

## ACTION-BASED DYNAMICAL MODELING 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 January 18, 2017*

### ABSTRACT

*RoadMapping* is a dynamical modeling machinery developed to constrain the Milky Way’s (MW) gravitational potential by simultaneously fitting an axisymmetric parametrized potential and an 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 modeling data drawn from an  $N$ -body simulation snapshot of a [disk-dominated galaxy with strong spiral arms \(but no bar\) of MW mass](#), exploring survey volumes with radii  $500 \text{ pc} \leq r_{\max} \leq 5 \text{ kpc}$ . The potential constraints are very robust, even though we use the simplest action-based DF, a 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, even for small volumes. For data from large survey volumes, *RoadMapping* finds axisymmetric models that average well over the spiral arms. Unsurprisingly, the models are slightly biased by the excess of stars in the spiral arms. Survey volumes with  $r_{\max} \geq 3 \text{ kpc}$  give predictions for an overall axisymmetric potential as good as those from larger volumes. However, a large radial survey extent,  $r_{\max} \sim 5 \text{ kpc}$ , is need to correctly recover the halo scale length. Potentials derived from data sets centered on inter-arm regions can be more reliably extrapolated than those centered on spiral arms. Our analysis implies that building axisymmetric models for the Galaxy with upcoming Gaia data will lead to sensible and robust approximations of the MW’s potential.

*Keywords:* Galaxy: disk — Galaxy: fundamental parameters — Galaxy: kinematics and dynamics — Galaxy: structure — galaxies: spiral

### 1. INTRODUCTION

An important basis for learning more about the Milky Way’s (MW) overall gravitational potential and orbit distribution function (DF) is to find the “best possible” axisymmetric model for the Galaxy. Given such a model the identification and characterization of non-axisymmetries like 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 the metallicity of each star, in addition to the distribution in orbit space.

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 application of *RoadMapping* was done by Bovy & Rix (2013). Trick et al. (2016), hereafter Paper I, subsequently performed a detailed analysis of the strengths and limitations of the approach. *RoadMapping* presumes that simple stellar populations in the MW disk—be it mono-abundance

populations (MAPs), i.e., stars with the same [Fe/H] and  $[\alpha/\text{Fe}]$  (Bovy et al. 2012a,c,b, 2016), 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). That MAPs are well described by the qDF was first shown by (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; see also Bovy (2014) for a general method to compute  $(\mathbf{x}, \mathbf{v}) \rightarrow (\mathbf{J}, \boldsymbol{\theta})$ ). Only if this assumed gravitational potential is close to the true potential, the action distribution of the stellar MAP in question will follow an orbit DF of qDF-shape. This is the idea on which *RoadMapping* builds, and which allows us 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). To avoid spiral arm effects, they did not use in-plane motions. Their potential model had only two free parameters (disk scale length and relative halo-to-disk contribution to the radial force at the solar 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  $R_s = 2.5 \text{ kpc}$  and was—in the regime  $R > 6.6 \text{ kpc}$ —later confirmed by Piffl et al. (2014) using a different action-based procedure.

<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

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, we found that *RoadMapping*’s modeling 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 that returns the correct forces, 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 modeling results.

The MW is, however, not axisymmetric. The bulge contains a strong bar (Liszt & Burton 1980; Blitz & Spergel 1991; Binney et al. 1991, 1997; Hammersley et al. 2000; Wegg & Gerhard 2013) and the disk itself is threaded by spiral arms (Oort et al. 1958; Georgelin & Georgelin 1976; Churchwell et al. 2009; Reid et al. 2014), 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. There is also kinematic evidence in the disk for moving groups (Dehnen 1998; Famaey et al. 2005; Bovy et al. 2009; Bovy & Hogg 2010)) and streaming motions (in 21 cm or velocities) (Bovy et al. 2015; Williams et al. 2013; Siebert et al. 2012), both of which are likely caused by non-axisymmetric perturbations to the gravitational potential.

As *RoadMapping* and related approaches can only build axisymmetric models, this is an important breakdown of modeling assumptions which was not investigated in Paper I. In this paper we want to understand in which respects *RoadMapping* will still give reliable constraints on the MW’s gravitational potential in the presence of spiral arms.

Our investigation makes use of an  $N$ -body simulation snapshot of a spiral galaxy with strong spiral arms presented in D’Onghia et al. (2013) (see their Figure 8, top left panel). From this snapshot we draw mock data in regions with different spiral arm strengths. We then apply the *RoadMapping* machinery to these data sets and test how well we recover the local and overall gravitational potential.

In Paper I we confirmed and tested separately the robustness of *RoadMapping* in the case that the data came from a different model family—for either the potential or DF—than assumed in the dynamical modeling. 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 potential and orbit DF model that we are using were picked as a pragmatic compromise between (i) being a reasonable choice given the initial axisymmetric set-up of the galaxy simulation, and (ii) because of their simplicity, computational advantages, and—in case of the qDF—because that’s what we are planning to use in the MW. Given that the simulation has also evolved away from its axisymmetric beginnings, we expect our chosen model to be reasonable, but not particularly well-suited to model this galaxy.

Spiral arms introduce another—but possibly minor—breakdown of the modeling assumptions: In a non-axisymmetric gravitational potential the three actions will not be strict integrals of motions anymore (Binney & Tremaine 2008; Minchev et al. 2011, 2012; Solway et al. 2012; Grand et al. 2012; Vera-Ciro & D’Onghia 2016). It will be interesting to see, if the action-based DF in *RoadMapping* modeling is still informative, even if it only uses the present actions estimated in an axisymmetric potential.

In the MW we expect the central bar to introduce additional non-axisymmetries in the Galactic disk—but presumably not stronger ones than the spiral arms. As the galaxy simulation in this work does not have a central bar, we do not investigate any bar effects here.

Though non-axisymmetry could be a severe problem for *RoadMapping*, we show in this paper that *RoadMapping* potential estimates are still surprisingly accurate, which makes us optimistic that they will also be so for the MW.

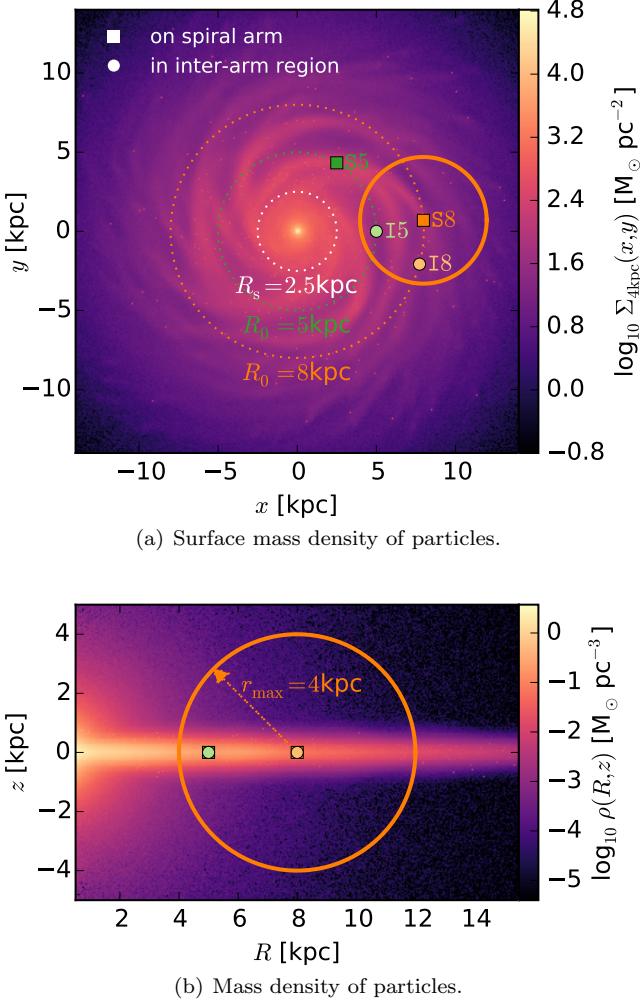
This paper is organized as follows. Section 2 describes the  $N$ -body simulation snapshot of a spiral galaxy of MW mass that we are going to model in this study, explains how we extract 6D stellar phase-space data from it, and how we quantify the spiral arm strength. Section 3 summarizes the *RoadMapping*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

## 2. DATA FROM A GALAXY SIMULATION

*RoadMapping* requires 6D phase-space coordinates  $(\mathbf{x}_i, \mathbf{v}_i)$  for a large set of stars that move independently in a collisionless galactic potential. If we want to test *RoadMapping* on a simulated galaxy, it is most convenient to apply it to an  $N$ -body simulation with a huge number of low-mass “star” particles. In that way, we can directly take the positions and the motions of individual particles as independent tracers of the potential, just as with the stars in the MW, without having to use an error-prone prescription to turn a single particle into many stars. The high-resolution simulations with its millions of particles by D’Onghia et al. (2013) satisfy this requirement.

### 2.1. Description of the galaxy simulation snapshot

The high-resolution  $N$ -body simulation snapshot of a disk galaxy by D’Onghia et al. (2013), which we use in this work, was carried out with the GADGET-3 code, and set up in the manner described in Springel et al. (2005). In this simulation overdensities with properties similar to giant molecular clouds induced four prominent spiral arms—and therefore a non-axisymmetric substructure—via the swing amplification mechanism. This galaxy simulation was also investigated by D’Onghia et al. (2013) (their Figure 8, top left panel) and D’Onghia (2015) (their Figure 2, top left panel). For details see



**Figure 1.** Simulation snapshot by D'Onghia et al. (2013). Shown are the surface mass density (in the  $(x, y)$ -plane, panel 1(a)) and mass density (in the  $(R, z)$ -plane, panel 1(b)) 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  $R_s = 2.5$  kpc (see Section 2.1) and the radii at which we center our test survey volumes in this investigation,  $R_0 = 8$  kpc and  $R_0 = 5$  kpc. The centers of the different survey volumes are marked with a square, if the survey volume is centered on a spiral arm (S8 and S5), or with a circle, if the volume is centered on an inter-arm region (I8 and I5). The orange circle with radius  $r_{\max} = 4$  kpc marks the survey volume in which we conduct the analysis discussed in detail in Section 4.1.

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

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

with total halo mass  $M_{\text{dm}} = 9.5 \cdot 10^{11} M_{\odot}$  and scale length  $a_{\text{dm}} = 29$  kpc.

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  $\sim 9.5 \cdot 10^5 M_{\odot}$ . The initial vertical mass distribution of the stars in the disk is specified by the profile of an isothermal sheet with a radially

**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.

constant scale height  $z_0$ , i.e.,

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

with total disk mass  $M_{*} = 0.04 \cdot M_{\text{dm}} = 3.8 \cdot 10^{10} M_{\odot}$ . The scale-length  $R_s$  is assumed to be 2.5 kpc and  $z_0 = 0.1 \cdot R_s$ . In this model the disk fraction within 2.2 scale lengths is 50% of the total mass, leading to a formation of approximately four arms (D'Onghia 2015).

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  $a_{\text{bulge}} = 0.1 \cdot R_s = 0.25$  kpc.

The initial velocity setup of the “disk star” particles assumes for simplicity Gaussian velocity dispersion profiles (Springel et al. 2005).

The simulation snapshot which we are using in this work has evolved in isolation for  $\sim 250$  Myr, which corresponds to approximately one orbital period at  $R \sim 8$  kpc. The mass density of simulation particles (without the DM halo) at this snapshot time is shown in Figure 1. The “molecular cloud perturbers”, which caused the formation of the four pronounced spiral arms, can be seen in Figure 1 as small overdensities in the disk. The spherical bulge and very flattened disk are shown in Figure 1(b).

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

## 2.2. Data selection and survey volume

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 analyzed in this study. For each one, we selected a volume with  $r_{\max} = [0.5, 1, 2, 3, 4, 5]$  kpc centered on a spiral arm (S) and on an inter-arm region (I) at both the equivalent of the solar radius,  $R_0 = 8$  kpc, and at  $R_0 = 5$  kpc, where the disk strongly dominates (see Figure 2), and the spiral arms are more pronounced than at  $R_0 = 8$  kpc (see Figure 3). The exact positions of the vantage points S8, I8, S5, and I5 are summarized in Table 1.

From within each volume we drew  $N_{*} = 20,000$  random “disk star” particles, and used their phase-space positions  $(\mathbf{x}_i, \mathbf{v}_i)$  within the simulated galaxy's rest-frame

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* modeling 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 significantly disturb the modeling. Secondly, in this study we want to isolate and investigate the deviations of the data from axisymmetry and the assumed potential and DF model independently of other effects.

### 2.3. True symmetrized potential

For a galaxy with pronounced spiral arms, an axisymmetric model matter distribution *per se* cannot reproduce the true matter distribution globally. We therefore obtain 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) 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 Hernquist profiles by construction (see Section 2.1).

The disk in this simulation snapshot deviates from its initial conditions in Equation (2): after 250 Myr pronounced spiral arms have formed, also causing some in-plane heating. Except of an overdensity around  $R \sim 6$  kpc (see Figure 6(d) in Section 4.1.2), the overall radial surface density profile (i.e., the azimuthal average) did not change by much, and appears smooth and exponential. We therefore chose a double exponential disk model to fit the particle distribution in the disk. The fit assumes the total disk mass to be known and be equal to the total mass of all disk particles. The best fit double exponential disk profile is found by maximizing the likelihood for all disk particles to be drawn from this axisymmetric density profile. In this way the fit does not depend on binning choices and is driven by the number and location of the stars—analogous to our *RoadMapping* procedure.

The best fit parameters for this reference potential, to which we will refer as the DEHH-Pot (Double-Exponential disk + Hernquist halo + Hernquist bulge) in the remainder of this work, are given in Table 2. As can be seen in Figure 6 in Section 4.1.2 below, the DEHH-Pot fits the overall true density distribution very well, with the exception of  $z \sim 0$  where the particle distribution is not as cuspy as the exponential disk.

Figure 2 shows the circular velocity curve of the DEHH-Pot, and its decomposition into disk, halo, and bulge contribution. The disk clearly dominates between  $R \sim 2$  kpc and  $R \sim 7$  kpc.

