{\text{halo}}, f_{\text{halo}}\}. \quad (16)$$

We will call this potential model the MNHH-Pot (Miyamoto-Nagai disk + Hernquist halo + Hernquist bulge) in the remainder of this work.

To estimate the stellar actions  $\mathbf{J} = (J_R, L_z, J_z)$  in the axisymmetric MNHH-Pot, we are using the *Stäckel fudge* algorithm by Binney (2012), which is the best compromise between speed and accuracy and especially well-suited for action estimations in a thin disk (see Sanders & Binney 2016). To speed up the code even more, we also interpolate the actions on a grid (Binney 2012; Bovy 2015) and used a fixed Stäckel focal length of  $\Delta = 0.45$  for the *Stäckel fudge*. We made sure, that the accuracy of the parameter estimates are not degraded by interpolation errors.

For the potential models and the qDF, as well as the *Stäckel fudge*, *RoadMapping* makes extensive use of the *galpy* python library by Bovy (2015)<sup>8</sup>.

The DEHH-Pot introduced in Section 2.3 with its double-exponential disk profile is better than the Miyamoto-Nagai disk in reproducing the overall radial density slope, as we will see in Figures 4 and ?? in the next section, and which is also suggested by the initial disk configuration in Equation (2). However, the closed form expression of the potential in Equation (13) has the crucial advantage of allowing very fast force and therefore action calculations in *RoadMapping* modelling. And the gain in computational speed as compared to using the DEHH-Pot, which requires the evaluation of integrals and Bessel functions to calculate the forces, is essential. In addition, by using a potential model where we already know that it is not be the optimal model for the galaxy's disk, we challenge *RoadMapping* even further.

How should one compare the *RoadMapping* results for the best-fit MNHH-Pot with the fixed DEHH-Pot (Table 2) in the remainder of this work? When considering a *global* best fit potential model for the galaxy, *RoadMapping* model can not be better than the DEHH-Pot, which is therefore a global upper limit for *RoadMapping*'s modelling success. We note that in deriving the DEHH-Pot in Section 2.3 we used the correct decomposition into bulge, disk and halo components. *RoadMapping* however is only sensitive to the composite gravitational effect of the potential. Given the flexibility of the MNHH-Pot model used in *RoadMapping* and the restriction of the modelling to a survey volume, it could be possible that *RoadMapping* can do better in fitting the local potential inside the survey volume at the price of recovering the global potential parameters to less accuracy. We will discuss this in the following sections.

## 4. RESULTS

We run the *RoadMapping* modelling on several data sets drawn from different survey volumes in the simulation snapshot introduced in Section 2.1. In Section 4.1 we will analyze one of these *RoadMapping* analyses in detail, and in Section 4.2 we compare the different results for the different data sets.

### 4.1. An axisymmetric galaxy model from *RoadMapping*

In this section we will discuss the results of the *RoadMapping* analysis of a single data set. This fiducial data set has  $N_* = 20,000$  stars that were drawn from the spherical volume with  $r_{\text{max}} = 4$  kpc centered on a spiral arm at the "solar" radius  $R = 8$  kpc shown as yellow sphere in Figure 1 (position S8). We use *RoadMapping* to fit a single qDF (see Section 3.2) and the potential model MNHH-Pot consisting of a Miyamoto-Nagai disk, a Hernquist halo and a (fixed) Hernquist bulge introduced in Section 3.3 to it. The best fit parameters are summarized in Table 3. The circular velocity  $v_{\text{circ}}(R_{\odot})$  and halo fraction  $f_{\text{halo}}$  are especially well recovered (compare to Table 2).

#### 4.1.1. Recovering the stellar distribution

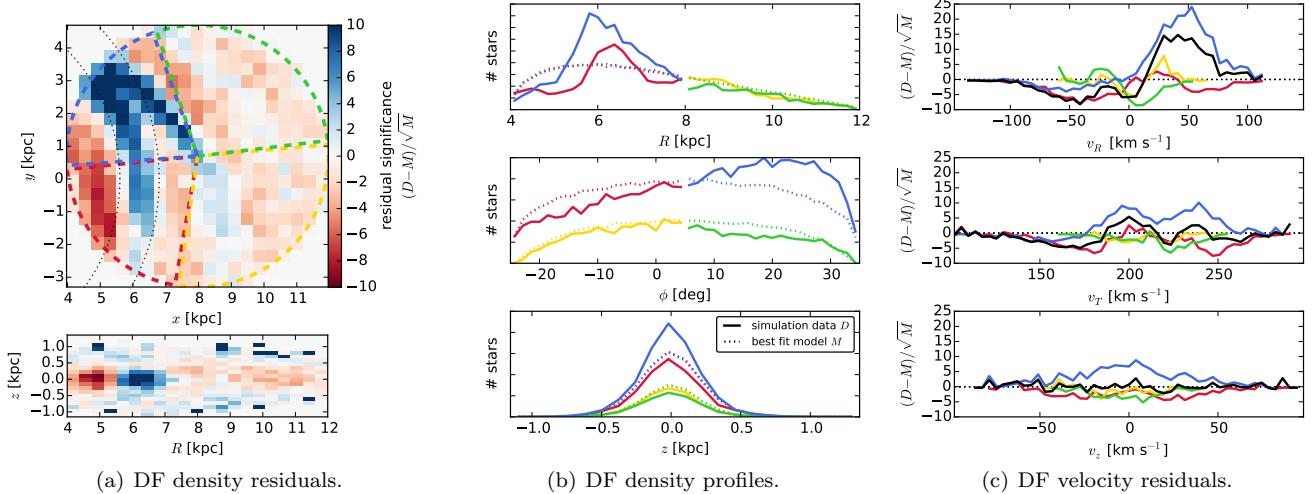
<sup>8</sup> The *galpy* python package by Bovy (2015) can be downloaded from <http://github.com/jobovy/galpy>.

**Table 3**

Best-fit MNHH-Pot and qDF parameters as recovered from the *RoadMapping* analysis of a survey volume with  $r_{\max} = 4$  kpc centered on a spiral arm at  $R_0 = 8$  kpc (position S8).

circular velocity	$v_{\text{circ}}(R_{\odot})$	$(223.0 \pm 0.1) \text{ km s}^{-1}$
Miyamoto-Nagai disk scale length	$a_{\text{disk}}$	$(3.62^{+0.06}_{-0.05}) \text{ kpc}$
Miyamoto-Nagai disk scale height	$b_{\text{disk}}$	$(0.26 \pm 0.02) \text{ kpc}$
halo fraction	$f_{\text{halo}}$	$(0.53 \pm 0.02)$
halo scale length	$a_{\text{halo}}$	$(21 \pm 2) \text{ kpc}$
bulge mass	$M_{\text{bulge}}$	$0.95 \cdot 10^{10} M_{\odot}$ (fixed)
bulge scale length	$a_{\text{bulge}}$	0.25 kpc (fixed)
qDF tracer scale length	$h_R$	$3.34^{+0.05}_{-0.04} \text{ kpc}$
qDF radial velocity dispersion	$\sigma_{R,0}$	$15.91 \pm 0.08 \text{ km s}^{-1}$
qDF vertical velocity dispersion	$\sigma_{z,0}$	$14.0^{+0.2}_{-0.1} \text{ km s}^{-1}$
qDF radial velocity dispersion scale length	$h_{\sigma,R}$	$(4.6 \pm 0.5) \text{ kpc}$
qDF vertical velocity dispersion scale length	$h_{\sigma,z}$	$(5.65^{+0.06}_{-0.07}) \text{ kpc}$

Note. — The bulge mass and scale length were fixed in the analysis to their true values, see Sections 2.1 and 3.3.



**Figure 3.** Comparison of the DF of stars in position-velocity space in the data set from the simulation snapshot (with  $N_* = 20,000$  and  $r_{\max} = 4$  kpc centered on position S8) and in the single-qDF best fit *RoadMapping* model from Table 3. Panel 3(a) shows the spatial density residual significance projected to the  $(x, y)$  and  $(R, z)$  plane. In the  $(x, y)$  panel the following regions are marked:  $R < 8$  kpc,  $\phi > 5^\circ$  (blue),  $R < 8$  kpc,  $\phi < 5^\circ$  (red),  $R > 8$  kpc,  $\phi > 5^\circ$  (green),  $R > 8$  kpc,  $\phi < 5^\circ$  (yellow). Panels 3(b) and 3(c) show the density profiles and velocity residual significance along each of the 6D phase-space coordinates separately for each of the four spatial regions. The blue region is very much dominated by the non-axisymmetric spiral arm. For the yellow region the axisymmetric single-qDF model is a good description. Overall the qDF is a good average axisymmetric model for the data. (In panel 3(a) we overplot the radii  $R_{\text{spiral}} \in [5.6, 6.8]$  kpc as black dotted lines to mark the approximate extent of the stronger spiral arm, to compare it with Figure 7.) [TO DO: Move legend to R panel and make it bigger] [TO DO: Jo writes: "I don't like boxes around legends"] [TO DO: Jo writes: "What is the black line?" — Mention in caption.]

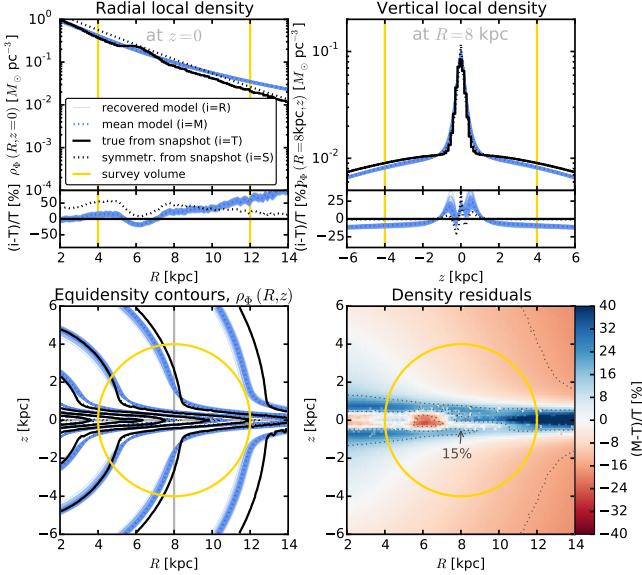
In Figure 3 we compare the configuration phase-space distribution of the 20,000 stars that entered the *RoadMapping* analysis as data with the best-fit stellar distribution, which is given by the action-based qDF and potential model found with *RoadMapping*. We note that the spiral arms introduce very strong non-axisymmetries in the data, both in the spatial and the velocity distribution (especially in  $v_R$ , where a lot more stars move outward than inward as compared to an axisymmetric model). We therefore compare in Figure 3 the data and fit separately for different spatial regions,  $R > 8$  kpc and  $R < 8$  kpc, and  $\phi < 5^\circ$  and  $\phi > 5^\circ$ . In the region where the spiral arm dominates (blue) the best fit *RoadMapping* model is actually a very poor model. However, what the model underestimates in the spiral arm, it slightly overestimates in the other regions and is there-

fore indeed something like a good average model for the overall distribution. Especially the region at  $R > 8$  kpc and  $\phi < 5^\circ$ , where neither the spiral arm nor the inter-arm regions dominate strongly, is very well described by the model. This is remarkable because we had no indication beforehand that a single qDF might be at all a good enough model to describe the overall stellar distribution in the simulation snapshot. But it obviously does—apart from the spiral arm of course, which we had no chance to capture anyway.

#### 4.1.2. Recovering the gravitational potential

As shown in the previous section the best fit *RoadMapping* model seems to reproduce the stellar phase-space distribution quite well. But is the corresponding potential close to the true potential?

Figures 4 and ?? compare the true simulation snapshot



**Figure 4.** Comparison of the true density distribution  $\rho_{\Phi,T}$  in the galaxy simulation snapshot (solid black line, averaged over  $\phi$ ) with the axisymmetric density distribution  $\rho_{\Phi,R}$  recovered with *RoadMapping* (solid blue lines) from  $N_* = 20,000$  stars in the survey volume S8 with  $r_{\max} = 4$  kpc (yellow line). The first two panels show density profiles along  $(R, z = 0)$  and  $(R = 8 \text{ kpc}, z)$ , together with the relative differences between true and recovered  $\rho_\Phi$ . The third panel displays equidensity contours of the matter distribution in the  $(R, z)$  plane. Overplotted are also the overall axisymmetric DEHH-Pot's  $\rho_{\Phi,S}$  (dotted black line) (see Section 2.3) and the  $\rho_{\Phi,M}$  of the recovered mean model in Table 3 (dotted blue line). The last panel shows the relative difference between the true  $\rho_{\Phi,T}$  (averaged over all  $\phi$ ) and the recovered mean model  $\rho_{\Phi,M}$ . Over wide areas even outside of the survey volume the relative difference is less than 15%. At  $R \gtrsim 8$  kpc and  $z \sim 0$  it becomes apparent that the chosen potential model cannot perfectly capture the structure of the disk. However, in the plane of the disk and at smaller radii within the survey volume, where most of the stars are located, the model gives good constraints on the density. **TO DO: Redo plot. Now I use the median of MCMC samples as best fit model.]** **[TO DO: Make survey volume dark orange.]** **[TO DO: Plot vertical profile from -5 to +5 only, to see disk better.]** **[TO DO: Mention different potential model names in legend]** **[TO DO: Maybe show the profiles not at  $R=8$  and  $z=0$ , but at  $R=6$  and  $z=h_z$  or so.]** **[TO DO: Decide if we want to include the local density (averaged over survey volume's wedge only) in this figure as well.]** **[TO DO: Jo writes: "I don't like boxes around legends"]** **[TO DO: Set R entry in legend by hand. It does not properly print, when it uses the thin line.]**

potential and the axisymmetric DEHH-Pot from Table 2 with the best fit MNHH-Pot from the *RoadMapping* analysis. In particular, Figure ?? the rotation curve, surface density profile and halo-to-disk ratio and Figure ?? compares the true and recovered (median) gravitational forces at the position of each star in the data set. In Figure ?? we explicitly show the true simulation snapshot circular velocity both averaged over the whole  $\Delta\phi = 2\pi$  (black lines), and over the survey volume's maximum radial extent  $\phi_0 \pm \arcsin(r_{\max}/R_0)$  (grey lines).

The recovery of density, surface density, circular velocity curve and disk-to-halo ratio is especially good in the region where most of the stars are located, around  $R \sim 6$  kpc and in the plane of the disk. In large regions inside the survey volume, and even outside, the density is recovered to within 15%. The circular velocity curve is even recovered to within 5%, which is also approxi-

mately the extent of perturbation that the spiral arms cause with respect to a smooth rotation curve.

There is however a very small ( $< 1.5\%$ ) underestimation of  $v_{\text{circ}}$  at larger radii. We suspect that this is a combination of a bias introduced by the spiral arms (see also discussion in Section 4.2.3) and the choice of potential model (as a similar bias shows up in the DEHH-Pot). The overall surface density profile is a bit overestimated ( $\sim 15\%$ ) at smaller radii; the reason is clearly the local spiral arm at  $R \sim 6$  kpc with its higher surface density and many stars entering the analysis, which bias the result and which was to be expected. At larger radii ( $R \sim 10 - 12$  kpc) the fit of radial local density and surface density profile is less good, which we account to the Miyamoto-Nagai disk having a shallower radial profile than an exponential disk, and not enough stars in this outer regions to give good constraints. The much stronger bias in the disk-to-halo ratio at larger radii is the result of a misjudgement of the halo scale length. As we will see later (see Figure 8) we seem to need an even larger survey volume to have enough radial coverage to constrain the halo scale length properly. But again, where most of the stars are located, our *RoadMapping* model is a very good average model for the true galaxy.

The aspect of the potential to which the stellar orbits are actually sensitive to, are the gravitational forces. In Figure ?? we compare therefore the true force that each star in the data set feels ( $F_{R,T}(*_i)$  and  $F_{T,M}(*_i)$ , calculated as sum of the individual contributions by each particle in the simulation and the analytic dark matter halo) with the force that the *RoadMapping* median model predicts for each star ( $F_{R,M}(*_i)$  and  $F_{z,M}(*_i)$ ). We scale the difference between truth and model by a typical radial or vertical force at the given radius for which we use

$$F_{R,\text{typ}}(R) \equiv v_{\text{circ},S}^2(R)/R \quad (17)$$

$$F_{z,\text{typ}}(R) \equiv F_{z,S}(R, z = h_z), \quad (18)$$

where  $v_{\text{circ},S}$  and  $F_{z,S}$  are the circular velocity and vertical force evaluated in the "true symmetric" DEHH-Pot in Table 2. As typical vertical force at a given radius we use the vertical force evaluated at the scale height  $h_z = 0.17$  kpc of the disk, because we found that

$$\begin{aligned} &\langle |F_{z,S}(*_i)|(R) \rangle \\ &\lesssim \langle |F_{z,T}(*_i)|(R) \rangle \\ &\lesssim F_{z,S}(R, z = h_z) \\ &\lesssim \langle J_z(*_i) \times \Omega_z(*_i)/h_r \rangle(R), \end{aligned}$$

where, e.g.,  $\langle |F_{z,T}(*_i)|(R) \rangle$  denotes the mean of the absolute value of the true vertical force of all stars at radius  $R$ , and  $J_z \times \Omega_z \sim \langle E_z \rangle$  is close to the mean vertical energy of a stellar orbit. Figure ?? shows therefore

$$\Delta F_R(*_i) \equiv \frac{|F_{R,M}(R_i, z_i)| - |F_{R,T}(x_i, y_i, z_i)|}{F_{R,\text{typ}}(R_i)} \quad (19)$$

$$\Delta F_z(*_i) \equiv \frac{|F_{z,M}(R_i, z_i)| - |F_{z,T}(x_i, y_i, z_i)|}{F_{z,\text{typ}}(R_i)} \quad (20)$$

for each star. The recovery is as expected: The true vertical force is stronger in the spiral arms (due to the higher surface density) and weaker in the inter-arm regions as compared to the axisymmetric best fit model. The radial force is well recovered where the majority of

the stars are located, i.e. in the wide inter-arm regions and in the peaks of the spiral arms. Misjudgments happen in the wings of the spiral arms: The true radial force (i.e., the pull towards the galactic center) is stronger at the outer edge/leading side of the spiral arm because of the additional gravitational pull towards the massive spiral arm, and for the same reason weaker at the inner edge/trailing side. Overall the recovered *RoadMapping* model appears to be a good mean model, averaging over spiral arms and inter-arm regions.

#### 4.1.3. Recovering the action distribution

In Sections 4.1.1 and 4.1.2 we have demonstrated the goodness of fit in the configuration space of the data, and of the recovered gravitational potential. What *RoadMapping* is actually fitting is however the distribution in action space. Figure 7 compares the data and the model action distribution given the best fit MNHH-Pot in Table 3. (We only use this axisymmetric potential to estimate the actions as they enter the fit, and do not calculate the true actions in the true potential.)

We note that the radial and vertical action distribution fits indeed quite well; the axisymmetric model however contains much more stars on close to circular orbits ( $J_R \sim 0, J_z \sim 0$ ) than the simulation does. In the data set there is an excess of stars in the galactic plane ( $J_z \sim 0$ ) that have more eccentric orbits than the axisymmetric model would have. In Figure 3(a) we have marked the radial extent of the stronger spiral arm with dotted lines ( $R_{\text{spiral}} \in [5.6, 6.8]$  kpc), and over-plotted the corresponding angular momenta  $L_z = R_{\text{spiral}} \times v_{\text{circ}}(R_{\text{spiral}})$  in Figure 7 as a rough estimate where in action space we expect the stars of this spiral arm. It is again obvious that this spiral arm contains (i) more stars in general and (ii) more stars with eccentric orbits ( $J_R > 0$ ) which are (iii) located mostly close to the plane ( $J_z \sim 0$ ), as compared to the axisymmetric model. All of this confirms our expectations for orbits in a spiral arm.

This exercise of comparing the data and the model actions in a best-fit axisymmetric potential should be performed for a future application to data in the Milky Way as well. It could help to learn about the approximate orbits that stars move on in real spiral arms and to subsequently construct [TO DO: Jo writes: "perturbed (?)"] orbit distribution functions for spiral arms.