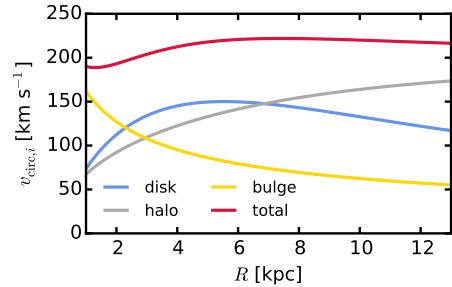
### 2.4. Quantifying the strength of spiral arms

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

**Table 2**  
Best fit parameters of the DEHH-Pot.

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 DEHH-Pot is introduced in Section 2.3, and we use it as the global best fit symmetrized potential model for the simulated galaxy. 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 (18) and (19).



**Figure 2.** Circular velocity curve of the DEHH-Pot, i.e., the symmetrized best fit to the  $N$ -body simulation, and its disk, halo and bulge components. To rotational support at 2.2 scale lengths is  $(v_{\text{circ},\text{disk}}/v_{\text{circ},\text{total}})^2 \sim 47\%$ . This demonstrates that the simulation is a disk-dominated spiral galaxy. [Comment to the referee: This is a new figure.]

the quantity

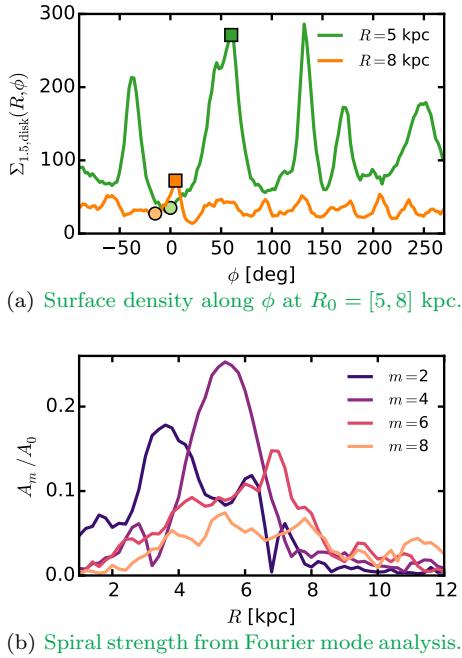
$$\Delta_{\text{Spiral}}(x_k, y_k) \equiv \frac{\Sigma_{1.5\text{kpc},\text{disk},T}(x_k, y_k)}{\Sigma_{1.5\text{kpc},\text{disk},S}(x_k, y_k)} - 1 \quad (3)$$

where  $\Sigma_{1.5\text{kpc},\text{disk},\alpha}$  is the true surface density of the disk component of the simulation snapshot ( $\alpha = T$  for “true”), or of the symmetrized snapshot model DEHH-Pot in Section 2.3 ( $\alpha = S$  for “symmetrized”),

$$\Sigma_{1.5\text{kpc},\text{disk},\alpha}(x_k, y_k) \equiv \int_{-1.5 \text{ kpc}}^{1.5 \text{ kpc}} \rho_{\text{disk},\alpha}(x_k, y_k, z) \, dz. \quad (4)$$

$(x_k, y_k)$  are the coordinates of regular grid points with spacing  $\delta = 0.25$  kpc.<sup>6</sup>  $(x_c = R_{0,c} \cdot \cos \phi_{0,c}, y_c = R_{0,c} \cdot \sin \phi_{0,c}, z_c = 0)$  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  $n \simeq \pi r_{\text{max}}^2 / 4\delta^2$  values of  $\Delta_{\text{Spiral}}(x_k, y_k)$  inside a given survey volume of radius  $r_{\text{max}}$  around position  $c$  and calculate

<sup>6</sup> We average the particle surface density of the true simulation potential over area element sizes of  $\delta \times \delta$  around  $(x_k, y_k)$ , when calculating  $\Sigma_{1.5\text{kpc},\text{disk},T}(x_k, y_k)$ .



**Figure 3.** Demonstrating the spiral arm strength at different radii. Panel 3(a) shows the surface density along the azimuth angle at the radii  $R_0$  on which we center our survey volumes. The also mark the corresponding  $\phi_{R_0}$  from Table 1. The difference between the surface density at S5 (■) and I5 (●) is 200% of the mean surface density at  $R_0$ ; for S8 (□) and I8 (○) the difference is 130%. Panel 3(b) shows the Fourier model amplitudes for  $m = 2, 4, 6, 8$  calculated as  $A_m / A_0 = |\sum_i M_i \exp(im\phi_i)| / \sum_i M_i$  for all disk particles at a given radius with mass  $M_i$  and azimuth position  $\phi_i$ . As can be seen, the simulation has overall four strong spiral arms, dominating between  $R = 4 \text{ kpc}$  and  $R = 7 \text{ kpc}$ . Inside of that there are two, and outside of that six or more arms. As the particle density increases with smaller radius, most tracers in the analysis will come from regions with only a few strong spiral arms. [Comment to the referee: This is a new figure.]

the mean and standard deviation,

$$\begin{aligned} \langle \Delta_{\text{Spiral}} \rangle &\equiv \langle \Delta_{\text{Spiral}}(r \leq r_{\max} | c) \rangle \\ &\equiv \frac{1}{n} \sum_{k=1}^n \Delta_{\text{Spiral}}(x_k, y_k) \quad (5) \\ \sigma_{\Delta_{\text{Spiral}}} &\equiv \sigma[\Delta_{\text{Spiral}}(r \leq r_{\max} | c)] \\ &\equiv \sqrt{\frac{1}{n} \sum_{k=1}^n [\Delta_{\text{Spiral}}(x_k, y_k) - \langle \Delta_{\text{Spiral}} \rangle]^2} \quad (6) \end{aligned}$$

$$\text{with } (x_k - x_c)^2 + (y_k - y_c)^2 \leq r_{\max}^2. \quad (7)$$

These quantities tell us if and how much a spiral arm or an inter-arm region dominates the survey volume ( $\langle \Delta_{\text{Spiral}} \rangle > 0$  for spiral arms,  $\langle \Delta_{\text{Spiral}} \rangle < 0$  for inter-arm regions) and how large the relative contrast between spiral arms and inter-arm regions is ( $\sigma_{\Delta_{\text{Spiral}}}$ ). For example, volumes will have a smaller relative spiral contrast  $\sigma_{\Delta_{\text{Spiral}}}$ , if they are either small and sitting completely within an inter-arm region, or if they are large volumes that contain—in addition to some spiral arms and depleted inter-arm regions—large areas of unperturbed disk.

Figure 4 shows  $\Delta_{\text{Spiral}}$  as function of  $(x_k, y_k)$ , a histogram over all  $\Delta_{\text{Spiral},k}$  within the galaxy, and  $\langle \Delta_{\text{Spiral}} \rangle$  and  $\sigma_{\Delta_{\text{Spiral}}}$  calculated for all test survey volumes in this

work (see Section 2.2), depending on position and size.<sup>7</sup>

### 2.5. Comparison of the $N$ -body simulation to the MW

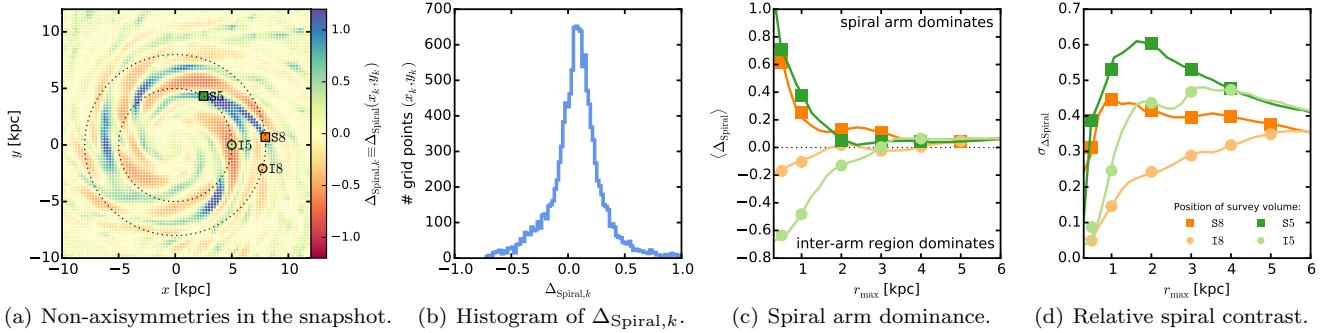
The  $N$ -body simulation in this work is not a perfect match to the MW. But to be suitable for this study it is most important that it satisfies our main requirements: It is a disk-dominated spiral galaxy with a very high number of star particles and strong spiral arms. In the following we discuss the differences between the simulation at hand and what we know about the MW.

**Bar.**— The MW has a central bar; our simulation does not. The Galactic bar can introduce non-axisymmetries around the co-rotation and Lindblad resonances (Dehnen 2000) [TO DO: more reference]. Bar effects are consequently not part of this study and remain a possible source of uncertainty in *RoadMapping*. We suspect however that its influence is small outside of the bar region.

**Mass components.**— The MW’s dark halo is estimated to have a mass  $M_{\text{dm},200} \sim 10^{12} \text{ M}_\odot$  (Bland-Hawthorn & Gerhard 2016); the simulation’s halo is slightly less massive with  $M_{\text{dm},200,\text{MW}} \sim 8 \cdot 10^{10} \text{ M}_\odot$  [TO DO: Check]. The stellar bulge mass of  $M_{\text{bulge}} = 9.5 \cdot 10^9 \text{ M}_\odot$  in this simulation is a bit smaller than the estimated stellar mass of the MW bulge with  $M_{\text{bulge,MW}} \sim 1.4 - 1.7 \cdot 10^{10} \text{ M}_\odot$  (Portail et al. 2015). The total stellar mass of the MW is estimated to be somewhere between  $M_{\text{stars,MW}} \sim 4.7 - 5.7 \cdot 10^{10} \text{ M}_\odot$  [TO DO: references, gelesen in B-H & G 2016], consistent with the total baryonic mass in the simulation of  $M_{\text{stars}} = 4.75 \cdot 10^{10} \text{ M}_\odot$ . The fraction of bulge mass to total stellar mass of the MW is  $M_{\text{bulge,MW}} / M_{\text{stars,MW}} \sim 0.3 \pm 0.06$ ; in our simulation it is 0.2.

**Disk.**— The best estimate for the MW’s thin disk scale length is  $h_r = 2.6 \pm 0.5 \text{ kpc}$  (Bland-Hawthorn & Gerhard 2016). The disk of this  $N$ -body simulation has a similar scale length,  $h_r = 2.5 \text{ kpc}$ . The stellar disk is however thinner than in the MW: scale height  $h_z = 220 - 450 \text{ pc}$  in the MW (Bland-Hawthorn & Gerhard 2016) versus  $h_z = 170 \text{ pc}$  in this simulation. Also, the MW has additional disk components, a gas disk and a stellar thick disk, which are not present in the simulation. This discrepancy does, however, not affect the objective of this study. The thick disk has a much higher velocity dispersion and is less prone to spiral perturbations. When we will apply *RoadMapping* to real MW data, the gas disk can be included as an additional component in the mass model, analogous to Bovy & Rix (2013), and if needed also the thick disk. By slicing data according to MAPs, the thick disk will be accounted for implicitly in the tracer selection.

<sup>7</sup> When considering the whole galaxy or a large survey volume,  $\langle \Delta_{\text{Spiral}} \rangle$  in Figures 4(b) and 4(c) is not exactly at 0, but slightly larger ( $< 0.05$ ). We account this small bias to the different functional forms of the DEHH-Pot and the initial axisymmetric disk in the simulation, Equation (2). This bias will however not affect our results.



**Figure 4.** Contrast and dominance of the spiral arms. Panel 4(a) shows the local spiral strength  $\Delta_{\text{Spiral},k}$  (calculated according to Equation (3) as described in Section 2.4) at regular grid points  $(x_k, y_k)$  with bin width 0.25 kpc. Marked are the centroids of the four test survey volumes of this study analogous to Figure 1. The histogram in panel 4(b) demonstrates the number of different  $\Delta_{\text{Spiral}}$  values in the region  $x, y \in [-14, 14]$  kpc. The panels 4(c) and 4(d) then show the dominance and relative contrast of spiral arms and inter-arm regions within each survey volume, depending on the volumes’ size,  $r_{\max}$ , and position (color-coded according to panel 4(a)). As measure for the spiral dominance we use the mean  $\langle \Delta_{\text{Spiral}}(r \leq r_{\max}) \rangle$ , and for the relative spiral contrast the standard deviation  $\sigma[\Delta_{\text{Spiral}}(r \leq r_{\max})]$ , calculated on the basis of all  $\Delta_{\text{Spiral},k}$  measurements within the given survey volume. 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  $R_0 = 5$  kpc than at  $R_0 = 8$  kpc. Also, inter-arm regions appear larger and smoother than spiral arms, as already inside a small volume centered 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.

*Disk fraction.*— The strength of the perturbations in the disk depends on the disk fraction (e.g., D’Onghia 2015). The total disk mass in the simulation might be only 4% of the halo mass, but within  $2.2R_s$  it is already 50% of the total mass. It is still under debate if the MW disk is maximal (i.e., rotational support of the disk at  $2.2R_s$  is  $\sim 55 - 90\%$ ; Sackett 1997; Binney & Tremaine 2008, §6.3.3). Bovy & Rix (2013) found, for example, a maximum disk with rotational support  $(v_{\text{circ,disk}}/v_{\text{circ,total}})^2 = (69 \pm 6)\%$  at  $2.2R_s$  with  $R_s = 2.15 \pm 0.14$  kpc. The DEHH-Pot (and therefore our  $N$ -body model) has  $(v_{\text{circ,disk}}/v_{\text{circ,total}})^2 = 47\%$  at  $2.2R_s$  and is therefore slightly sub-maximal (see Fig. 2). The decomposed rotation curve in Figure 2 is qualitatively similar to the MW model MI by (Barros et al. 2016) (their Figure 5, left panels). Their disk dominates between  $R \sim 5 - 9$  kpc, while our simulation’s disk dominates between  $R \sim 2 - 7$  kpc. We account for this difference by also drawing mock data sets from regions around  $R_0 = 5$  kpc (see Section 2.2).