[TO DO: Jo writes: "If you make the equivalent of Figure 7 for fits that are especially good and those that are especially bad, is the difference between the qDF prediction and the data that went into the analysis very different? If so, that might be nice to point out and then say that we will investigate this in later work."]

### 4.2. The influence of spiral arms in *RoadMapping* modelling

In the previous section we showed that for a large enough survey volume ( $r_{\text{max}} = 4$  kpc) *RoadMapping* can construct a good average axisymmetric potential (and DF) model for a galaxy with spiral arms. In the following we want to investigate how this modelling success depends on the position and the size of the survey volume within the galaxy and with respect to the spiral arms.

#### 4.2.1. Test suite

We center our test survey volumes at the positions marked in Figures 1 and 2 (see also Table 1) and consider volume sizes with  $r_{\text{max}} \in [0.5, 1, 2, 3, 4, 5]$  kpc for  $R_0 = 8$  kpc and  $r_{\text{max}} \in [0.5, 1, 2, 3, 4]$  kpc for  $R_0 = 5$  kpc (to avoid the galactic center). As demonstrated in Figure 2 the spiral arm strength is very different in these test volumes. Each data set that we draw from the simulation contains a random selection of  $N_* = 20,000$  stars inside the given spherical volume and we fit a single qDF and MNHH-Pot to it.

#### 4.2.2. Discussion of the model parameter recovery

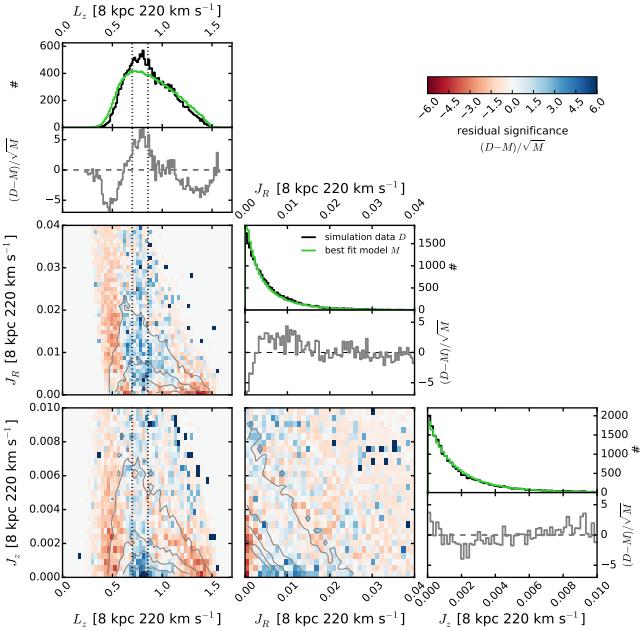
Figure 8 shows the best fit parameters for the different *RoadMapping* analyses. We also overplot the parameters of the reference DEHH-Pot from Table 2. Overall the statistical random errors on the parameter recovery are very small for  $N_* = 20,000$  and possible systematic errors dominate.

There are only a few exceptions ( $r_{\text{max}} = 1$  kpc at I5,  $r_{\text{max}} = 2$  kpc at I8), which we will discuss later.

Let's first consider the parameters of the gravitational potential: All volumes recover  $v_{\text{circ}}(R_\odot)$  within a few  $\text{km s}^{-1}$ ; in the largest volumes, where the circular velocity curve is probed over several kpc the estimate is the most accurate. The halo fraction  $f_{\text{halo}}$  of the radial force at the Sun is very well recovered, especially for  $r_{\text{max}} \gtrsim 2$  kpc. The estimate that we get for the best fit Miyamoto-Nagai disk scale height  $b_{\text{disk}}$  seems to be also approximately independent of the size of the volume. We can even recover the true halo scale length  $a_{\text{disk}}$ , however only for a volume as large as  $r_{\text{max}} = 5$  kpc. The models at  $r_{\text{max}} = 500$  kpc appear to be too small to constrain the halo at all. Smaller volumes that underestimate  $a_{\text{halo}}$  get slightly larger estimates for the disk scale length  $a_{\text{disk}}$  and the overall radial density slope is then probably closer to the truth, even if the individual parameters are not. Outliers can be often explained by having a look at the data: The large disk scale length recovered from the  $r_{\text{max}} = 2$  kpc volume at S5 for example mirrors the comparably flat matter distribution caused by two spiral arms close together and dominating the volume (see the large  $\sigma_\kappa$  for this analysis in the right panel of Figure 2, and Figure A3).

The right column of Figure 8 compares the recovered qDF parameters for the different survey volumes with the qDF parameters we got from fixing the potential model to the DEHH-Pot and fitting the qDF only in a  $r_{\text{max}} = 5$  kpc volume. Even though the qDF parameters for small volumes are widely different for different positions within the galaxy, they all approach the values recovered with the DEHH-Pot for larger volumes. There seems therefore to be indeed an overall best-fit qDF describing the average tracer distribution in the galaxy's disk. The only difference is in the  $h_{\sigma,z}$  parameter, where the models with free MNHH-Pot recover a slightly larger value than the models using the known DEHH-Pot. The reason is that the Miyamoto-Nagai disk has a different radial profile than the double exponential-disk (see Figure 4), which leads to a less steep radial decline in the vertical forces and therefore mean vertical orbital energies  $\langle E_z \rangle \sim \nu \times J_z$  and therefore to a slightly longer  $h_{\sigma,z}$  scale length. In general, volumes centered on spiral arms have larger velocity dispersion parameters  $\sigma_{R,0}$  and  $\sigma_{z,0}$  as compared to volumes at the same radius  $R_0$ .





**Figure 7.** Comparison of the stellar action distribution of the data set  $D$  used in the analysis and the recovered axisymmetric stellar distribution  $M$  (see Figure 3 for the distribution in configuration space). All actions, of data set and best fit distribution, were calculated in the best fit MNHH-Pot in Table 3. The upper panel in each column compares the one-dimensional histograms of the angular momentum  $L_z$ , the radial action  $J_R$  and the vertical action  $J_z$  of both data and best fit. The other one-dimensional and two-dimensional distributions show the residual significance  $(D - M) / \sqrt{M}$ . The two-dimensional residuals are over-plotted with equidensity contours of the data  $D$ 's two-dimensional action distribution. In Figure 3(a) we have marked the approximate radial extent of the stronger spiral arm with black dotted lines ( $R_{\text{spiral}} \in [5.6, 6.8]$  kpc); the dotted lines in the  $L_z$  distributions in this figure correspond to  $L_z = R_{\text{spiral}} \times v_{\text{circ}}(R_{\text{spiral}})$ . This comparison gives a first impression of how the approximate action distribution in spiral arms might look like. [TO DO: Jo wants me to move the following sentence to text: "An open task that the Galactic dynamical modelling community faces in the future, is to construct orbit distribution functions for spiral arms."] [TO DO: Calculate reference data new. Now I use the median of MCMC samples as best fit model.] [TO DO: Add in legend that simulation data was calculated in best fit potential too.]

but centered on an inter-arm region, and the volumes at  $R_0 = 5$  kpc with their stronger spiral arms have larger velocity dispersions than those at  $R_0 = 8$  kpc—which is what we expect. Most volumes recover similar tracer scale lengths  $h_R \sim 2.5 \pm 0.5$  kpc close to the known disk scale length; only the volumes centered on the inter-arm region at  $R_0 = 8$  kpc (position I8) recover much longer  $h_R$ . This might be related to the fact, that volumes at I8 are dominated by a especially large inter-arm region.

There are a few survey volumes for which the recovered parameters show some peculiarities: The models from volumes with  $r_{\max} = 1$  kpc at I5 and  $r_{\max} = 2$  kpc at I8 reject the dark matter halo completely, i.e.,  $f_{\text{halo}} = 0$ . The corresponding halo scale lengths  $a_{\text{halo}}$  are therefore also pretty unconstrained, while the corresponding disk scale heights  $b_{\text{disk}}$  are grossly overestimated to account for the missing contribution of the spherical halo. We have investigated the reason for this fitting result and found, that for the way how the spiral arms affect the circular velocity curve in these volumes, the recovered models with unusual radial profiles are indeed a better

description for the data (see Figure A2-A3). Also, while most analyses average the vertical forces radially over the spiral arms (see Figures ??, lower right panel), for these analyses the averaging happens vertically, i.e., at some height above the plane [TO DO: check if this is approximately the scale height] the model's vertical forces are equally good at all radii (in spiral arms and between), while at small and large  $|z|$  the model is bad. *RoadMapping* therefore found a good average fit model also for the stars in these volumes.

[TO DO: Jo writes: "Stress importance of large volume already here."]

[TO DO: HW suggests to include the 500pc8Void and 1kpc8Void analyses, mark them with red and mention that the MCMCs did not properly converge, because the tracer scale length cannot be constrained because of the position between the spiral arms.]

[TO DO: Make the same diagnostic plots for 1kpc5Void as for 2kpc8Void and check that my explanation of  $\text{halo}=0$  in this text is indeed correct. Also, discuss separately that 500pc8Void and 1kpc8Void MCMCs did not converge.]

#### 4.2.3. Recovery of the gravitational forces depending on the size of the survey volume

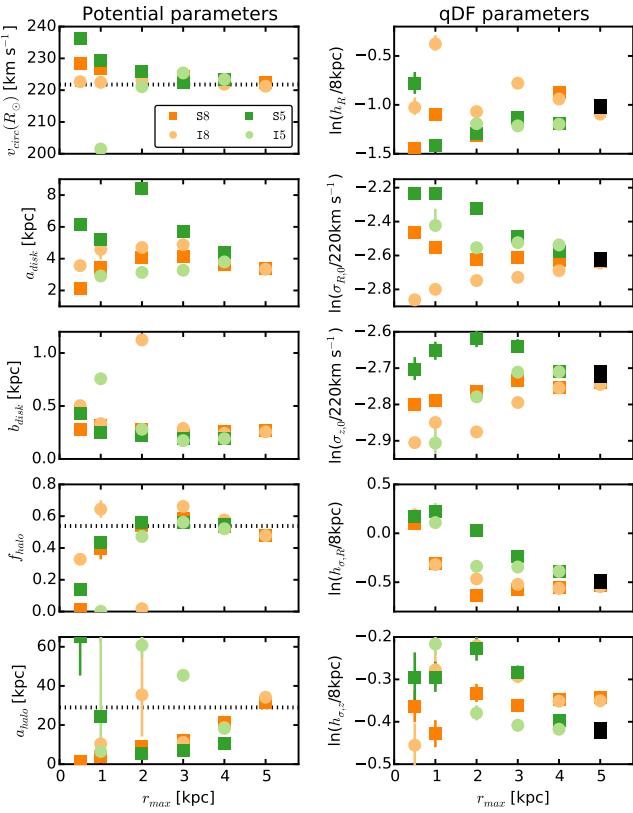
In the previous section we found that the potential and qDF parameters recovered from different survey volumes can be quite different. While the differences can be qualitatively explained, it is not clear yet how good the corresponding potential constraints actually are in a quantitative sense. To test this we calculate again  $\Delta F_R(*_i)$  and  $\Delta F_z(*_i)$  from Equations (19)-(20) at the position of each star  $*_i$  in each data set (see also Figure ??). From the corresponding histograms we derive the median and the 16th and 84th percentiles ( $1\sigma$  range) and show them in the left panels of Figure 9. We chose this diagnostic because the forces at the positions of the stars are the quantities of the potential to which our modelling is sensitive.

In addition it is interesting to see, how good a recovered potential describes the overall gravitational potential of the galaxy, i.e. its predictive power. We introduce another diagnostic, that uses a cylindrical grid centered on the respective positions in Table 1, having always a radius of  $r_{\max} = 5$  kpc and a height of  $z = 1.5$  kpc both above and below the plane. In the  $(x, y)$  plane the regular grid points have a distance of 0.25 kpc and in  $z$  they have a distance of 0.125 kpc to better sample the thin disk. (We throw out grid points close to the galactic center with  $R < 0.125$  kpc, however.) [comment: This was decided completely arbitrary. I guess we can think of a slightly different grid, that is more motivated.] We then evaluate at the position  $g_j \equiv (x_j, y_j, z_j)$  of each regular grid point the force residuals

$$\Delta F_R(g_j) \equiv \frac{|F_{R,M}(R_j, z_j)| - |F_{R,T}(x_j, y_j, z_j)|}{F_{R,\text{typ}}(R_j)} \quad (21)$$

$$\Delta F_z(g_j) \equiv \frac{|F_{z,M}(R_j, z_j)| - |F_{z,T}(x_j, y_j, z_j)|}{F_{z,\text{typ}}(R_j)}, \quad (22)$$

analogous to Equations (19)-(20). The two panels on the right in Figure 9 show the [16th,84th] percentile range and the median of the grid points' distribution in  $\Delta F_R(g_j)$  and  $\Delta F_z(g_j)$ .



**Figure 8.** [TO DO: Check that caption is still fine.] Overview over the model parameter estimates (MNHH-Pot parameters on the left, qDF parameters on the right) recovered with *RoadMapping* from data sets drawn from the simulation snapshot at different positions in the galaxy (colour-coded) and of different size ( $r_{\max}$  as indicated on the  $x$ -axis). The black dotted line shows the known model parameters  $a_{\text{disk}}$  and  $b_{\text{disk}}$  are related but not directly comparable to an exponential disk scale length and height). The black squares denote the qDF parameters we recovered by fixing the potential to the DEHH-Pot, centering a survey volume with  $r_{\text{kpc}} = 5$  on the spiral arm at  $R = 8$  kpc, and fitting the qDF only. A survey volume with a radial coverage as large as  $r_{\max} = 5$  kpc is required to properly recover the "true" model parameters. [TO DO: Preliminary. Add more results. Make sure that the "true" potential parameters are the final version. [TO DO: Add additional y-axis on right side of plot with non-log values.] [TO DO: Jo writes: "I don't like boxes around legends"]]

The first and most important thing to learn from Figure 9 is, that we get very close to recovering the true forces at the positions of the stars in the survey volume, no matter how large the survey volume is.

Second, if we consider not the stars inside the survey volume, but the spatial average of the forces in the fixed cylindrical volume ( $r_{\max} = 5$  kpc, as described above), the radial forces are overall very well recovered, especially for large survey volumes. There is however a overestimation of  $\sim 5 - 20\%$  in the vertical forces (depending on volume size and position) which is induced by the spiral arms (see explanation below).

Third, the constraints we get on the spatially averaged forces inside  $r < 5$  kpc are almost as good for the survey volume of  $r_{\max} = 3$  kpc as compared to survey volumes of  $r_{\max} = 4$  or 5 kpc. If we had to decide between a  $r_{\max} = 3$  kpc volume with good data quality and a larger volume with worse data quality, we would loose nothing in terms of force recovery when using the smaller volume. (Only

the halo scale length might not be as well constrained, see Figure 8).

Now let's discuss the reason for the biases that we observe [TO DO: Jo writes: "Too colloquial."]. The peak of the distribution in  $\Delta F_R(*_i)$  and  $\Delta F_R(g_j)$  is slightly biased towards an underestimation of  $|F_{R,M}|$  in our *RoadMapping* models. We believe the explanation for this to be the following: Spiral arms are very thin. If a spiral arm crosses the observation volume both its leading side (at large radii) and its trailing side (at small radii) are also in the volume. Stars in the trailing side feel a lower gravitational pull towards the galaxy center than they would if there was no spiral arm. Because there are in general more stars at smaller radii, the *RoadMapping* fit is slightly biased to reproduce in general slightly weaker radial forces (see also Figures ?? and ??). [TO DO: Mention that this condition was not the case for the 500pc8Void and 1kpc8Void volumes, in which we had difficulties modelling the distribution.]

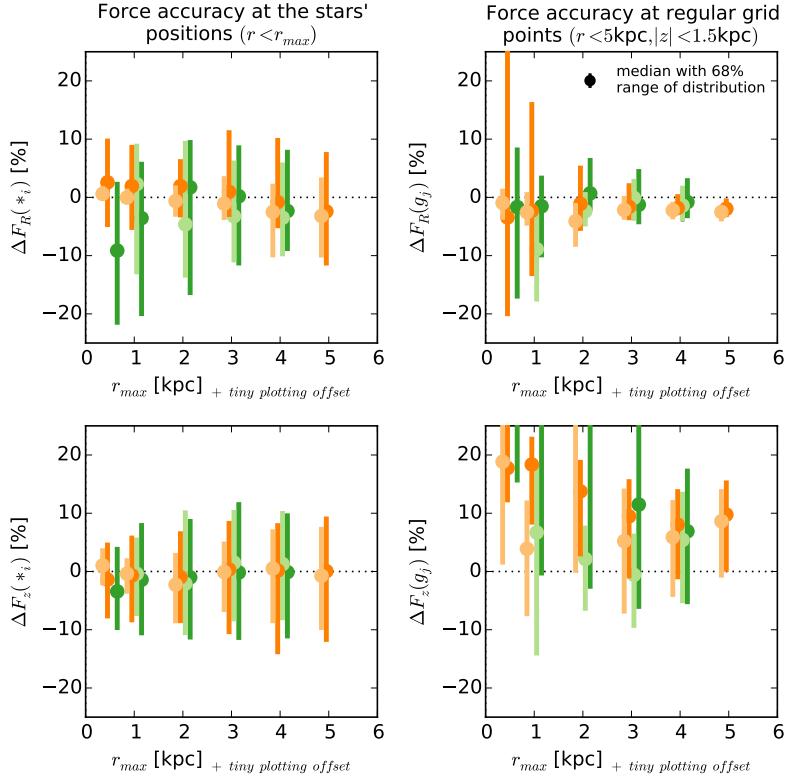
The peak of  $\Delta F_z(*_i)$  is approximately at 0, while the peak of  $\Delta F_z(g_j)$  is strongly biased towards an overestimation of  $|F_{z,M}|$ . There are much more stars in the spiral arms than in the inter-arm regions, and the stars in the spiral arm feel stronger vertical forces because of the higher disk mass. *RoadMapping* finds a model that in general has much stronger vertical forces than expected for a smooth potential. While the actual vertical forces that the many stars in the spiral arms feel are very well recovered, it becomes obvious when looking at the grid points regularly distributed in space, that the *RoadMapping* vertical forces are much stronger. As expected the overestimation is especially strong ( $\sim 20\%$ ) for small survey volumes dominated by spiral arms, while small volumes dominated by an inter-arm region result in much better estimates for the spatially averaged  $F_z(g_j)$  ( $\sim 5\%$  bias). Large volumes lie somewhere in between (bias of  $\sim 10\%$ ).

[TO DO: Jo writes: "Maybe comment, here or later, that this further shows that the procedure of BR13 was good, i.e., constrain  $F_z$  at one radius that corresponds to typical radius."]

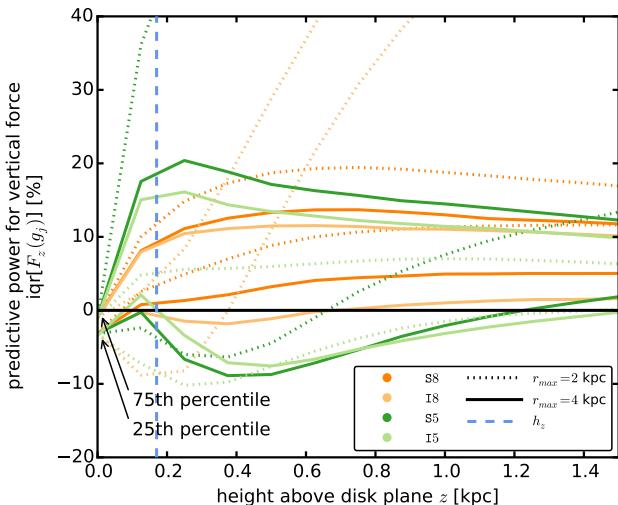
Because the stellar number asymmetry in the trailing vs. leading sides of spiral arms is much smaller than the stellar number asymmetry in spiral arm vs. inter-arm region, the bias is visible in the distribution of  $\Delta F_{R,*_i}$  (the  $F_R$  recovery is biased only by a few stars which leads to a bias which is visible for the majority of stars) and not in  $\Delta F_z(*_i)$  (the majority of stars biases the fit and we therefore recover  $F_z$  also for the majority of stars), but becomes really pronounced for  $\Delta F_z g_j$  (the inter-arm regions dominate when averaging spatially which leads to a large overestimation of  $F_z$  spatially) and stays small for  $\Delta F_R g_j$  (trailing and leading sides of spiral arms are similarly important when averaging spatially so it becomes visible that the bias is actually not that big).