*Number of spiral arms.*— The exact number of spiral arms in the MW is still under debate. The distribution of star-forming regions traced, e.g., by HII regions (Georgelin & Georgelin 1976), maser sources (Reid et al. 2009, 2014), and young massive stars (Urquhart et al. 2014), suggest that the MW has four major spiral arms (see also Vallée (2008, 2014) and references therein). Observations in the infrared, e.g., of old red-clump giant stars in the Spitzer/GLIMPSE survey (Churchwell et al. 2009) or of stellar NIR-emission by the COBE satellite (Drimmel & Spergel 2001) indicate that the MW is a grand-design two-armed spiral. [TO DO: More references???] Our simulation is overall a four-armed spiral galaxy, with the mode  $m = 4$  dominating between  $R = 4 - 7$  kpc,  $m = 2$  at smaller and  $m \geq 6$  at larger radii (see Figure 3(b)).

*Strength of spiral arms.*— As demonstrated in Figure 3(a)

the spiral arms introduce strong peak-to-peak differences in the stellar surface density (e.g.,  $\sim 200\%$  at  $R = 5$  kpc). The disk dominates inside  $R = 8$  kpc (see Figure 2), and in Section 4.1.2, Figure 7, we will see that the spiral arms introduce relative perturbations in the total gravitational forces of up to 30%. Figure 5(c), which we will discuss in Section 4.1.1, suggests an excess of stars with radial velocities up to  $50 \text{ km s}^{-1}$  in our simulated galaxy as compared to an axisymmetric model. This can be compared to Reid et al. (2014), who measured, for example, that typical peculiar non-circular motions in the MW spiral arms were around  $10 - 20 \text{ km s}^{-1}$  for  $R \gtrsim 4$  kpc. We therefore expect the strength of the perturbation to the total potential due to the spiral arms in this simulation—especially inside  $R = 8$  kpc where most of our mock data is drawn from—to be similar or even stronger than those in the MW.

### 3. ROADMAPPING MODELING

In this section we summarize the mathematical ingredients of *RoadMapping*, and motivate the DF and potential model that we are going to fit to the data. *RoadMapping* makes extensive use of the `galpy` python library by Bovy (2015)<sup>8</sup>. For full details on the *RoadMapping* machinery see also Paper I.

#### 3.1. Likelihood

As already laid out in Section 2.2, we use as data the 6D  $(\mathbf{x}_i, \mathbf{v}_i)$  coordinates of  $N_*$  stars within a spherical survey volume. The corresponding, purely spatial selection function  $\text{SF}(\mathbf{x})$  is

$$\text{SF}(\mathbf{x}) \equiv \begin{cases} 1 & \text{if } |\mathbf{x} - \mathbf{x}_0| \leq r_{\max} \\ 0 & \text{otherwise} \end{cases}, \quad (8)$$

with  $\mathbf{x}_0 = (R_0, \phi_0, z = 0)$  from Table 1.

Given a parametrized potential model  $\Phi(R, z)$  with parameters  $p_\Phi$ , the probability that the  $i$ -th star is on an

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

orbit with the actions

$$\mathbf{J}_i \equiv \mathbf{J}[\mathbf{x}_i, \mathbf{v}_i | p_\Phi], \quad (9)$$

is proportional to the given orbit distribution function DF( $\mathbf{J}$ ) with parameters  $p_{\text{DF}}$ ,

$$\text{DF}(\mathbf{J}_i | p_{\text{DF}}) \equiv \text{DF}(\mathbf{x}_i, \mathbf{v}_i | p_\Phi, p_{\text{DF}}). \quad (10)$$

The joint likelihood of a star being within the survey volume and on a given orbit is therefore

$$\mathcal{L}_i \equiv \mathcal{L}(\mathbf{x}_i, \mathbf{v}_i) = \frac{\text{DF}(\mathbf{x}_i, \mathbf{v}_i | p_\Phi, p_{\text{DF}}) \cdot \text{SF}(\mathbf{x}_i)}{\int \text{DF}(\mathbf{x}, \mathbf{v} | p_\Phi, p_{\text{DF}}) \cdot \text{SF}(\mathbf{x}) d^3x d^3v}. \quad (11)$$

The details of how we numerically evaluate the likelihood normalization to a sufficiently high precision are discussed in Paper I.<sup>9</sup>

In the scenario considered in this paper it can happen that there are a few ( $\sim 1$  in 20,000) stars entering the catalog 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 an overall smooth galactic disk. To avoid that such single stars with very low likelihood have a strong impact on the modeling we introduce here a simple strategy to ensure a robust likelihood,

$$\mathcal{L}_i \rightarrow \max(\mathcal{L}_i, \epsilon \cdot \text{median}(\mathcal{L})), \quad (12)$$

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 robust likelihood method was not used in Paper I. For future applications to real MW data an outlier model similar to the one in Bird et al. (2013) could be added: a stellar outlier distribution with constant spatial number density and velocities following a broad Gaussian.

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  $\text{pdf}(p_\Phi, p_{\text{DF}} | \text{data}) \propto \prod_{i=1}^{N_*} \mathcal{L}_i \cdot \text{prior}(p_\Phi, p_{\text{DF}})$  using a nested-grid approach and then explore the full shape of the  $\text{pdf}$  using a Monte Carlo Markov Chain (MCMC)<sup>10</sup>. Full details on this procedure are given in Paper I.

### 3.2. Distribution function model

The most simple action-based orbit DF is the quasi-isothermal DF (qDF) introduced by Binney (2010) and Binney & McMillan (2011), which has been a successful ingredient in Paper I and many disk modeling approaches (Bovy & Rix 2013; Piffl et al. 2014; Sanders & Binney 2015). The exact functional form of the qDF( $J_R, L_z, J_z | p_{\text{DF}}$ ) is given, for example, in Binney & McMillan (2011), or in Equations (2)-(4) of Paper I.

The qDF is expressed in terms of actions, frequencies, and scaling profiles for the radial stellar tracer density  $n(R_g)$ , and velocity dispersion profiles  $\sigma_z(R_g)$  and

<sup>9</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 normalization, 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>10</sup> We use the MCMC software *emcee* by Foreman-Mackey et al. (2013).

$\sigma_R(R_g)$ . The latter are functions of the guiding-center radius  $R_g$ , i.e., the radius of a circular orbit with given angular momentum  $L_z$  in a given potential. We set the scaling profiles to

$$n(R_g | p_{\text{DF}}) \propto \exp\left(-\frac{R_g}{h_R}\right) \quad (13)$$

$$\sigma_R(R_g | p_{\text{DF}}) = \sigma_{R,0} \times \exp\left(-\frac{R_g - R_\odot}{h_{\sigma,R}}\right) \quad (14)$$

$$\sigma_z(R_g | p_{\text{DF}}) = \sigma_{z,0} \times \exp\left(-\frac{R_g - R_\odot}{h_{\sigma,z}}\right). \quad (15)$$

The free model parameters of the qDF are

$$p_{\text{DF}} \equiv \{\ln h_R, \ln \sigma_{R,0}, \ln \sigma_{z,0}, \ln h_{\sigma,R}, \ln h_{\sigma,z}\}. \quad (16)$$

There is no stellar abundance or age information in the simulation snapshot used in this work. We therefore cannot define stellar sub-populations, as we normally would for the MW (see Paper I and Bovy & Rix 2013). However, the disk of the galaxy simulation was originally set up as a single axisymmetric flattened particle population whose density decreases exponentially with radius (see Section 2.1). This is actually very similar to the stellar distribution generated by a single qDF (see, e.g., Ting et al. 2013). As all particles in the disk have evolved for the same  $\sim 250$  Myr since its axisymmetric set-up, we can consider the disk essentially as a mono-age population. All of this motivates us therefore to use one single qDF to model the whole disk.

Locally, the current particle distribution in the snapshot at hand might be dominated by non-axisymmetries which evolved later in the simulation. We have no indication if for small survey volumes the qDF is still a good model for the data. We will use it anyway—to see how far we can get with the simplest model possible.

### 3.3. Potential model

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),

$$\Phi_{\text{disk}}(R, z) = -\frac{GM}{\sqrt{R^2 + (a_{\text{disk}} + \sqrt{z^2 + b_{\text{disk}}^2})^2}}, \quad (17)$$

where  $a_{\text{disk}}$  and  $b_{\text{disk}}$  are the equivalents of a disk scale length and scale height. Using 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 modeling an unfair advantage. The free model parameters of the halo are the halo scale length  $a_{\text{halo}}$  and the halo fraction, i.e., the relative halo-to-disk contribution to the radial force at the solar radius, defined as

$$f_{\text{halo}} \equiv \left. \frac{F_{R,\text{halo}}}{F_{R,\text{disk}} + F_{R,\text{halo}}} \right|_{\substack{R=R_\odot \\ z=0}}. \quad (18)$$

**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.

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

$$v_{\text{circ}}(R_\odot = 8 \text{ kpc}) \equiv \sqrt{R \frac{\partial \Phi}{\partial R}} \Big|_{z=0}. \quad (19)$$

The total set of free potential model parameters is therefore

$$p_\Phi \equiv \{v_{\text{circ}}(R_\odot), a_{\text{disk}}, b_{\text{disk}}, a_{\text{halo}}, f_{\text{halo}}\}. \quad (20)$$

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 use the *Stäckel fudge* algorithm by Binney (2012) with fixed focal length  $\Delta = 0.45$ , and interpolate the actions on a grid (Binney 2012; Bovy 2015). We made sure that the accuracy of the parameter estimates are not degraded by interpolation errors.

Galaxy disks in general, as well as the simulated disk in this work, have exponential radial density profiles. A single Miyamoto-Nagai disk is more massive at large radii than an exponential disk (see e.g., Smith et al. 2015). By construction, the DEHH-Pot introduced in Section 2.3 is therefore better suited to reproduce the overall density distribution in the simulation than the MNHH-Pot (as we will see in Figure 6 in the next section). However, the closed form expression of the Miyamoto-Nagai potential in Equation (17) has the crucial advantage of allowing much faster force and therefore action calculations. 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.

#### 4. RESULTS

At the core of this work is a suite of 22 data sets consisting of the phase-space coordinates of stellar tracer particles, drawn from the spiral galaxy simulation snapshot introduced in Section 2.1. Each data set comes from a different survey volume within the galaxy’s disk (see Section 2.2). We modeled all data sets with *RoadMapping* as described in Section 3, by fitting to it a single qDF (see Section 3.2) and the potential model MNHH-Pot (introduced in Section 3.3). This resulted in 22 independent measurements of the simulated galaxy’s potential and DF.

We present our results in two steps. In Section 4.1 we look at one of these *RoadMapping* models in detail. In Section 4.2 we then compare all 22 *RoadMapping* results, and discuss their differences in the context of spiral arms.

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

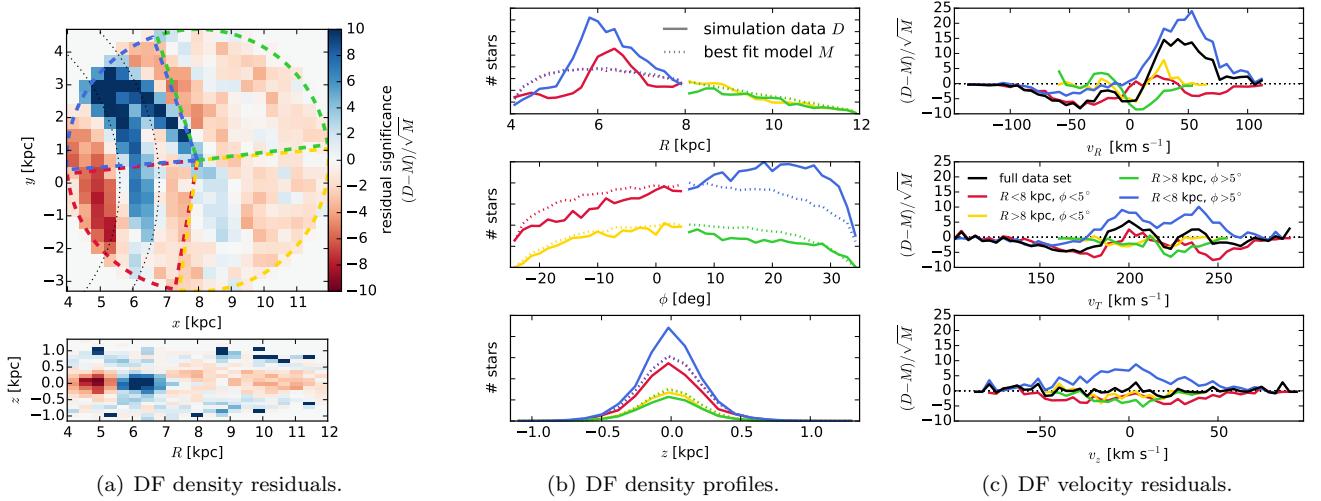
In this section we will discuss all aspects of a *RoadMapping* model for one single data set. This data set has  $N_* = 20,000$  stars that were drawn from the spherical volume with  $r_{\max} = 4$  kpc centered on a spiral arm at the “solar” radius  $R_0 = 8$  kpc. This volume is shown in orange in Figure 1 (position S8). We chose this volume because of its position centered on a smaller spiral arm, similar to our Sun being located in the Orion spiral arm. It is a bit larger than our conservative guess for the survey volume size for which we currently expect unbiased potential estimates from the final Gaia data release (DR),  $r_{\max} = 3$  kpc from the Sun (see Paper I and Section 5.3). But improvements in *RoadMapping* with respect to the treatment of measurement errors might ultimately allow to also model larger volumes. This example survey volume ranging from  $R = 4 - 12$  kpc also has the advantage of being crossed by several spiral arms of different spiral strength (see Figures 1(a) and 3). In this section our goal is now to investigate the ability of the best fit *RoadMapping* model to serve as an overall axisymmetric model for the galaxy.

The parameters of the best fit MNHH-Pot and qDF recovered with *RoadMapping* from this data set 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

The *RoadMapping* fit itself takes place in action space. However, an important sanity check to decide if the fit was successful, is to test if the best fit *RoadMapping* model (i.e., best fit action-based DF in best fit potential and in given selection function) generates a stellar distribution that reproduces the distribution of data points in observable phase-space,  $(\mathbf{x}, \mathbf{v})$ . This comparison is shown in Figure 5.