[TO DO: Jo writes: "I do think it would be good to look a little more deeply in why the predictive  $F_Z$  fails so much by looking at where most of the difference is (close to the plane or far away from it)." —> TO DO: Include figure Fg vs z and discuss it in text.]

#### 4.2.4. Influence of spiral arms on the recovery of the gravitational forces



**Figure 9.** Accuracy of the radial (upper panels) and vertical (lower panels) gravitational forces recovered with *RoadMapping* from the suite of data sets introduced in Section 4.2.1, depending on the size  $r_{\max}$  of the survey volume. The left panels investigate how well *RoadMapping* recovers the forces for the ensemble of stars in the data set ( $\Delta F_{R/z}(*_i)$  in Equations (19)-(20)) and the right panels test the predictive power of each of the data sets' best-fit potential by showing how well the forces are recovered at regular grid points in a large cylinder of  $r_{\max} = 5$  kpc and height  $|z| = 1.5$  kpc around the survey volumes center ( $\Delta F_{R/z}(g_j)$  in Equations (21)-(22)). For each distribution of  $\Delta F_{R/z}(*_i)$  and  $\Delta F_{R/z}(g_j)$  we show here the median as dot with the [16th,84th] percentile range as bar. We find that the forces are very well recovered at the positions of the stars independent of the size of the volume. The spiral arms introduce some biases in the overall recovered potential and we need at least a survey volume of  $r_{\max} = 3$  kpc to get a potential with a reasonable predictive power. [TO DO: Make sure all final analyses are in this plot.] [TO DO: Add legend to plot.] [TO DO: Jo suggested to make the right lower plot for slices at different heights, because it is very likely that most of the misjudgement is at high z. Alternatively we could use a different reference volume that's not going up to 1.5kpc but maybe only up to 3-4 scale heights?] [TO DO: Recalculate statistics with [16,84] percentile range.] [TO DO: Jo writes: "Yes!" → Different volume or showing the plot for different slices?]



**Figure 10.** [TO DO] [TO DO: Jo writes: "I don't like boxes around legends"]

Figure 9 in Section 4.2.3 gave an overview of the accuracy of the gravitational forces recovery for our whole

suite of data sets depending on the size of the different survey volumes. In the following we want to relate the same quantities, i.e., local force recovery at the positions of the stars that entered the analysis (see Equations (19)-(20)) and power of a recovered model to predict the forces in a fixed large volume (see Equations (21)-(22)), to the actual strength and dominance of spiral arms in the respective survey volumes (see Section 2.4). In Figure 11 we find a connection between the local force recovery and the relative spiral contrast, and Figure 12 shows a trend that the predictive power of the recovered model depends on the dominance of a spiral arm in the survey volume. The former was to be expected, but the latter is a very interesting result. [TO DO: Jo writes: "Why surprising?"]

Let's have a closer look [TO DO: Jo writes: "Too colloquial!"] at Figure 11, where we relate Equations (19)-(20) (see also Figure 9, left panels) to Equation (6) (see also Figure 2, right panel). The average fraction of stars for which the recovery of radial or vertical force is bad (i.e., larger than 10%) decreases with decreasing spiral contrast  $\sigma_\kappa \equiv \sigma[\kappa(r \leq r_{\max})]$  and is overall very similar (see the linear fits in Figure 11) for a given  $\sigma_\kappa$ , but there is scatter when considering individual analyses, which is

stronger for the radial forces. The overall trend is however obvious and as expected: Volumes, in which the steep surface density drop around a strong spiral arm ( $\kappa_j \neq 0$ ) is not balanced by larger areas with less perturbations ( $\kappa_j \sim 0$ ), have both a large relative spiral contrast  $\sigma_\kappa$  and a large relative number of stars affected by the non-axisymmetric kinematics of the spiral arms, for which the axisymmetric *RoadMapping* model cannot properly predict the forces. We see again the trend that volumes at  $R_0 = 5$  kpc have a stronger spiral contrast and a worse force recovery than the volumes centered at  $R_0 = 8$  kpc. Volumes centered on inter-arm regions have, as expected, slightly lower  $\sigma_\kappa$  than volumes centered on spiral arms and therefore a better local forces recovery.

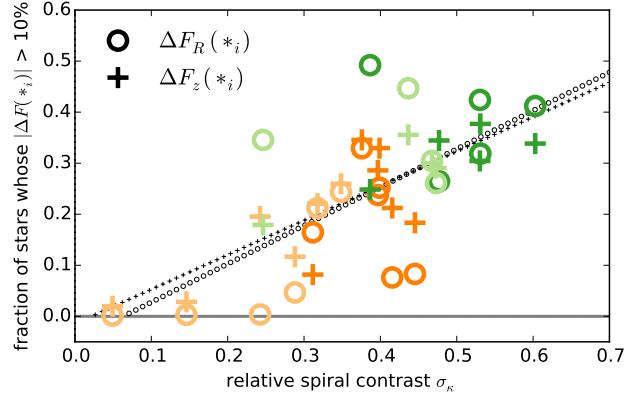
[TO DO: So far the reference cylinder is actually a reference sphere, cut at  $z=-1.5, 1.5$ . Redo with proper cylinder.]

The result in Figure 12 is especially interesting. Here we relate the volume averaged predictive power ((21)-(22) (see also Figure 9, right panels) to the dominance of the spiral arm in Equation (5) (see also Figure 2, third panel). Firstly we note that  $F_z(g_j)$  is recovered much worse than  $F_R(g_j)$  and the reason is, as laid out in Section 4.2.3), that the reference cylinder contains some regions at high  $|z|$  that are not probed by stars. The main result of this figure is the following: The less a volume is dominated by a spiral arm (smaller  $\langle \kappa \rangle$ ), the better the model recovered from the corresponding volume can predict the forces in a large cylinder of  $r_{\max} = 5$  kpc around the volumes midpoint. For large volumes where spiral arms and inter-arm regions average out ( $\langle \kappa \rangle \sim 0$ ) this is true and as expected. For smaller volumes the picture is different: Small volumes dominated by spiral arms ( $\langle \kappa \rangle >> 0$ ) contain a large fraction of stars in spiral arms that bias the potential estimates strongly. The line fitted to the S8 and S5 analyses in Figure 12 suggests a relation between the spiral arm dominance and the predictive power of the analysis. For small volumes dominated by inter-arm regions ( $\langle \kappa \rangle < -0.1$ ) there is no such clear relation. As long as the inter-arm region is not unusually depleted of stars and filling the whole volume (as in the case of the I5 analysis at  $\langle \kappa \rangle \sim -0.5$  with  $r_{\max} = 1$  kpc), all volumes centered on inter-arm regions have a comparable or even better predictive power than the analyses at  $\langle \kappa \rangle \sim 0$ . It appears that between spiral arms the distribution is smooth and close enough to the axisymmetric average model, such that the potentials recovered from these volumes have real predictive power for much a much larger volume. This is an important result.

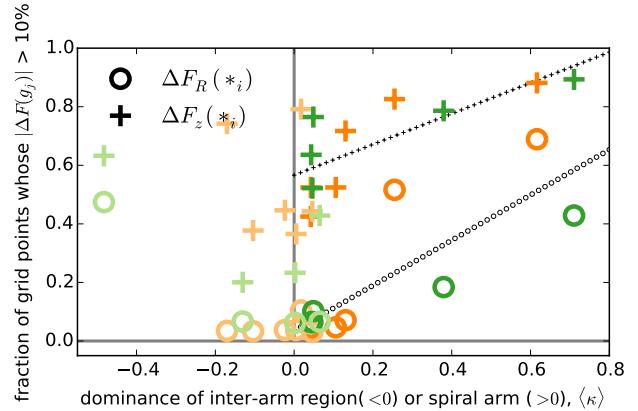
[TO DO: Jo writes: "okay, I see now why it is surprising, but maybe only mention that this is interesting here."]

#### 4.2.5. Circular velocity curves and surface density profiles

In this section we will illustrate even further that even though the recovered potential parameters are very different for different survey volumes, the actual potential parameter recovery is very good—and not just the gravitational forces on average. In Appendix A in Figures A1-A6 we show the circular velocity curves and surface density profiles within two times the disk scale height ( $|z| \leq 2 \times h_z = 0.34$  kpc) for all data sets in this work and compare them with the actual matter distribution in the angular wedge  $\phi_0 \pm \arcsin(r_{\max}/R_0)$  containing the survey volume. We also mark the approximate radial



**Figure 11.** Influence of the spiral arm contrast within the survey volume on the recovery of the gravitational forces at the positions of the stars that entered the *RoadMapping* analysis. The relative spiral contrast on the  $x$ -axis was quantified as  $\sigma_\kappa \equiv \sigma[\kappa(r \leq r_{\max})]$  according to Equation (6) in Section 2.4. On the  $y$ -axis the fraction of stars is shown for which the radial (circle symbols) and vertical (cross symbols) force residual calculated from Equations (19)-(20) is larger than 10% (i.e., at 0 all stars have good force measurements). The dotted lines (small circles and crosses) are linear fits to the radial and vertical force residual fraction, respectively, and are guides to the eye that show the clear and expected trend that in volumes with smaller spiral arm contrast, where comparably fewer stars are located in spiral arms, the axisymmetric best-fit model can recover the true gravitational forces also for more stars. [TO DO: Make sure all final analyses are in this plot.]



**Figure 12.** Influence of the spiral arm dominance within the survey volume on the predictive power of the models recovered from each survey volume. The spiral arm dominance on the  $x$ -axis was quantified as  $\langle \kappa \rangle \equiv \langle \kappa(r \leq r_{\max}) \rangle$  from Equation (5) in Section 2.4. The  $y$ -axis shows the fraction of a large and fixed cylindrical volume with  $r_{\max} = 5$  kpc and  $|z| \leq 1.5$  kpc centred around the respective survey volume positions given in Table 1 (i.e., fraction of regular grid points) with radial (circle symbols) and vertical (cross symbols) force residuals larger than 10%. This quantifies how well the model can predict the potential in a large portion of the galaxy and at 0 the predictive power is best. The dotted lines (small circles and crosses) are guides to the eyes fitted to the analyses dominated by spiral arms at positions S8 (dark orange) and S5 (dark green) in Table 1. This figure demonstrates that the predictive power of a model derived from a survey volume that is either large enough for spiral arms not to dominate the volume ( $\langle \kappa \rangle \sim 0$ , compare Figure 2), or that are derived from volumes centered on inter-arm regions ( $\langle \kappa \rangle \lesssim 0$ , volumes at I8, bright orange, and I5, bright green) are less biased by the kinematics in the spiral arms and have therefore a higher predictive power. (The only exception is the volume at I5 with  $r_{\max} = [TODO]$  and  $\langle \kappa \rangle \sim -0.5$  that is especially strongly dominated by a depleted inter-arm region.) [TO DO: Make sure all final analyses are in this plot.] [TO DO: correct legend, it is supposed to read  $\Delta F_R(g_j)$ .]

regions in which the stellar number density per radial bin  $\Delta R = 200$  pc is higher than average in each data set. Even though the curves vary extremely between the individual data sets, it becomes very obvious that it is indeed the regions in which the majority of stars is located that drives the *RoadMapping* fit. And that, no matter if this region is dominated by a spiral arm or inter-arm region and no matter if this region is only as small as  $r_{\max} = 500$  kpc, *RoadMapping* is indeed constraining the local potential where most of the stars of the data set are located. Also, the constraints are not only most accurate but also most precise in these regions.

Only in the two volumes with  $r_{\max} = [0.5, 1]$  kpc at position S8 *RoadMapping* has some difficulties fitting the circular velocity curve; the model expects a flat or falling rotation curve and is presented with a steeply rising rotation curve due to the spiral arm dominating the region. But given *RoadMapping* recovers a good average surface density profile and the circular velocity at least at the center of the small volume, the fit is still quite successful.

In an application to real data in the MW we would also have the possibility to impose some informative prior information on the potential shape (e.g., on the rotation curve), to avoid very unrealistic results (see also discussion in Section 5.4).

[TO DO: Jo writes: "It would be interesting to figure out a little more how we can determine whether we expect the fit to be strongly biased because we are using a volume that it sitting right on top of a massive spiral arm, although that is perhaps best kept for a later paper."]

## 5. DISCUSSION

### 5.1. On the informativeness of an orbit distribution function

The qDF is surprisingly informative; even though there was absolutely no indication beforehand that the qDF would be actually a good model for the simulation snapshot; and even though small survey volumes dominated by non-axisymmetric spiral arms are definitely not well-described by the qDF. It appears however that a potential model that does not fit the gravitational forces acting on the stars must lead to such an unrealistic orbit/action distribution, that a fit with even such a simple orbit DF as the the qDF is impossible. This demonstrates once more how powerful the concept of an orbit DF is.

### 5.2. On the restrictiveness of the parametrized potential model

We do, on the one hand, use a bulge and halo model that reproduce the true bulge and halo better than we can hope to use one in reality for the MW. The fact that for all except of the largest volumes the true halo scale length is not remotely recovered (and some small volumes even have  $f_{\text{halo}} = 0$ ), and that the contribution of the bulge to the overall potential is small (i.e., the bulge contribution to the total radial force at  $R = 8$  kpc is only  $\sim 9 - 10\%$ ) remedies this apparent advantage.

We use on the other hand a disk model, the Miyamoto-Nagai disk, that we chose purely for its convenient parametric form and of which we know that it is not a good model for the simulation snapshot, especially not for the

radial density profile. As we saw in most figures in this work this might lead to biases in predicting the potential at radii where we have only a few or no stars. But because the spiral arms are such strong perturbations in the overall potential, a better disk model would not give much better results.

It appears that a potential model with a reasonable shape and flexibility (here: disk+bulge+halo structure with 5 free parameters) can do well enough in finding a good fit, both locally for small volumes and overall for large volumes.

### 5.3. Precision of Gaia measurement errors

Considering measurement uncertainties of distances and proper motions, we found in Paper I that for a survey volume with  $r_{\max} = 3$  kpc, distance uncertainties of  $\pm 10\%$  and proper motion uncertainties of [TO DO] *RoadMapping* still gives unbiased parameter results. Even if the proper motion errors are not perfectly known.

At a distance of 3 kpc the measurement uncertainties of bright stars in Gaia lie even below these limits.

In paper I we focussed on recovering completely unbiased model parameters and found that *RoadMapping* is robust to moderate deviations of the model assumptions. In this work we released the condition that the model parameters itself had to be recovered accurately, but allowed *RoadMapping* to simply find an overall best fit for the data strongly affected by spiral arms—which was surprisingly successful in recovering the local potential even if the model parameters were not recovered.

We therefore presume that in reality we probably have an even larger margin of error before the measurement uncertainties mess up the constraints noticeably than we found in Paper I. And Gaia uncertainties are already below our formal error.

In addition, we found in this work that a volume of  $r_{\max} = 3$  kpc should be already big enough to find an overall best fit axisymmetric model for the galaxy. At larger distances dust starts affecting the measurements. And inside of  $R = 3 - 4$  kpc the stellar motions become increasingly non-axisymmetric, possibly because of the Galactic bar (Reid et al. 2014).

Overall we should therefore be very well off by applying *RoadMapping* to Gaia data within  $r_{\max} = 3$  kpc only.

### 5.4. Location of the Sun in the Milky Way

[TO DO: Jo writes: "Perhaps section 5.4 could be quite a bit shorter, because it is quite speculative and seems oddly detailed given that it's speculative."]

The Sun is located in one of the smaller spiral arms of the MW, the local Orion spur/arm (Morgan et al. 1953). The Orion arm has long been considered a minor arm within the MW (Blaauw 1985). Two of the MW's major spiral arms pass by the Sun within a few kpc: The Perseus arm is  $\sim 2$  kpc from the Sun (towards the outer MW) (Xu et al. 2006), and the Sagittarius arm at  $\sim 1$  kpc (Sato et al. 2010).

Although, recent observations of stronger star formation (SF) in the Orion arm (Xu et al. 2013) and lower SF in the Perseus arm (Zhang et al. 2013) suggest that the designation which arms in the MW are "major arms" could be slightly different than previously thought.

The Gaia DR1 in September 2016 will have parallax measurements better than 10% within a maximum range

of  $\sim 500$  pc from the Sun, the Gaia DR2 at the end of 2017 will cover 2 kpc with good parallaxes [TO DO: My only reference for this is HW. Some paper?].

How reliable *RoadMapping* results from the DR1 will be, depends on the strength of the local Orion arm, which will dominate the survey volume. If previous assessments that it is only a weak spiral arm are correct, we can hope to get already pretty reliable measurements for the overall MW potential. If it is a stronger spiral arm, we might run into some difficulties when modelling such a small volume dominated by a spiral arm (see Figure A1).

However, a recent measurement of the MW's rotation curve by Reid et al. (2014) confirms again that it is flat. We could impose this condition as a prior constraint in *RoadMapping* (or fix the rotation curve slope as Bovy & Rix (2013) did in their *RoadMapping* analysis), which should greatly help *RoadMapping* to find a good model even from a small volume on a strong spiral arm.

In the Gaia DR2 with a coverage of  $r_{\max} \sim 2$  kpc we would in addition to the Orion arm also have the Sagittarius and Perseus arms cross the survey volume. *RoadMapping* should do much better with this volume, because it should become easier to average over spiral arms and inter-arm regions if there are several of them.

The scatter and errors of the rotation curve measurements in Figure 4 by Reid et al. (2014) could hide spiral arm signatures as strong as those of the simulated spiral galaxy in this work. It is however more likely that the MW spiral arms are actually weaker than in this simulation: Reid et al. (2014) measured for example that typical peculiar non-circular motions in spiral arms were around  $10 \text{ km s}^{-1}$  (and around  $20 \text{ km s}^{-1}$  in the Perseus arm), while Figure 3(c) about the non-axisymmetric velocity distribution in our simulated galaxy suggests an excess of stars with radial velocities around  $50 \text{ km s}^{-1}$ .

Reid et al. (2014) also found that within  $R \sim 4$  kpc the peculiar motions become much bigger, which could be due to the Galactic bar that stirs up the stellar motions in the center. For *RoadMapping* a volume around the Sun of  $r_{\max} = 3 - 4$  kpc should however already give good results.

### 5.5. Recovering non-axisymmetric structures in the potential from modelling small volumes

As we saw in Section 4.2.5 and the Figures in Appendix A, *RoadMapping* makes a very good attempt at fitting the local surface density and circular velocity curve also for volumes as small as  $r_{\max} = 500$  pc or 1kpc. The modelling results might not always be informative about global properties of the potential (i.e., the recovered model does not have a halo, a too thick disk or a overall completely wrong rotation curve slope), but it does tell us about the local potential. It for example correctly predicts a higher surface density in survey volumes dominated by spiral arms, or correctly estimates the circular velocity curve at the median radial position of the stars.

Instead of discarding these globally wrong potential models from small volumes we could actually make use of this interesting property of *RoadMapping* modelling. First of all it implies that we can to a certain extent already trust modelling results from the Gaia DR1 with its small spatial coverage—in particular the gravitational

forces at the solar position. At later stages, where we have the full Gaia catalogue with its huge survey volume, we could split the data set not only into different MAPs analogous to Bovy & Rix (2013), but also into different spatial bins in the  $(x, y)$  plane of the MW and model each of the smaller volumes separately. This approach would probe the local potential only, even when using an axisymmetric potential model, and should be sensitive to the spiral arms. In this way it should be possible to build up a non-axisymmetric map of the MW potential—with very large spatial pixels however—with constraints from dynamical modelling only. It would be in particular interesting to compare and correlate these results from action-based dynamical modelling with results from dynamical Jeans models of the MW disk, stellar number counts and gas disk measurements to also learn more about the non-axisymmetric distribution of dark matter in the MW disk.

## 6. CONCLUSION

*RoadMapping* is a dynamical modelling machinery that simultaneously fits an action-based DF and axisymmetric potential model to the 6D phase-space coordinates of stellar populations in the MW disk and builds heavily on previous work by Binney & McMillan (2011); Binney (2012); Bovy (2015). In this work we have continued a throughout investigation of the capabilities of *RoadMapping* to recover the MW's gravitational potential. *RoadMapping* was first applied by Bovy & Rix (2013) and improved and tested in detail against the breakdowns of its modelling assumptions by Trick et al. (2016).