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 significantly larger number of stars move outward than inward



**Figure 5.** Comparison of the true and best fit stellar DF( $\mathbf{x}, \mathbf{v}$ ) in position-velocity space. The true DF( $\mathbf{x}, \mathbf{v}$ ) (i.e., the data  $D$ ) is the distribution of all the stars in the data set drawn from the simulation snapshot (with  $N_* = 20,000$  and  $r_{\max} = 4$  kpc centered on position S8). The best fit DF( $\mathbf{x}, \mathbf{v}$ ) (i.e., the model  $M$ ) is generated by MC sampling of the best fit qDF( $\mathbf{J}$ ) in the best fit potential model from *RoadMapping* in Table 3, given the known selection function. Panel 5(a) shows the spatial density residual significance  $(D - M)/\sqrt{M}$  of the projection to the  $(x, y)$  and  $(R, z)$  plane (the MC sampled  $M$  is the expected number of stars per bin, and  $\sqrt{M}$  the expected error due to Poisson statistics). 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 5(b) and 5(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 5(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 8.)

as compared to an axisymmetric model). We therefore compare 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 in Figure 5) 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 therefore indeed something like a good average model for the overall distribution. The region at  $R > 8$  kpc and  $\phi < 5^\circ$  (yellow in Figure 5), where neither the spiral arm nor the inter-arm regions dominate strongly, is especially well-described by the model.

Overall the qDF appears to be a good model for unperturbed regions of the disk and averages over spiral arms.

#### 4.1.2. Recovering the gravitational potential

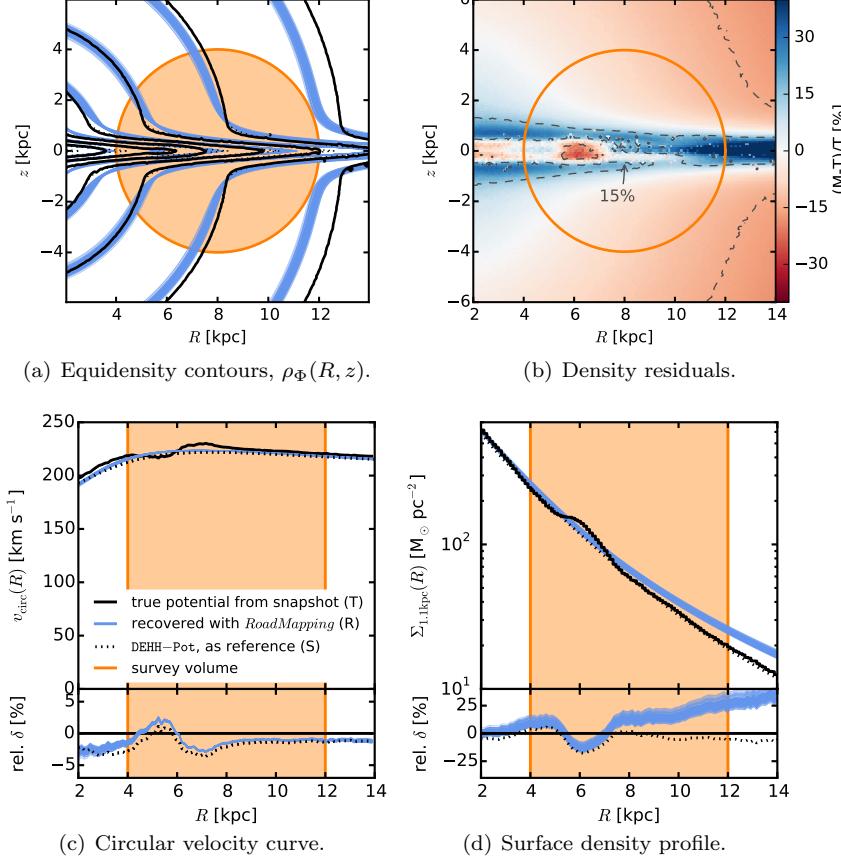
As shown in the previous section the best fit *RoadMapping* model seems to reproduce the average stellar phase-space distribution quite well. But is the corresponding potential close to the true potential?

Figure 6 compares the true potential from the simulation snapshot (symmetrized by averaging over the whole  $\Delta\phi = 2\pi$ ) and the axisymmetric reference DEHH-Pot from Table 2 with the best fit MNHH-Pot from the *RoadMapping* analysis. In particular, Figure 6 illustrates the overall matter density distribution, the rotation curve and the surface density profile. Figure 7 compares the true and recovered (median) gravitational forces at the position of each star in the data set.

The recovery of density, surface density and circular velocity curve 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 vol-

ume, and even outside, the density is recovered to within 15%. The circular velocity curve is recovered to within 5%, which is also approximately 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 discussion in Section 4.2.4) and potentially the choice of potential model (see discussion in Section 4.1.4). The overall surface density profile is a bit overestimated ( $\sim 15\%$ ) at smaller radii; this is clearly due to 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 the local density in the disk and surface density profile starts to flare strongly due to the choice of the flaring Miyamoto-Nagai disk (see Section 4.1.4). 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 is the gravitational forces. In Figure 7 we therefore compare the true force that each star in the data set feels (i.e., the radial force,  $\mathbf{F}_{R,T}(\mathbf{x}_{*,i})$ , and vertical force,  $\mathbf{F}_{z,T}(\mathbf{x}_{*,i})$ , calculated as the sum of the individual contributions by each particle in the simulation and the analytic dark matter halo at the position of star  $\mathbf{x}_{*,i}$ ) with the force that the *RoadMapping* median model predicts for each star ( $\mathbf{F}_{R,M}(\mathbf{x}_{*,i})$  and  $\mathbf{F}_{z,M}(\mathbf{x}_{*,i})$ ). We scale the difference between truth and model by a typical radial or vertical force at the given radius for which we



**Figure 6.** Comparison of the true and recovered gravitational potential  $\Phi$ . Panel 6(a) compares in the  $(R, z)$  plane equidensity contours of the overall matter density distribution  $\rho_\Phi$ , generating the potential. Panel 6(c) and 6(d) show the potential's circular velocity curve and radial surface density profile within  $|z| = 1.1$  kpc, respectively. We compare the true  $\rho_\Phi$ ,  $v_{\text{circ}}$ , and  $\Sigma_{1.1\text{kpc}}$ , of the galaxy simulation (azimuthally averaged over the whole galaxy; black solid lines), with 100 MNHH-Pot potentials drawn from the pdf of the best fit *RoadMapping* model (blue lines). This model was derived from  $N_* = 20,000$  stars in the spherical survey volume at S8 with  $r_{\text{max}} = 4$  kpc. (The extent of the survey volume is marked in orange.) The best fit parameters are given in Table 3, and Panel 6(b) shows the residuals between the matter density corresponding to the mean values in this table,  $\rho_{\Phi,\text{M}}$ , and the true density  $\rho_{\Phi,\text{T}}$ . Overplotted in Panels 6(a), 6(c), and 6(d) is also the reference DEHH-Pot (see Section 2.3; black dotted line). Over wide areas even outside of the survey volume the relative difference between true and recovered density 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. The circular velocity curve is recovered to less than 5%, especially at larger radii. [Comment to the referee: To make the paper shorter, I removed two panels from the original figure 4 (radial and vertical cuts through density distribution) and one panel from the original figure 5 (disk-to-halo ratio). The remaining panels of these two figures were now put together into this figure.] [TO DO: Rewrite colorbar label in (b) to include rho. Remove letters from legend.] [TO DO: Make survey volume similar to figures later in paper with majority of stars in bright.]

use

$$F_{R,\text{typ}}(R) \equiv v_{\text{circ},S}^2(R)/R \quad (21)$$

$$F_{z,\text{typ}}(R) \equiv F_{z,S}(R, z = h_z), \quad (22)$$

where  $v_{\text{circ},S}$  and  $F_{z,S}$  are the circular velocity and vertical force evaluated in the “true symmetric” reference potential DEHH-Pot in Table 2. As the typical vertical force at a given radius we use  $F_{z,S}$  evaluated at the scale height  $h_z = 0.17$  kpc of the disk. This is of the same order as the true vertical force averaged over all stars at this radius. Figure 7 shows therefore

$$\Delta F_R(\mathbf{x}_{*,i}) \equiv \frac{|F_{R,M}(R_i, z_i)| - |F_{R,T}(x_i, y_i, z_i)|}{F_{R,\text{typ}}(R_i)} \quad (23)$$

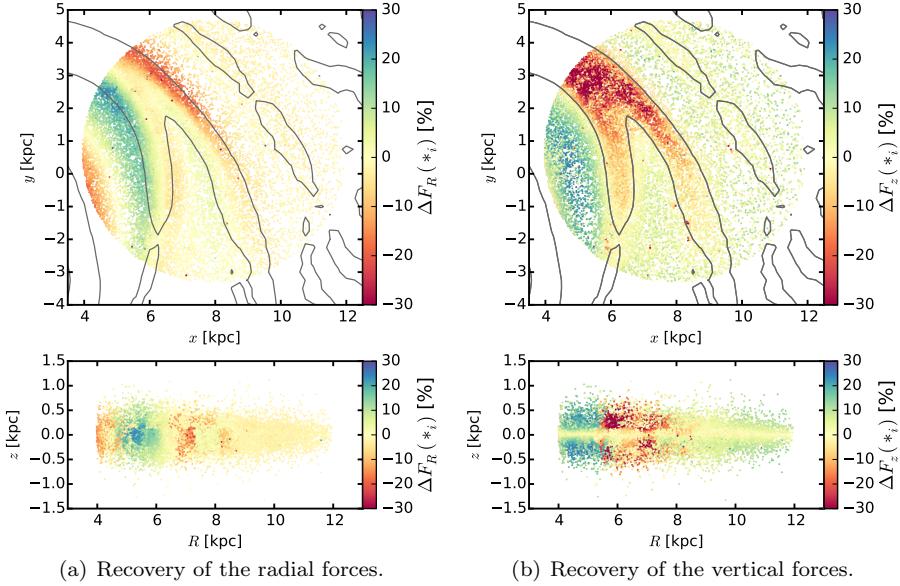
$$\Delta F_z(\mathbf{x}_{*,i}) \equiv \frac{|F_{z,M}(R_i, z_i)| - |F_{z,T}(x_i, y_i, z_i)|}{F_{z,\text{typ}}(R_i)} \quad (24)$$

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 the fit in the configuration space of the



**Figure 7.** Recovery of the gravitational forces with *RoadMapping*. We compare the true gravitational forces with the forces estimated from the *RoadMapping* best fit potential in Table 3 at the  $(x, y)$  positions (upper panels) and  $(R, z)$  positions (lower panels) of the stars that entered the analysis. In particular, we color-code the positions of the stars according to the radial (panel 7(a)) and vertical (panel 7(b)) force residuals scaled by a typical force, i.e., we show  $\Delta F_R(\mathbf{x}_{*,i})$  and  $\Delta F_z(\mathbf{x}_{*,i})$  in Equations (23)-(24). The overplotted grey contours correspond to  $\Sigma_{1.5\text{kpc},\text{disk},T}/\Sigma_{1.5\text{kpc},\text{disk},S} = 1.15$ , i.e., the true vs. the symmetric disk surface density, and mark the position of the spiral arms. The red dots mark stars for which the best fit model underestimates the (absolute value of) force. This is the case for the radial force in the leading sides of the spiral arms and the vertical force within the spiral arms, which cannot be reproduced. Blue marks correspond to stars for which the force is overestimated. Overall the radial forces are very well recovered, which is related to the good recovery of the circular velocity curve in Figure 6(c). There are more problems with the vertical force, which is related to the higher surface densities in spiral arms which slightly biases the overall *RoadMapping* model. [TO DO: Make colorbar labels include  $\mathbf{x}_{*,i}$ ]

data, and of the recovered gravitational potential. What *RoadMapping* is actually fitting, however, is the distribution in action space. Figure 8 compares the data and the model action distribution (generated by the best fit qDF) given the best fit MNHH-Pot in Table 3. (We use this axisymmetric potential to calculate the actions which lead to the best fit model, and do not attempt to estimate the true actions in the true potential.)

We note that the radial and vertical action distribution fits quite well; the axisymmetric model however contains many more stars on close-to-circular orbits ( $J_R \sim 0, J_z \sim 0$ ) than the simulation. 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 predict. In Figure 5(a) we have marked the radial extent of the stronger spiral arm with dotted lines ( $R_{\text{spiral}} \in [5.6, 6.8] \text{ kpc}$ ), and overplotted the corresponding angular momenta  $L_z = R_{\text{spiral}} \times v_{\text{circ}}(R_{\text{spiral}})$  in Figure 8. This serves as a rough estimate for the region in action space where we expect the stars of this spiral arm to be located. 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) mostly located 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.

One of the open tasks that the Galactic dynamical modeling community faces is the description of the orbit distribution of spiral arms. The above exercise of comparing the data and the model actions in a best-fit axisymmetric potential should therefore be performed for any future application to data in the Milky Way as well. It could help to learn more about the approximate orbits

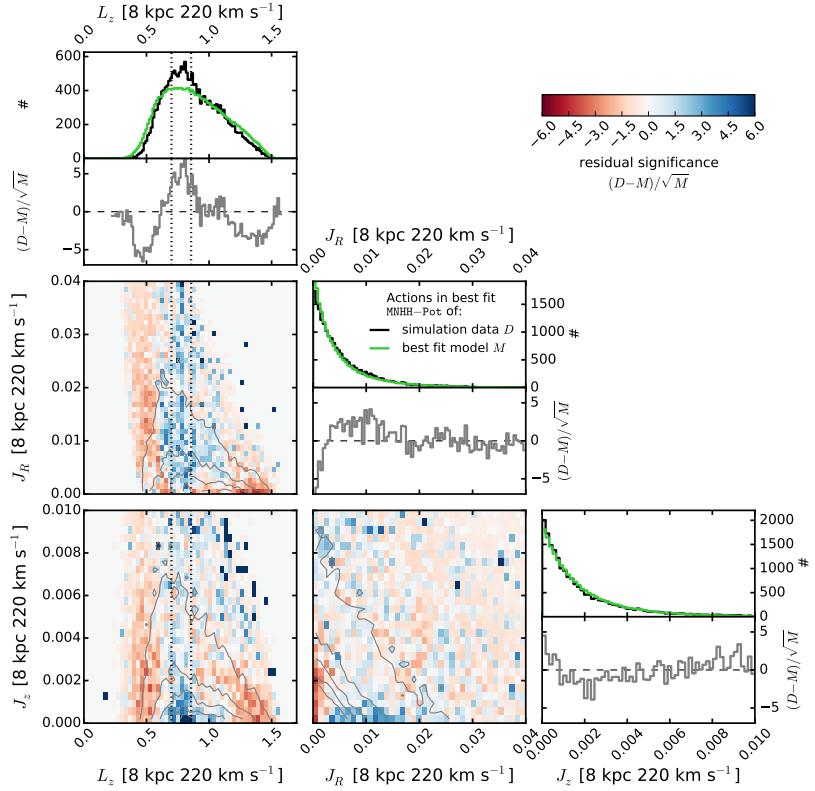
that stars move on in real spiral arms, and how spiral arms perturb axisymmetric action DFs.

#### 4.1.4. Calibrating the method with modelling a snapshot without spiral arms

To better understand the effect of spiral arms on the *RoadMapping* modeling in the previous sections, we also performed a calibration test run. For that we applied *RoadMapping* to a mock data set drawn from the same spherical survey volume as in Section 4.1 (centered at S8 with maximum radius  $r_{\text{max}} = 4 \text{ kpc}$ ), but from the initial axisymmetric snapshot of the galaxy simulation ( $t = 0 \text{ Myr}$ ), in which no spiral arms had evolved yet.

The spatial tracer distribution was very well recovered by *RoadMapping*, as well as the distribution of tracers in  $v_R$  and  $v_z$ . There were some deviations in the  $v_T$  distribution between data and best fit model, however, with our model predicting more asymmetric drift. We accounted this to the simple initial setup of the disk's stellar velocities, which does not follow a physical distribution function but only assumes a triaxial Gaussian velocity distribution (Springel et al. 2005). We expect the modeling in this calibration run to be slightly biased by this. Other than that, this confirms our expectation that the qDF is indeed a good model for the unperturbed disk, and that the deviations between data and model in Figure 5 were purely due to the spiral arms.

The surface density profile generated by the best fit potential was a perfect fit inside of  $R \sim 7 \text{ kpc}$ , but it started flaring further out. This was expected, because the Miyamoto-Nagai disk is known to be more massive at larger radii than an exponential disk. Also, if the fit is driven by the majority of stars, the fit is expected to be



**Figure 8.** Comparison of the stellar action distribution of the data set  $D$  used in the analysis and the recovered axisymmetric action distribution  $M$  (see Figure 5 for the comparison in configuration space). All actions of the data set and best fit distribution were calculated in the best fit MNHH-Pot in Table 3. The upper panel in each column shows one-dimensional histograms of the  $D$  and  $M$  distribution of angular momentum,  $L_z$ , the radial action,  $J_R$ , and the vertical action,  $J_z$ . The other panels display the residual significance  $(D - M)/\sqrt{M}$ , as both one-dimensional and two-dimensional distribution (the model  $M$  was constructed by MC sampling the best fit qDF,  $\sqrt{M}$  is the expected noise due to Poisson statistics, and  $(D - M)/\sqrt{M}$  therefore describes how significant any difference between  $D$  and  $M$  is). The two-dimensional residuals are overplotted with equidensity contours of the data  $D$ 's two-dimensional action distribution (grey solid lines). In Figure 5(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.

better at smaller radii with its higher tracer number density. The same flaring in the surface density also showed up in Figure 6. Overall our chosen potential model will therefore systematically bias vertical forces at large radii to be too strong as compared to the truth.

The recovered potential parameters for the initial snapshot,  $a_{\text{disk}} = (3.73 \pm 0.04)$  kpc,  $b_{\text{disk}} = (0.34 \pm 0.03)$  kpc,  $f_{\text{halo}} = (0.52 \pm 0.03)$ , and  $a_{\text{halo}} = (24 \pm 2)$  kpc, are consistent with the spiral arm-affected measurements in Table 3 to within 2, 4, 1, and 1.5 times the statistical error, respectively. While the halo scale length in this calibration run is also consistent with the truth within 3 times the error, we expect it to be underestimated to at least partly account for the Miyamoto-Nagai disk being too massive at large radii. As we will see later (in Section 4.2.3 and Figure 12) we seem to need an even larger survey volume to have enough radial coverage to constrain the halo scale length properly.

Compared to the truth, the circular velocity curve derived from the initial snapshot was underestimated by  $\sim 2.5\%$ ;  $v_{\text{circ}}(R_{\odot}) = (219.36 \pm 0.08)$  km s $^{-1}$ . For the snapshot with spiral arms (see Figure 6) we observe an underestimation of only  $\sim 1.5\%$ , so the bias in the initial snapshot might be related to its slightly unphysical setup of the  $v_T$  velocities. But because this underesti-

mation is present for both snapshots, it could also be a systematic bias due to the chosen Miyamoto-Nagai disk model. On the other hand, the bias also shows up in the DEHH-Pot's circular velocity curve in Figure 6. The DEHH-Pot has a more realistic shape and was found as a direct fit to the spiral arm affected disk particle distribution. The  $\sim 1.5\%$  underestimation could therefore also be introduced by the spiral arms (see Section [TO DO]).

Because the simulation immediately develops spiral arms after the initial snapshot, there was no snapshot, being both still axisymmetric and already in a dynamical steady state, to resolve this further.

#### 4.2. The influence of spiral arms in RoadMapping modeling

In the previous section we showed that for a large 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 modeling 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. A suite of data sets drawn from spiral arms and inter-arm regions

To investigate a range of data sets affected in different proportions by spiral arms, we center our test survey volumes at the positions marked in Figures 1 and 4 (see also Table 1) and consider volume sizes with  $r_{\max} \in [0.5, 1, 2, 3, 4, 5]$  kpc for  $R_0 = 8$  kpc and  $r_{\max} \in [0.5, 1, 2, 3, 4]$  kpc for  $R_0 = 5$  kpc (to avoid the galactic center). As demonstrated in Figure 4 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. Recovering the circular velocity curves and surface density profiles

It turns out that *RoadMapping* was successful in finding reasonable and even very good best fit potential models for each one of the 22 test data sets independent of size and location—given the data and limitations of the model. To illustrate this and to make this encouraging result immediately obvious, we explicitly show the circular velocity curves and surface density profiles of all analyses in Figures 9-11.

In contrast to Figure 6, these Figures only show the true profiles for the region within the galaxy where the data comes from: averaged over the angular wedge covering the radial extent of the survey volume,  $\Delta\phi = \phi_0 \pm \arcsin(r_{\max}/R_0)$ , and within  $|z| \leq 2 \cdot h_z = 0.34$  kpc, i.e., twice the scale height of the disk, which contains most of the disk mass. This is the matter distribution in which the stars are currently moving, and therefore the potential to which the modeling should be sensitive. In Figures 9-11 we also mark the survey volume and the radial bins of size  $\Delta R = 200$  pc with the highest number of stars.

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. Furthermore, whether this region is dominated by a spiral arm or an inter-arm region, and even if this region is only as small as  $r_{\max} = 500$  pc, *RoadMapping* indeed constrains 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 that *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).

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

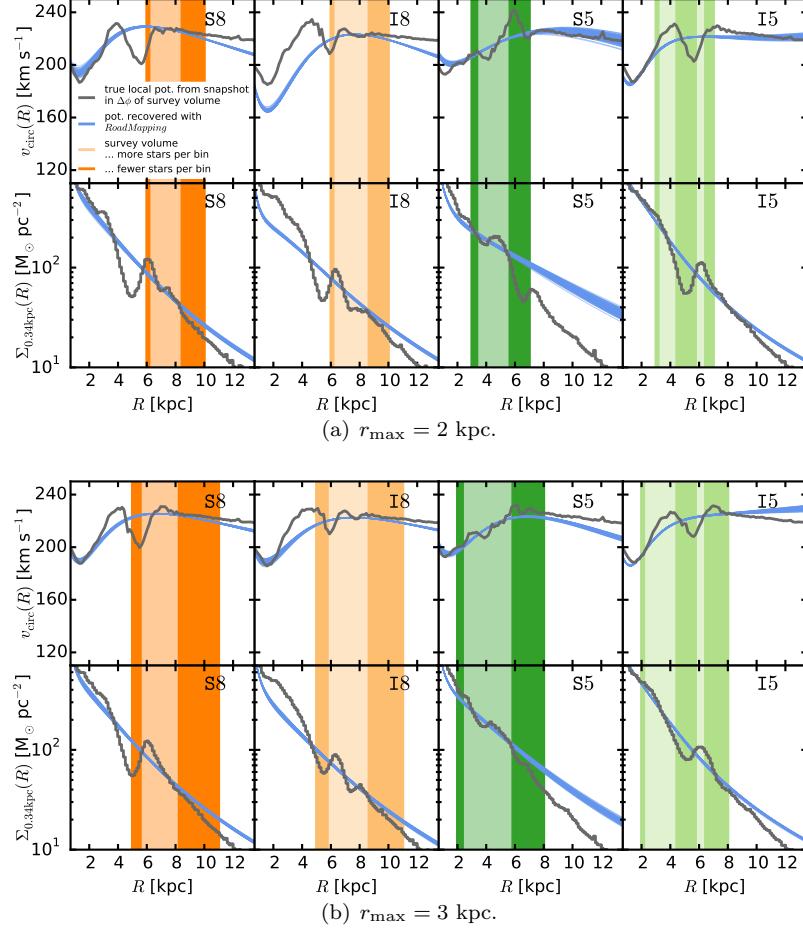
Having illustrated in Section 4.2.2 that the potential recovery is qualitatively successful, we will investigate

this result more quantitatively in the following two sections. We start here by discussing the recovered best fit potential and qDF parameters for all analyses, which are displayed in Figure 12, and compared to 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 ( $r_{\max} = 500$  pc,  $r_{\max} = 1$  kpc at I5,  $r_{\max} = 2$  kpc at I8), which we will discuss later.

We will first consider the parameters of the gravitational potential: All volumes recover  $v_{\text{circ}}(R_0)$  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  $r_{\max} \gtrsim 2$  kpc. The estimate that we get for the best fit Miyamoto-Nagai disk scale height  $b_{\text{disk}}$  seems to also be 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  $r_{\max} = 5$  kpc. The models at  $r_{\max} = 500$  pc appear to be too small to constrain the halo at all, and the MCMCs diverged completely for this parameter. 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 often be explained by having a look at the data: The large disk scale length recovered from the  $r_{\max} = 2$  kpc volume at S5, for example, mirrors the comparably flat matter distribution caused by two spiral arms close together and dominating the volume (see Figure 9(a) and the large  $\sigma_{\Delta S_{\text{spiral}}}$  for this analysis in Figure 4(d)).

The right column of Figure 12 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  $r_{\max} = 5$  kpc volume at S8. 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 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 fitting a 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 6 [TO DO: Change this explanation. Also: different halo scale length. And figure does not necessarily illustrate only the flaring MN-disk.]), 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  $R_0$  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



**Figure 9.** Comparison of the true local circular velocity curve and surface density within  $|z| \leq 2 \times h_z = 0.34 \text{ kpc}$  with the recovered *RoadMapping* models from survey volumes of size  $r_{\max} = 2 \text{ kpc}$  and  $r_{\max} = 3 \text{ kpc}$ . The blue lines show the *RoadMapping* potential models recovered from these data sets (each line is one of 100 potentials drawn from the full *pdf* sampled with the MCMC). The grey curves show the true profiles as derived from the galaxy simulation snapshot, averaged over the angular wedge  $\phi_0 \pm \arcsin(r_{\max}/R_0)$  that encloses the corresponding survey volume (see Table 1 for all  $R_0$  and  $\phi_0$  values). In other words, we show the true profiles only for the region of the spiral galaxy that was actually probed by the data. The orange and green colored regions mark the radial extent of the survey volume. We sorted the stars of each data set into radial bins of size  $\Delta R = 200 \text{ kpc}$ . The radial bins with a higher than average number of stars are marked with a lighter shade of the corresponding color, and the bins with a lower than average number with a darker shade. 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.

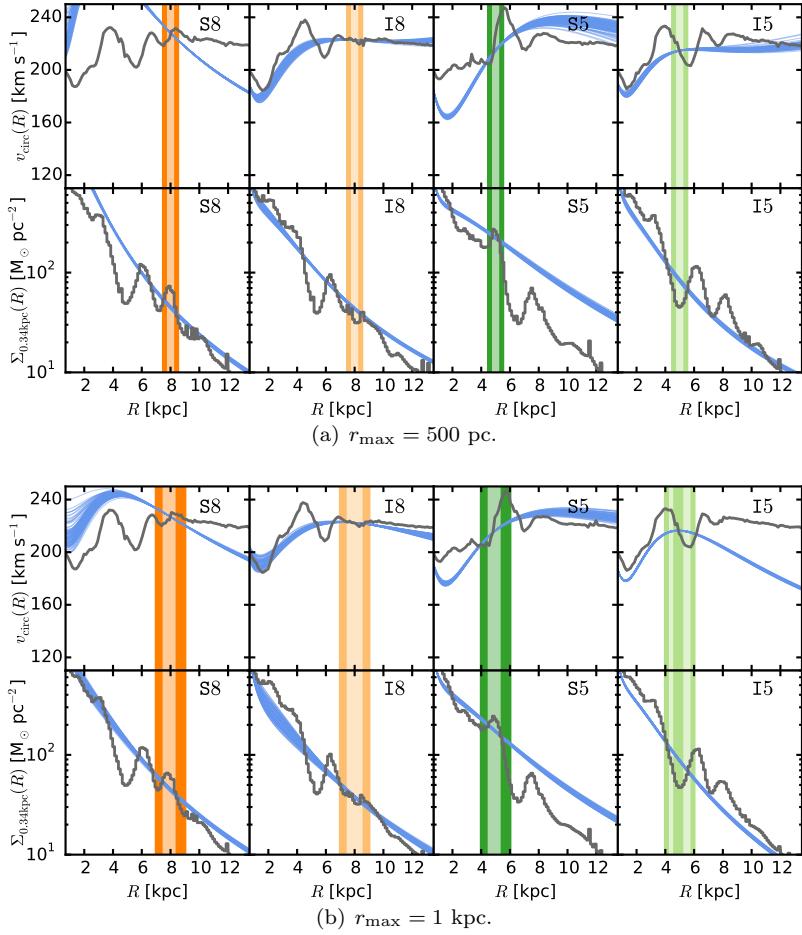
centered on the inter-arm region at  $R_0 = 8 \text{ kpc}$  (position I8) recover much longer  $h_R$ . This might be related to the fact that volumes at I8 are dominated by an especially extended inter-arm region. The volumes at I5 with  $r_{\max} = [0.5, 1] \text{ kpc}$  were not able to constrain the tracer scale length at all because of the unfortunate position between the rising density wings of two strong spiral arms (see Figure 10).