Here, we investigated the important aspect of modelling data affected by spiral arms in the galactic disk with the axisymmetric *RoadMapping* modelling. All similar studies using action-based modelling assume the influence of spiral arms not to be problematic, but here we test this assumption for the first time explicitly by modelling a simulated spiral galaxy by D'Onghia et al. (2013) (which has stronger spiral arms as we expect in the MW) and by comparing the modelling results with the true potential. It turns out, that *RoadMapping*-like action-based dynamical modelling is indeed very robust against perturbations of spiral arms, especially if the survey volume is large enough such that spiral-arms and inter-arm regions average out. In Section 4.1 we demonstrated this in detail for a single *RoadMapping* analysis of a data set with a spatial coverage of radius  $r_{\max} = 4$  kpc around the Sun.

We have investigated analyses in different survey volumes in the simulated galaxy in Section 4.2, with different sizes and at different positions with respect to the spiral arms and inter-arm regions. It turns out that independent of the size and position of the survey volume the gravitational forces are mostly well recovered at the locations of the stars that entered the analysis. Starting from survey volume sizes  $r_{\max} = 3$  kpc the recovered potential model becomes already a pretty good average potential model for the whole galaxy. For a convenient position, e.g., in a smooth and not too depleted inter-arm region, also smaller models can give good overall constraints. If a small volume is dominated by a very strong spiral arm the constraints become less reliable, as expected. The correct dark matter halo scale length was however only

recovered for a survey volume as large as  $r_{\max} = 5$  kpc. This overall robustness of *RoadMapping* is in particular notable, as the breakdown of the assumption of axisymmetry implies a breakdown of several model assumptions simultaneously: (i) orbital actions are not conserved anymore, which dilutes the whole idea behind action-based DF modelling somewhat, (ii) the axisymmetric potential model is not expected to be an optimal model for the galaxy, (iii) the quasi-isothermal DF is used for its functional simplicity and was not expected to describe the orbit distribution within spiral arms. However, the qDF seems to be informative enough to guide the fit to potential shapes that correctly measure the average surface density within  $\sim 2 \times$  the disk scale height and the circular velocity where most of the stars that entered the analysis are located—even for small volumes with  $r_{\max} = 500$  kpc dominated by spiral arms.

All these results make us very optimistic that *RoadMapping* will be well-suited to making new measurements of the MW's gravitational potential with the upcoming Gaia releases. It might work potentially even with the Gaia DR1, that only covers a small region  $r_{\max} = 500$  kpc (where the distance errors are small enough), because the local Orion arm, in which the Sun lies, is thought to be only a minor spiral arm in the MW and should not disturb the modelling too much.

## REFERENCES

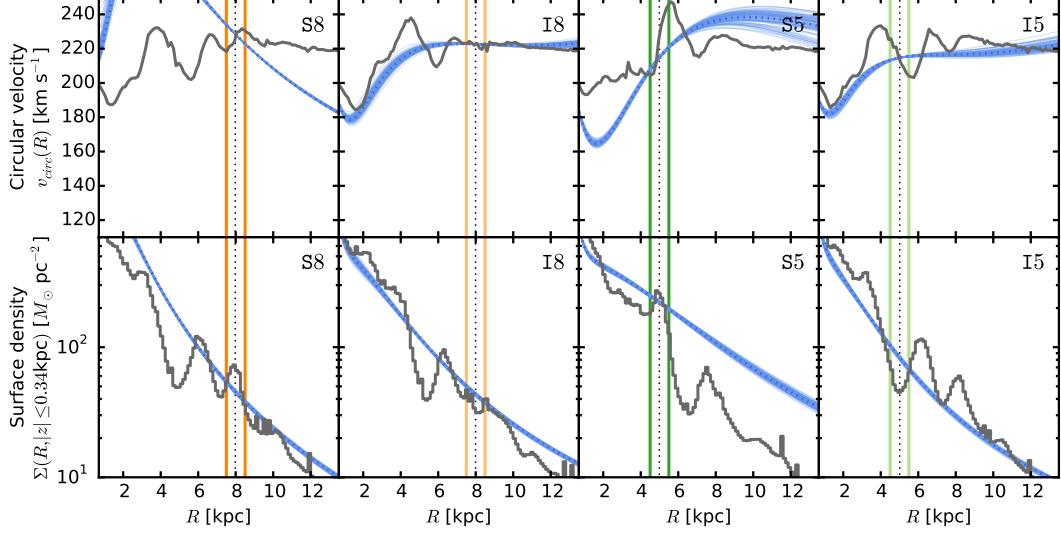
- Binney, J. 2010, MNRAS, 401, 2318  
 —. 2012, MNRAS, 426, 1324  
 Binney, J., & McMillan, P. 2011, MNRAS, 413, 1889  
 Bird, J. C., Kazantzidis, S., Weinberg, D. H., et al. 2013, ApJ, 773, 43  
 Blaauw, A. 1985, in IAU Symposium, Vol. 106, The Milky Way Galaxy, ed. H. van Woerden, R. J. Allen, & W. B. Burton, 335–341  
 Blitz, L., & Spergel, D. N. 1991, ApJ, 379, 631  
 Bovy, J. 2015, ApJS, 216, 29  
 Bovy, J., & Rix, H.-W. 2013, ApJ, 779, 115  
 Bovy, J., Rix, H.-W., & Hogg, D. W. 2012a, ApJ, 751, 131  
 Bovy, J., Rix, H.-W., Hogg, D. W., et al. 2012b, ApJ, 755, 115  
 Bovy, J., Rix, H.-W., Liu, C., et al. 2012c, ApJ, 753, 148
- Bovy, J., Rix, H.-W., Schlafly, E. F., et al. 2016, ApJ, 823, 30  
 Churchwell, E., Babler, B. L., Meade, M. R., et al. 2009, PASP, 121, 213  
 Das, P., & Binney, J. 2016, MNRAS, 460, 1725  
 D'Onghia, E., Vogelsberger, M., & Hernquist, L. 2013, ApJ, 766, 34  
 Eyer, L., Holl, B., Pourbaix, D., et al. 2013, Central European Astrophysical Bulletin, 37, 115  
 Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306  
 Georgelin, Y. M., & Georgelin, Y. P. 1976, A&A, 49, 57  
 Hammersley, P. L., Garzón, F., Mahoney, T. J., López-Corredoira, M., & Torres, M. A. P. 2000, MNRAS, 317, L45  
 Hernquist, L. 1990, ApJ, 356, 359  
 Jurić, M., Ivezić, Ž., Brooks, A., et al. 2008, ApJ, 673, 864  
 Liszt, H. S., & Burton, W. B. 1980, ApJ, 236, 779  
 Martig, M., Minchev, I., & Flynn, C. 2014, MNRAS, 442, 2474  
 Martig, M., Fouesneau, M., Rix, H.-W., et al. 2016, MNRAS, 456, 3655  
 Minchev, I., Chiappini, C., & Martig, M. 2014, A&A, 572, A92  
 Miyamoto, M., & Nagai, R. 1975, PASJ, 27, 533  
 Morgan, W. W., Whitford, A. E., & Code, A. D. 1953, ApJ, 118, 318  
 Ness, M., Hogg, D. W., Rix, H.-W., et al. 2016, ApJ, 823, 114  
 Newberg, H. J., Yanny, B., Rockosi, C., et al. 2002, ApJ, 569, 245  
 Piffl, T., Binney, J., McMillan, P. J., et al. 2014, MNRAS, 445, 3133  
 Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2014, ApJ, 783, 130  
 Sanders, J. L., & Binney, J. 2015, MNRAS, 449, 3479  
 —. 2016, MNRAS, 457, 2107  
 Sato, M., Hirota, T., Reid, M. J., et al. 2010, PASJ, 62, 287  
 Springel, V., Di Matteo, T., & Hernquist, L. 2005, MNRAS, 361, 776  
 Ting, Y.-S., Rix, H.-W., Bovy, J., & van de Ven, G. 2013, MNRAS, 434, 652  
 Trick, W. H., Bovy, J., & Rix, H.-W. 2016, ArXiv e-prints, arXiv:1605.08601  
 Vera-Ciro, C., & D'Onghia, E. 2016, ApJ, 824, 39  
 Wegg, C., & Gerhard, O. 2013, MNRAS, 435, 1874  
 Xu, Y., Newberg, H. J., Carlin, J. L., et al. 2015, ApJ, 801, 105  
 Xu, Y., Reid, M. J., Zheng, X. W., & Menten, K. M. 2006, Science, 311, 54  
 Xu, Y., Li, J. J., Reid, M. J., et al. 2013, ApJ, 769, 15  
 Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, AJ, 137, 4377  
 Zhang, B., Reid, M. J., Menten, K. M., et al. 2013, ApJ, 775, 79

## APPENDIX

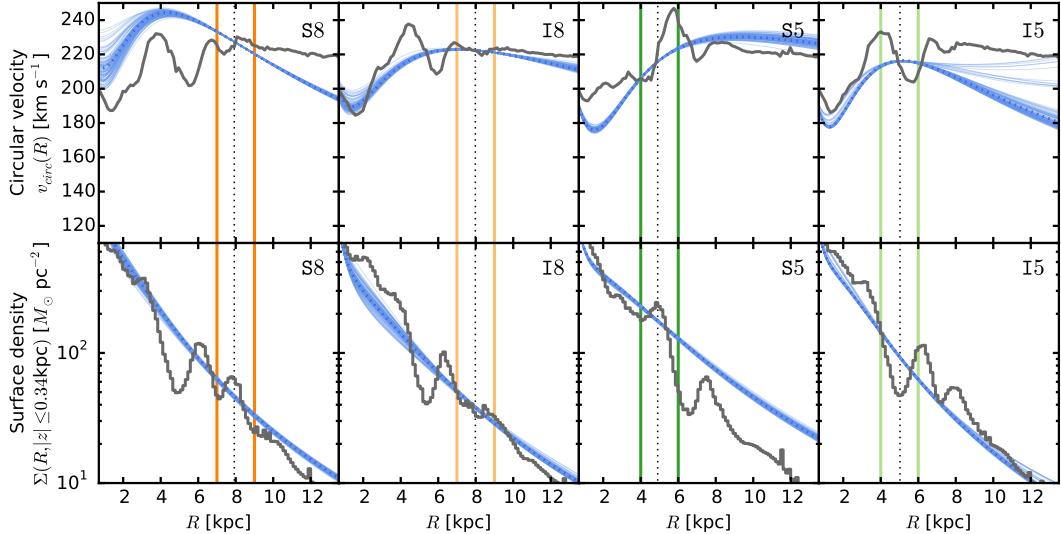
### A. CIRCULAR VELOCITY CURVES AND SURFACE DENSITY PROFILES FOR DIFFERENT SURVEY VOLUMES

[TO DO: Jo writes: "I'm also not sure that the appendix is really necessary, although there are few disadvantages to keeping it."]