There are a few survey volumes for which the recovered parameters show some peculiarities: The models from volumes with  $r_{\max} = 1 \text{ kpc}$  at I5 and  $r_{\max} = 2 \text{ kpc}$  at I8 reject the dark matter halo completely, i.e.,  $f_{\text{halo}} = 0$ . The corresponding halo scale lengths  $a_{\text{halo}}$  are therefore unconstrained,<sup>11</sup> 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

<sup>11</sup> For these analyses and the  $r_{\max} = 500 \text{ pc}$  analyses the fit could not constrain  $a_{\text{halo}}$  and the MCMC was diverging. We had to stop the MCMC after some time, so the  $a_{\text{halo}}$  might in truth be even less constrained than shown in Figure 12.

for the way in which 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 10(b) and 9(a)). Also, while most analyses average the vertical forces radially over the spiral arms (see Figures 7, lower right panel), for these analyses the averaging happens vertically, i.e., at approximately one scale height above the plane where 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 (see also Figure 14(a) in Section 4.2.4). *RoadMapping* therefore also found a good average fit model for the stars in these volumes.

Overall we find that if the volume is large enough to average over several spiral arms and inter-arm regions, an unlucky positioning with respect to the spiral arms does not lead to strong biases in the parameter recovery. We stress again that for particularly large volumes,  $r_{\max} = 5 \text{ kpc}$ , we were able to recover all model parameters, including the halo scale length  $a_{\text{halo}}$ .



**Figure 10.** Same as Figure 9, but for all small survey volumes with  $r_{\max} = 500 \text{ pc}$  and  $r_{\max} = 1 \text{ kpc}$ .

#### 4.2.4. Recovering the gravitational forces

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 explained qualitatively, it is not yet clear how good the corresponding potential constraints actually are in a quantitative sense. To test this we calculate again  $\Delta F_R(\mathbf{x}_{*,i})$  and  $\Delta F_z(\mathbf{x}_{*,i})$  from Equations (23)-(24) at the position of each star  $\mathbf{x}_{*,i}$  in each data set (see also Figure 7). From the corresponding histograms we derive the median and the 16th and 84th percentiles ( $1\sigma$  range) and show them in Figure 13(a). We chose this diagnostic because the forces at the positions of the stars are the quantities of the potential to which our modeling is sensitive.

In addition, it is interesting to see how well the extrapolation of a recovered potential describes the overall gravitational potential of the galaxy, i.e., its predictive power. We introduced another diagnostic, which uses a cylindrical grid centered on the respective positions in Table 1, always having a radius of  $r_{\max} = 5 \text{ kpc}$  and a height of  $z = 1.5 \text{ kpc}$  both above and below the plane. In the  $(x, y)$  plane the regular grid points have a distance of  $0.25 \text{ kpc}$  and in  $z$  they have a distance of  $0.125 \text{ kpc}$  to better sample the thin disk (we throw out grid points close to the galactic center with  $R < 0.125 \text{ kpc}$ , however). We then evaluate at the position  $g_j \equiv (x_j, y_j, z_j)$

of each regular grid point the force residuals

$$\Delta F_R(g_j) \equiv \frac{|F_{R,M}(R_j, z_j)| - |F_{R,T}(x_j, y_j, z_j)|}{F_{R,\text{typ}}(R_j)} \quad (25)$$

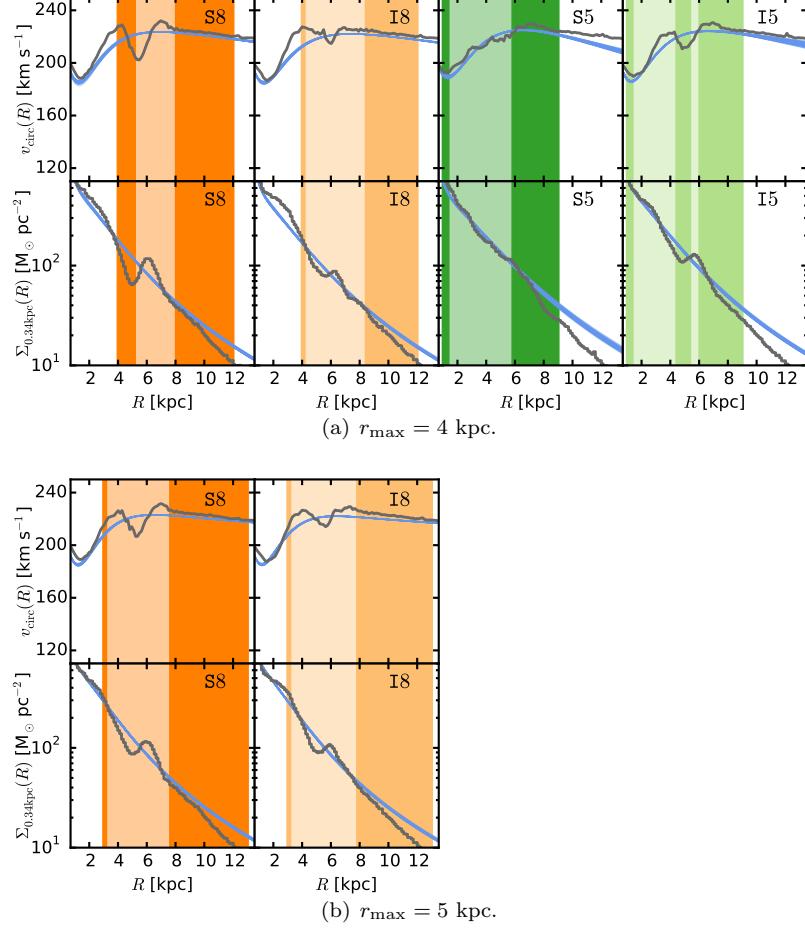
$$\Delta F_z(g_j) \equiv \frac{|F_{z,M}(R_j, z_j)| - |F_{z,T}(x_j, y_j, z_j)|}{F_{z,\text{typ}}(R_j)}, \quad (26)$$

analogous to Equations (23)-(24). The two panels in Figure 13(b) show the [16th,84th] percentile range and the median of the grid points' distribution in  $\Delta F_R(g_j)$  and  $\Delta F_z(g_j)$ .

The first and most important thing to learn from Figure 13(a) 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 ( $r_{\max} = 5 \text{ kpc}$ , as described above; Figure 13(b)), the radial forces are overall very well recovered, especially for large survey volumes. There is however an 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  $r < 5 \text{ kpc}$  are almost as good for the survey volume of  $r_{\max} = 3 \text{ kpc}$  as compared to survey volumes of  $r_{\max} = 4$  or  $5 \text{ kpc}$ . If we had to decide between a  $r_{\max} = 3 \text{ kpc}$  volume with good data quality and a larger volume



**Figure 11.** Same as Figure 9, but for all big survey volumes with  $r_{\max} = 4 \text{ kpc}$  and  $r_{\max} = 5 \text{ kpc}$ .

with worse data quality, we would lose nothing in terms of predictive power when using the smaller volume (only the halo scale length might not be as well constrained, see Figure 12).

What are the reasons for the biases that we observe in Figure 13?

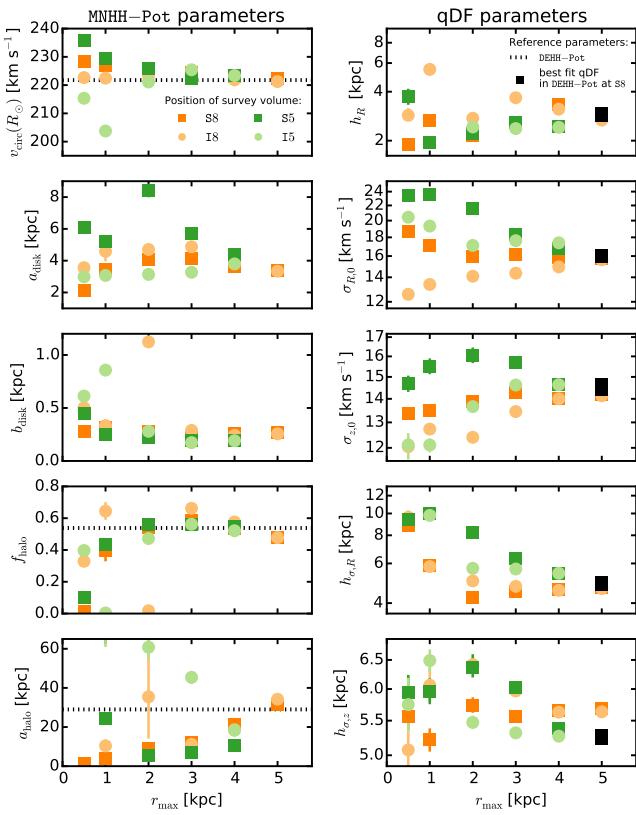
The peak of the distribution in  $\Delta F_R(\mathbf{x}_{*,i})$  and  $\Delta F_R(g_j)$  is slightly biased towards an underestimation of  $|F_{R,M}|$  in our *RoadMapping* models. We believe the explanation for this to be the fact that spiral arms are very thin. If a spiral arm crosses the observation volume, both its leading side (at large radii) and its trailing side (at small radii) are also in the volume. Stars on 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 6 and 7).<sup>12</sup>

The peak of  $\Delta F_z(\mathbf{x}_{*,i})$  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

<sup>12</sup> In the special case that the survey volume coincidentally only contains part of a spiral arm—as was the case with the analyses for  $r_{\max} = [0.5, 1] \text{ kpc}$  at position I5 (see Figure 10, right panels)—the fitting behaves differently anyway, as was already discussed in Section 4.2.3.

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 too strong on a spatial average. We illustrate this also in Figure 14: Regions far away from the strongest spiral arms (at large  $R$  and large  $|z|$ , see Figure 1) are not well described by a potential derived mostly from stars in spiral arms. As expected, the overestimation of  $|F_{z,T}(g_j)|$  in Figure 13(b) 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  $\sim 10\%$ ).

The stellar number asymmetry in the trailing vs. leading sides of spiral arms is much smaller than the stellar number asymmetry in the spiral arm vs. the inter-arm region. The bias is therefore visible in the distribution of  $\Delta F_R(\mathbf{x}_{*,i})$  (because the  $F_R$  recovery is biased only by a few stars, which leads to a bias that is visible for the majority of stars) and not in  $\Delta F_z(\mathbf{x}_{*,i})$  (because the majority of stars bias the fit and we therefore also recover  $F_z$  for the majority of stars). The bias becomes particularly pronounced for  $\Delta F_z(g_j)$  (because the inter-arm



**Figure 12.** Overview of the model parameter estimates (MNHH-Pot parameters on the left, qDF parameters on the right) recovered with *RoadMapping* from 22 different data sets. All data sets were drawn from the same simulation snapshot, but from survey volumes at different positions in the galaxy (color-coded) and of different sizes ( $r_{\text{max}}$  as indicated on the  $x$ -axis). Note that all five qDF parameters are shown here on a logarithmic scale, because *RoadMapping* uses a logarithmically flat prior for them in the fit (see Equation (16)). 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 (position S8), and fitting the qDF only. A survey volume with a radial coverage as large as  $r_{\text{max}} = 5$  kpc is required to properly recover all “true” model parameters.

regions dominate when averaging spatially which leads to a large average overestimation of  $F_z$ ) and stays small for  $\Delta F_R(g_j)$  (because 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: Replace predictive power with extrapolability.]

#### 4.2.5. The influence of the spiral arm strength

Figure 13 in Section 4.2.4 gave an overview of how accurately we recover the gravitational forces for our whole suite of data sets (depending on the size of the survey volume). In this figure some subtle clues were also hidden, which suggest that the quality of the force recovery could be correlated with the position of the data set with respect to the spiral arms. In the following we will investigate this further.

We proceed by relating the same quantities, i.e., local force recovery at the positions of the stars that entered

the analysis and power of a recovered model to predict the forces in a fixed large volume, to the actual strength and dominance of the spiral arms in the respective survey volumes (see Section 2.4). In Figure 15 we find a connection between the local force recovery and the relative spiral contrast, and Figure 16 shows a trend that the predictive power of the recovered model depends on the dominance of a spiral arm in the survey volume and is better for volumes centered on inter-arm regions.

First, we will take a closer look at Figure 15. Here, we relate  $\Delta F_R(\mathbf{x}_{*,i})$  and  $\Delta F_z(\mathbf{x}_{*,i})$  (Equations (23)-(24); see also Figure 13(a)) to the spiral contrast  $\sigma_{\Delta\text{Spiral}} \equiv \sigma[\Delta_{\text{Spiral}}(r \leq r_{\text{max}})]$  (Equation (6); see also Figure 4(d)). 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  $\sigma_{\Delta\text{Spiral}}$ . This is as expected: Volumes in which the steep gradient in surface density around a strong spiral arm ( $\Delta_{\text{Spiral},k} \neq 0$ ) is not balanced by larger areas with less perturbations ( $\Delta_{\text{Spiral},k} \sim 0$ ) have (i) a large relative spiral contrast  $\sigma_{\Delta\text{Spiral}}$ , and (ii) a large relative number of stars affected by the non-axisymmetric kinematics of the spiral arms. And for these stars, the axisymmetric *RoadMapping* model cannot properly predict the forces.