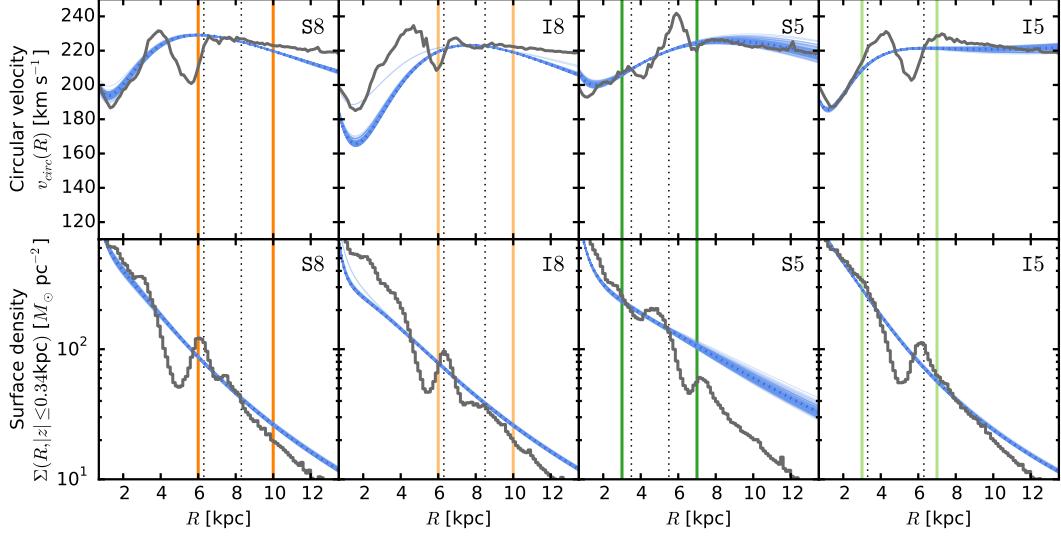
This appendix presents in Figures A1-A6 all circular velocity curves and surface density profiles that were recovered in this work from survey volumes of different sizes  $r_{\max} \in [0.5, 1, 2, 3, 4, 5]$  kpc and positions (see Table 1). The surface density profiles are calculated within twice the disk scale height  $|z| \leq 2 \times h_z = 0.34$  kpc, because this encloses most of the mass in the disk, most tracers and is therefore a height that should be well constrained by *RoadMapping*. [TO DO: Redo plots A3-A6 with a fixed radial bin size for the histograms.]



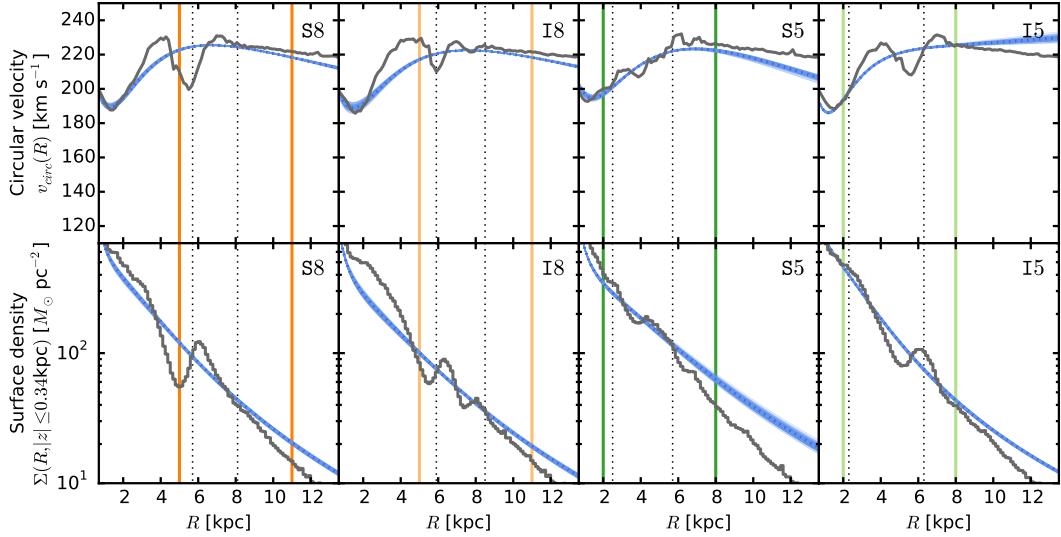
**Figure A1.** Comparison of the true local circular velocity curve and surface density within  $|z| \leq 2 \times h_z = 0.34$  kpc with the recovered *RoadMapping* models from survey volumes of size  $r_{\max} = 500$  pc. The grey curves show the profiles as derived from the galaxy simulation snapshot and averaged over an angular wedge  $\phi_0 \pm \arcsin(r_{\max}/R_0)$  that encloses the corresponding survey volume (see Table 1 for the corresponding  $R_0$  and  $\phi_0$  values, i.e., that part of the spiral galaxy that was probed by the data. The vertical orange and green lines mark the radial extent of the survey volume and the dotted vertical line marks the mean radius of all stars that entered the analysis. The blue lines finally show the *RoadMapping* potential models recovered from these stars; the blue dotted line corresponds to the model with the median potential parameters, and the thin blue lines correspond to 100 potential models drawn from the full posterior probability distribution sampled with the MCMC. It turns out that the constraints of highest accuracy and precision are always where most of the stars are located—within the survey volume and in particular at the peak of the distribution. [TO DO: Make sure all final analyses are in this plot.] [TO DO: *RoadMapping* analysis of the data set centered on I5 is still running.]



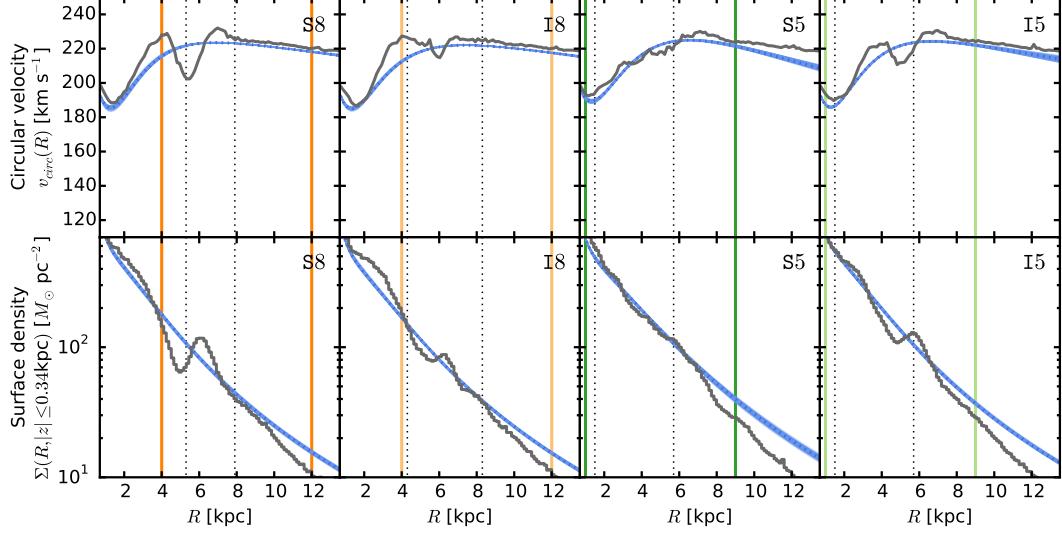
**Figure A2.** Same as Figure A1, but for all survey volumes with  $r_{\max} = 1$  kpc. [TO DO: Make sure all final analyses are in this plot.]



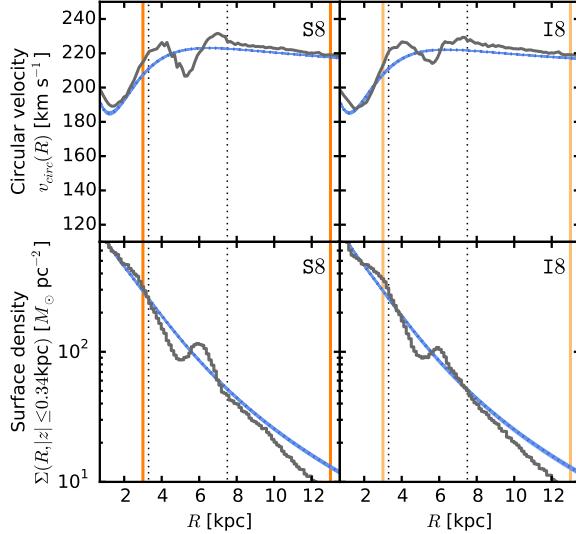
**Figure A3.** Same as Figure A1, but for all survey volumes with  $r_{\max} = 2$  kpc. Instead of marking the median of the radial stellar distribution, the dotted lines correspond here to the radial range within which the number of stars per radial bin  $\Delta R = 200$  pc is higher than average. This appears to be also the range within which the constraints are best. [TO DO: Make sure all final analyses are in this plot.]



**Figure A4.** Same as Figure A3, but for all survey volumes with  $r_{\max} = 3$  kpc. [TO DO: Make sure all final analyses are in this plot.]



**Figure A5.** Same as Figure A3, but for all survey volumes with  $r_{\max} = 4$  kpc. [TO DO: Make sure all final analyses are in this plot.]



**Figure A6.** Same as Figure A3, but for all survey volumes with  $r_{\max} = 5$  kpc. [TO DO: Make sure all final analyses are in this plot.]