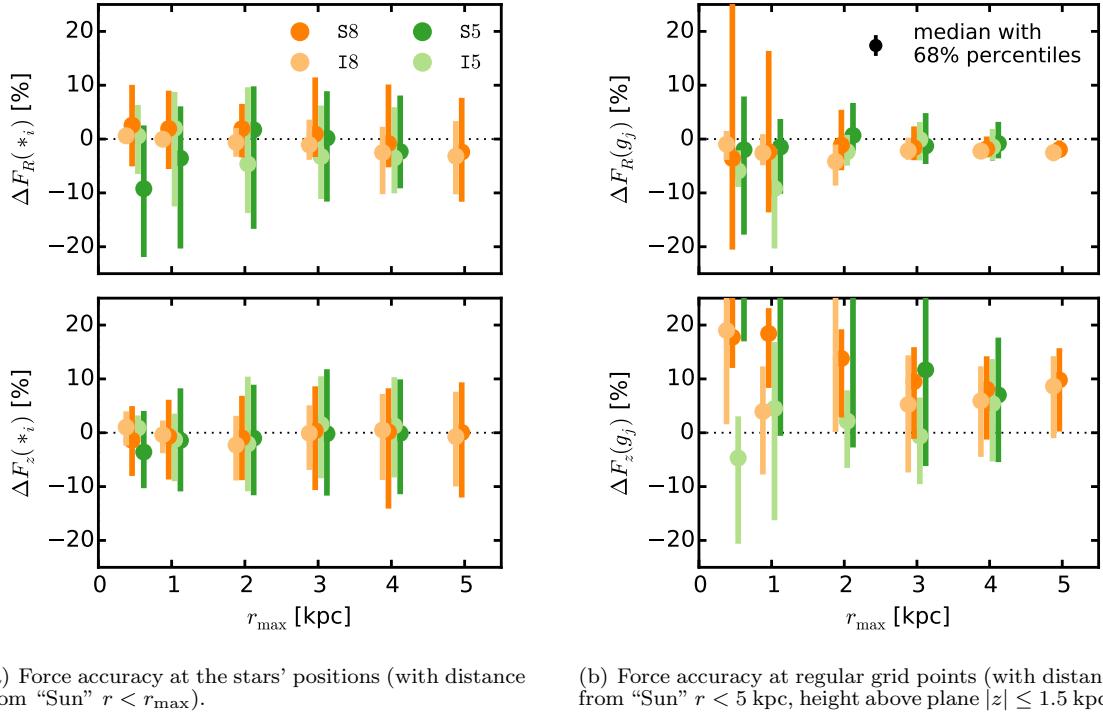
Interestingly, and even though there is some scatter, the force recovery at a given  $\sigma_{\Delta\text{Spiral}}$  is on average very similar for the radial and vertical forces (compare the linear fits in Figure 15). This means that *RoadMapping* attempts to fit both the radial and vertical forces at the positions of the stars, and is not particularly sensitive to just one of them.

As we saw in Figure 4(d), the spiral contrast  $\sigma_{\Delta\text{Spiral}}$  increases for the different test volume positions approximately in this order: I8 → S8 → I5 → S5. From Figure 15 it follows that this is also the order in which the accuracy of the force recovery decreases. (We did not include this piece of additional information in Figure 15, but it can be seen in Figure 13(a), especially for the smaller volumes.)

[TO DO: Rewrite this whole paragraph.] Figure 16 stresses even more that the position of the survey volume with respect to the spiral arms matters for the force recovery. Here, we relate the volume-averaged predictive power,  $\Delta F_R(g_j)$  and  $\Delta F_z(g_j)$  (Equations (25)-(26); see also Figure 13(b)) to the dominance of the spiral arm  $\langle \Delta_{\text{Spiral}} \rangle$  (Equation (5); see also Figure 4(c)). We show this separately for the volumes centered on spiral arms (S8 and S5; in Panel ??) and the analyses that used data sets drawn from inter-arm regions (I8 and I5; in Panel ??). We derive the fraction of grid points  $g_j$  in the reference volume (see Section 4.2.4) with forces that are misjudged by more than 10%, and plot them against  $\langle \Delta_{\text{Spiral}} \rangle$  in Panel ??, and against  $-\langle \Delta_{\text{Spiral}} \rangle$  in Panel ???. In that way the  $x$ -axis quantifies how much the spiral arms or the inter-arm regions dominate the corresponding survey volume.

Firstly, we note that there is a clear trend that the predictive power gets worse if a spiral arm strongly dominated the survey volume from which the potential constraint was derived. The same trend can be seen for the dominance of inter-arm regions, but it is weaker and less clear.

Secondly, we note that  $F_z(g_j)$  is recovered much worse than  $F_R(g_j)$ . The reason for this was already laid out in



**Figure 13.** 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 show how well *RoadMapping* recovers the forces for the ensemble of stars in the data set ( $\Delta F_R(\mathbf{x}_{*,i})$  and  $\Delta F_z(\mathbf{x}_{*,i})$  in Equations (23)-(24)) 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 ( $\Delta F_R(g_j)$  and  $\Delta F_z(g_j)$  in Equations (25)-(26)). For each distribution of  $\Delta F(\mathbf{x}_{*,i})$  and  $\Delta F(g_j)$  we show here the median as a dot with the [16th,84th] percentile range as a bar. (For presentation purposes we added a small offset  $\ll 1$  to  $r_{\max}$  on the  $x$ -axis.) 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.

Section 4.2.4).

The main result of Figure 16 is, however, the following: The predictive power of data sets drawn from survey volumes centered on inter-arm regions appears to be in general much better than that of data sets centered on spiral arms. We suspect that the reason for this is that the stellar distribution between spiral arms is smoother, more extended, and closer to the overall axisymmetric average model, such that the potentials recovered from these volumes have real predictive power for a much larger volume. This is an important result.

## 5. DISCUSSION

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

The qDF appears to be very informative. We did expect it to be at least a reasonable model for the overall symmetrized disk of the galaxy simulation, considering its initial set-up as an axisymmetric, exponentially decreasing particle distribution that subsequently evolved as a mono-age population (see Sections 2.1 and 3.2). In Section 4.1.1 and 4.1.3 we demonstrated that the qDF is indeed a good average model for the tracer distribution in a large survey volume—even though the spiral arms did introduce considerable deviations.

We had, however, no indications beforehand of how well the axisymmetric qDF would perform in a small survey volume completely dominated by non-axisymmetric

spiral arms. It would not have been surprising if *RoadMapping* had failed. But in Section 4.2.2 it turned out that the potential measurements were reliable even in most of the small volumes with  $r_{\max} = [0.5, 1]$  kpc. And the corresponding qDF parameters were tightly constrained and reasonable as well.

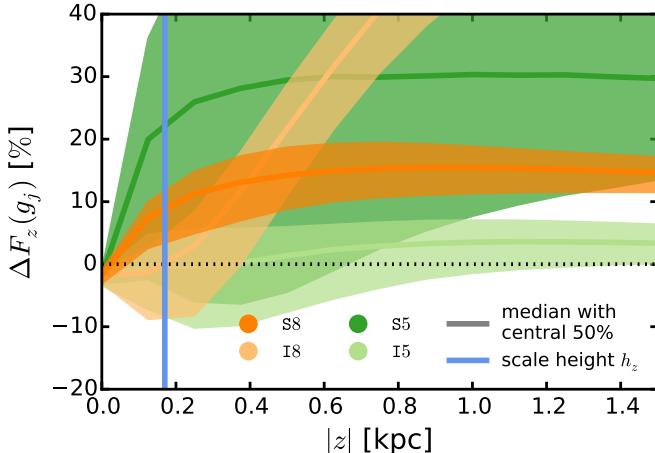
We deduce that the qDF is indeed flexible and robust enough to work with data affected by non-axisymmetries.

That the corresponding potential constraints were reliable as well leads to the following conclusion: A potential model that does not fit the gravitational forces acting on the stars appears to lead to such an unrealistic orbit/action distribution, that a fit with even such a simple orbit DF as the qDF is impossible. This demonstrates once more how powerful the concept of an orbit DF is.

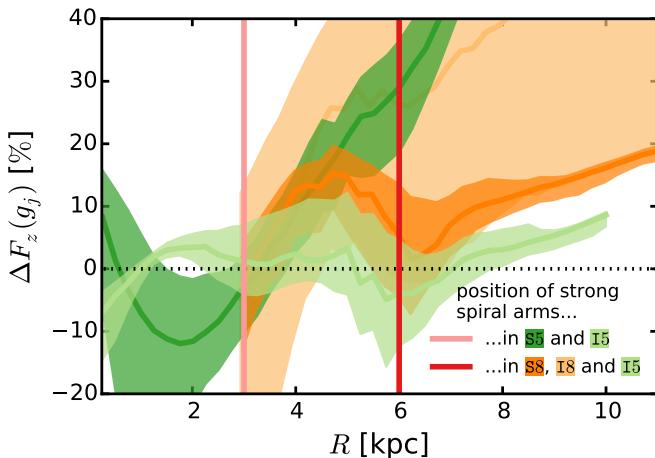
### 5.2. On the restrictiveness of the parametrized potential model

How much does the choice of potential model matter for the success of the modeling?

We used, on the one hand, a bulge and halo model that reproduces the true bulge and halo better than we can hope to use in reality for the MW. The fact that for all except 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



(a) Vertical force prediction vs. height above the plane.

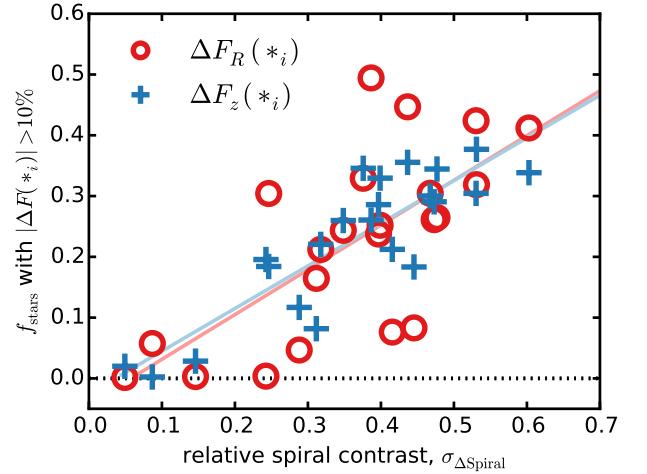


(b) Vertical force prediction vs. galactocentric radius.

**Figure 14.** RoadMapping’s ability to predict the vertical gravitational force as a function of the height above the plane,  $|z|$ , and galactocentric radius,  $R$ . In particular, we show the predictive power, calculated from Equation (26), for the RoadMapping potential models derived from the four data sets with  $r_{\max} = 2$  kpc at positions S8, I8, S5, and I5 (color-coded; see also Figure 1). The colored bands show the distribution (median with the central 50% percentiles) of  $\Delta F_z$ ’s of all regular grid points  $g_j$  at a given  $|z|$  or  $R$ . This figure demonstrates the origin of the strong bias in the vertical force prediction, which we found in Figure 13(b). As the fit is driven by the excess of stars that feel stronger vertical forces in the spiral arms, e.g., at  $R \sim 3$  kpc or  $R \sim 6$  kpc, we get the vertical force right at these radii in the plane of the disk. We consequently overestimate it in the smooth outer regions of the galaxy and at large heights above the plane, where the spiral arms don’t play a role. (The potential model with  $r_{\max} = 2$  kpc at I8, light orange, is a special case that was discussed in Section 4.2.3.)

to the total radial force at  $R = 8$  kpc is only  $\sim 9 - 10\%$ ) remedies this apparent advantage.

On the other hand, we used 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 paper 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 probably not give much better results.



**Figure 15.** Influence of the spiral arm contrast on the recovery of the gravitational forces at the positions of the stars that entered the RoadMapping analysis. Each circle/cross pair corresponds to one of our 22 data sets. The relative spiral contrast on the  $x$ -axis is quantified as  $\sigma_{\Delta\text{Spiral}} \equiv \sigma[\Delta_{\text{Spiral}}(r \leq r_{\max})]$  calculated within each survey volume according to Equation (6) in Section 2.4. On the  $y$ -axis the fraction of stars ( $f_{\text{stars}}$ ) in each data set is shown for which the radial (red circles) and vertical (blue crosses) force residual calculated from Equations (23)-(24) is larger than 10% (i.e., at 0 all stars have good force measurements, at 1 everything went wrong). The red and blue lines 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. On average, the radial and vertical forces are equally well-recovered at a given spiral contrast. [TO DO: make this figure analogous to other figure (i.e. legend and axis labels)]

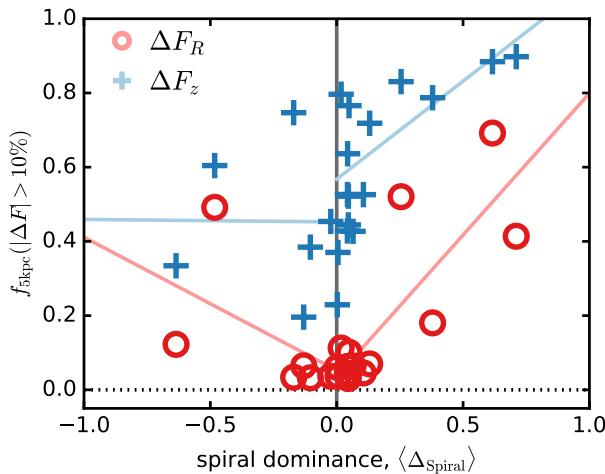
It appears that a potential model with a reasonable shape and flexibility (here: disk+bulge+halo structure with 5 free parameters) can do well enough in finding a good fit, both locally for small volumes and overall for large volumes.

This is in agreement with one of our key results of Paper I, where we managed to successfully fit data from a MW-like (but axisymmetric) galaxy model with a bulgeless potential of a restrictive Stäckel form. This was illustrated in Figure 16 of Paper I. Considering that we used there the same number of stars,  $N_* = 20,000$ , the potential uncertainties were much greater than in the analogous figure of this work, Figure 6(a). We believe that this how RoadMapping accounts for an inconvenient potential parametrization—by increasing the uncertainty of the model estimate—which is exactly as it should be.

### 5.3. Gaia measurement errors and choosing the survey volume size

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  $< 10\%$  and proper motion uncertainties of less than  $3 \text{ mas yr}^{-1}$  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 (at least the final data release of) Gaia lie even below these limits.



**Figure 16.** Influence of the spiral arm dominance on the extrapolability of the *RoadMapping* potential models recovered from each survey volume. Each circle/cross pair corresponds to one of our 22 *RoadMapping* analyses. How much a spiral arm dominates within a given survey volume is quantified by  $\langle \Delta_{\text{Spiral}} \rangle$  from Equation (5) and shown on the  $x$ -axis.  $\langle \Delta_{\text{Spiral}} \rangle \sim 0$  means spiral arms and inter-arm regions dominate equally in the survey volume. The more negative  $\langle \Delta_{\text{Spiral}} \rangle$  is, the more dominated by an inter-arm region is the survey volume. The  $y$ -axis shows the volume fraction within which the radial (red circles) and vertical (blue crosses) gravitational force residuals are larger than 10%. In particular, we extrapolate each potential model within a cylindrical volume with  $r_{\text{max}} = 5$  kpc and  $|z| \leq 1.5$  kpc, centered on the respective survey volume positions given in Table 1.  $f_{5\text{kpc}}(|\Delta F| > 10\%)$  therefore quantifies how well a *RoadMapping* potential model can be extrapolated out to 5 kpc from the center of the survey volume. At  $f_{5\text{kpc}}(|\Delta F| > 10\%) \sim 0$  the extrapolation works best. The red and blue lines are linear fits and only serve as guide to the eye. This figure demonstrates that the extrapolability of the model gets better the less a spiral arm dominates the data set. For the dominance of inter-arm regions this trend is less pronounced. Overall, the radial force can be much better extrapolated than the vertical force, and models from data sets centered on inter-arm regions can be more reliably extrapolated than those centered on spiral arms. [Comment to the referee: The previous figure with two panels was condensed in one panel, because this appears to be easier to understand.]

In paper I we focused on recovering completely unbiased model parameters and found that *RoadMapping* is robust against moderate deviations of the model assumptions. In this work we released the condition that the model parameters themselves 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 than we found in Paper I, and before the measurement uncertainties noticeably muddle the constraints. Also, Gaia uncertainties are already below our formal error anyway.

In addition, we found in this work that a volume of  $r_{\text{max}} = 3$  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 (e.g., Reid et al. 2014; Bovy et al. 2015, and others, see Introduction 1).

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

How well we can do with the first few Gaia data releases remains to be seen. The Gaia DR1 in September 2016 will have parallax measurements of  $\sim 16\%$  for red clump giants at a distance of  $\sim 500$  pc from the Sun (de Bruijne et al. 2014; Michalik et al. 2015), which is not precise enough for *RoadMapping*. We could, however, use photometric distances, which should be precise enough for red clump stars and even extend over a larger volume. The Gaia DR2 at the end of 2017 might however already cover  $\sim 2$  kpc with good parallaxes for fainter stars as well.

#### 5.4. Spiral arms in the solar neighbourhood

The Sun is located in one of the smaller spiral arms of the MW, the local Orion spur/arm (Morgan et al. 1953). 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) (towards the Galactic center). It is, however, still under dispute which arms are actually major arms of the MW (Blaauw 1985; Xu et al. 2013; Zhang et al. 2013).

How reliable the *RoadMapping* results from the Gaia DR1 will be depends on the strength of the local Orion arm, which will dominate the survey volume. *RoadMapping* had, for example, some difficulties recovering the circular velocity curve for small volumes (see Figure 10).

However, recent measurements of the MW’s rotation curve (Bovy et al. 2012d; Reid et al. 2014) confirm 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). Based on independent information on the location and strength of the MW spiral arms, we could also use the present approach to estimate systematic uncertainties.

For later Gaia data releases, where the tracers extend further into the Galaxy ( $\sim 2$  kpc from the Sun), the Sagittarius and Perseus arms could also play a role. In general, *RoadMapping* should do much better with larger volumes, and if several spiral arms and inter-arm regions within the survey volume average each other out.

This should work especially well if the MW is—like the simulation in this study—a four-armed spiral. The exact number of spiral arms is however still disputed (see Section 2.5). If Gaia will confirm studies that the MW is a two-armed grand design spiral instead, having a very large survey volume would become even more important to achieve a good axisymmetric average *RoadMapping* model for the MW. Local measurements of the potential as in Figure 10 should however still be possible.

#### 5.5. Absence of a central bar in the simulation

The simulations we analyzed do not have a prominent bar, and so we have not explicitly explored the impact of such a feature. Bars can play an important role in the dynamics when very small volumes near a resonance are considered (e.g. Dehnen 2000). When considering volumes of  $\gtrsim 1$  kpc, we have no reason to believe that this should severely affect the robustness of such an analysis.

### 5.6. Interpreting RoadMapping results

The two main findings of this work give a clear directive regarding how we should deal with any future result about the MW's potential derived with *RoadMapping*. (i) If the data spans a large volume, a significant proportion of the disk (at least  $R \sim 5 - 11$  kpc), and averages over several spiral arms and inter-arm regions, we can trust and use the resulting model as an overall axisymmetric potential for the MW. For  $R \sim 3 - 13$  kpc we should even be able to make definite statements about the dark matter distribution. (ii) If the data do not span such a large volume, we can still believe the local constraints. In particular, the surface density within 1-2 disk scale heights, the circular velocity within the survey volume, and the average gravitational forces where the majority of stars is located.

Fortunately, this is consistent with the procedure by Bovy & Rix (2013), who used *RoadMapping* to constrain the vertical force  $F_z$  for each MAP at only one radius that corresponds to a typical radius. It will be interesting to see whether *RoadMapping* potential constraints from Gaia data will agree with their findings.

In addition, one should always compare the distribution of the data and the recovered model in configuration space ( $\mathbf{x}, \mathbf{v}$ ) and in action space, as we did in Section 4.1. This is not only a sanity check to confirm the goodness of the fit, but it might also reveal some substructure in the data that only becomes visible when comparing it to an axisymmetric smooth model.

### 5.7. Recovering non-axisymmetric structures in the potential from modeling small volumes

As we saw in Sections 4.2.2 and 4.2.4, *RoadMapping* also makes a very good attempt at constraining the local potential for volumes as small as  $r_{\max} = 500$  pc or 1 kpc. It recovers, for example, the higher surface density in volumes dominated by spiral arms, or correctly estimates the circular velocity at the median radial position of the stars (see Figure 10).

At a time after the final Gaia data release, when we have data of high accuracy covering a large proportion of the disk, we could make use of this interesting property of *RoadMapping* modeling. 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 only probe the local potential, even when using an axisymmetric potential model, and should be sensitive to the overdensities induced by the spiral arms. In this way it should be possible to build up a non-axisymmetric map of the MW potential—albeit with very large spatial pixels—with constraints from dynamical modeling only.

## 6. CONCLUSION

*RoadMapping* is a well-tested axisymmetric dynamical modeling machinery that simultaneously fits an action-based DF and gravitational potential to the individual 6D phase-space coordinates of stellar populations in the MW disk. *RoadMapping* builds on previous work by Binney & McMillan (2011); Binney (2012); Bovy (2015) and was first applied by Bovy & Rix (2013). *RoadMapping* was improved and tested in detail against the breakdowns of its modeling assumptions by Trick et al. (2016).

In this paper we investigated the robustness of *RoadMapping* when modeling a non-axisymmetric system. We explore this for the first time explicitly, by modeling a simulated spiral galaxy from D'Onghia et al. (2013), which has stronger spiral arms than we expect in the MW, and by comparing the results to the true potential. We find that *RoadMapping*-like action-based dynamical modeling is very robust against perturbations of spiral arms, especially if the survey volume is large enough to encompass both spirals and inter-arm regions. 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 a position equivalent to that of the Sun.

In Section 4.2 we have investigated the role of survey volumes differing in size and position with respect to the spiral arms and inter-arm regions within the simulated galaxy. We find that the gravitational forces are mostly well-recovered at the locations of the stars that entered the analysis. For survey volume sizes  $r_{\max} \geq 3$  kpc the recovered potential model already becomes a good average potential model for the whole galaxy. For some positions of the survey volume center, e.g., in a smooth and not-too-depleted inter-arm region, smaller volumes can also 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 particularly notable, as the breakdown of the assumption of axisymmetry implies a breakdown of several model assumptions simultaneously: (i) orbital actions are not fully conserved anymore, (ii) the true potential is not spanned by the family of model potentials, (iii) the quasi-isothermal DF need not, or will not, 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  $r_{\max} = 500$  pc dominated by spiral arms.

The results of this paper imply that *RoadMapping* should be well-suited to making new measurements of the MW's gravitational potential with the upcoming Gaia data releases. It might even potentially work with the Gaia DR1 with its smaller coverage of the disk ( $r_{\max} \sim 1$  kpc) because the local Orion arm, in which the Sun is located, is thought to be only a minor spiral arm in the MW and should not significantly disturb the modeling.

## 7. ACKNOWLEDGEMENTS

W.H.T. thanks Stephen Pardy for helpful advice on handling  $N$ -body simulations. W.H.T. and H.-W.R. acknowledge funding from the European Research Council under the European Unions Seventh Framework Programme (FP 7) ERC Grant Agreement n. [321035]. J.B. received support from the Natural Sciences and Engineering Research Council of Canada for financial support of this project. E.D. gratefully acknowledges the support of the Alfred P. Sloan Foundation and ATP NASA Grant No NNX14AP53G, and support by Sonderforschungs-

bereich SFB 881 "The Milky Way System" of the German Research Foundation (DFG).

## REFERENCES

- Barros, D. A., Lépine, J. R. D., & Dias, W. S. 2016, *A&A*, 593, A108
- Binney, J. 2010, *MNRAS*, 401, 2318
- . 2012, *MNRAS*, 426, 1324
- Binney, J., Gerhard, O., & Spergel, D. 1997, *MNRAS*, 288, 365
- Binney, J., Gerhard, O. E., Stark, A. A., Bally, J., & Uchida, K. I. 1991, *MNRAS*, 252, 210
- Binney, J., & McMillan, P. 2011, *MNRAS*, 413, 1889
- Binney, J., & Tremaine, S. 2008, *Galactic Dynamics: Second Edition* (Princeton University Press)
- 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
- Bland-Hawthorn, J., & Gerhard, O. 2016, *ARA&A*, 54, 529
- Blitz, L., & Spergel, D. N. 1991, *ApJ*, 379, 631
- Bovy, J. 2014, *ApJ*, 795, 95
- . 2015, *ApJS*, 216, 29
- Bovy, J., Bird, J. C., García Pérez, A. E., et al. 2015, *ApJ*, 800, 83
- Bovy, J., & Hogg, D. W. 2010, *ApJ*, 717, 617
- Bovy, J., Hogg, D. W., & Roweis, S. T. 2009, *ApJ*, 700, 1794
- 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., Schlaufly, E. F., et al. 2016, *ApJ*, 823, 30
- Bovy, J., Allende Prieto, C., Beers, T. C., et al. 2012d, *ApJ*, 759, 131
- Churchwell, E., Babler, B. L., Meade, M. R., et al. 2009, *PASP*, 121, 213
- Das, P., & Binney, J. 2016, *MNRAS*, 460, 1725
- de Bruijne, J. H. J., Rygl, K. L. J., & Antoja, T. 2014, in *EAS Publications Series*, Vol. 67, *EAS Publications Series*, 23–29
- Dehnen, W. 1998, *AJ*, 115, 2384
- . 2000, *AJ*, 119, 800
- D'Onghia, E. 2015, *ApJ*, 808, L8
- D'Onghia, E., Vogelsberger, M., & Hernquist, L. 2013, *ApJ*, 766, 34
- Drimmel, R., & Spergel, D. N. 2001, *ApJ*, 556, 181
- Eyer, L., Holl, B., Pourbaix, D., et al. 2013, *Central European Astrophysical Bulletin*, 37, 115
- Famaey, B., Jorissen, A., Luri, X., et al. 2005, *A&A*, 430, 165
- 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
- Grand, R. J. J., Kawata, D., & Cropper, M. 2012, *MNRAS*, 421, 1529
- 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
- Michalik, D., Lindegren, L., & Hobbs, D. 2015, *A&A*, 574, A115
- Minchev, I., Chiappini, C., & Martig, M. 2014, *A&A*, 572, A92
- Minchev, I., Famaey, B., Combes, F., et al. 2011, *A&A*, 527, A147
- Minchev, I., Famaey, B., Quillen, A. C., et al. 2012, *A&A*, 548, A127
- 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
- Oort, J. H., Kerr, F. J., & Westerhout, G. 1958, *MNRAS*, 118, 379
- Piffl, T., Binney, J., McMillan, P. J., et al. 2014, *MNRAS*, 445, 3133
- Portail, M., Wegg, C., Gerhard, O., & Martinez-Valpuesta, I. 2015, *MNRAS*, 448, 713
- Reid, M. J., Menten, K. M., Zheng, X. W., et al. 2009, *ApJ*, 700, 137
- Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2014, *ApJ*, 783, 130
- Sackett, P. D. 1997, *ApJ*, 483, 103
- 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
- Siebert, A., Famaey, B., Binney, J., et al. 2012, *MNRAS*, 425, 2335
- Smith, R., Flynn, C., Candlish, G. N., Fellhauer, M., & Gibson, B. K. 2015, *MNRAS*, 448, 2934
- Solway, M., Sellwood, J. A., & Schönrich, R. 2012, *MNRAS*, 422, 1363
- 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, *ApJ*, 830, 97
- Urquhart, J. S., Figura, C. C., Moore, T. J. T., et al. 2014, *MNRAS*, 437, 1791
- Vallée, J. P. 2008, *AJ*, 135, 1301
- . 2014, *AJ*, 148, 5
- Vera-Ciro, C., & D'Onghia, E. 2016, *ApJ*, 824, 39
- Wegg, C., & Gerhard, O. 2013, *MNRAS*, 435, 1874
- Williams, M. E. K., Steinmetz, M., Binney, J., et al. 2013, *MNRAS*, 436, 101
- 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