HAL
open science

Weighing the Galactic disk using phase-space spirals. IV. Tests on a three-dimensional galaxy simulation

A. Widmark, J. A. S. Hunt, C. F. P. Laporte, G. Monari

To cite this version:

A. Widmark, J. A. S. Hunt, C. F. P. Laporte, G. Monari. Weighing the Galactic disk using phase-space spirals. IV. Tests on a three-dimensional galaxy simulation. Astronomy and Astrophysics - A&A, 2022, 663, 10.1051/0004-6361/202243173 . insu-03765351

HAL Id: insu-03765351

https://insu.hal.science/insu-03765351

Submitted on 1 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A&A 663, A16 (2022)
https://doi.org/10.1051/0004-6361/202243173
© ESO 2022

Astronomy & Astrophysics

Weighing the Galactic disk using phase-space spirals

IV. Tests on a three-dimensional galaxy simulation

A. Widmark[1], J. A. S. Hunt[2], C. F. P. Laporte[3], and G. Monari[4]

[1] Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Jagtvej 128, 2200 Copenhagen N, Denmark
e-mail: axel.widmark@nbi.ku.dk
[2] Center for Computational Astrophysics, Flatiron Institute, 162 5th Av., New York City, NY 10010, USA
[3] Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, 08028 Barcelona, Spain
[4] Université de Strasbourg, CNRS UMR 7550, Observatoire astronomique de Strasbourg, 11 rue de l'Université, 67000 Strasbourg, France

Received 21 January 2022 / Accepted 9 March 2022

ABSTRACT

In this fourth article on weighing the Galactic disk using the shape of the phase-space spiral, we have tested our method on a billion particle three-dimensional *N*-body simulation, comprised of a Milky Way like host galaxy and a merging dwarf satellite. The main purpose of this work was to test the validity of our model's fundamental assumptions that the spiral inhabits a locally static and vertically separable gravitational potential. These assumptions might be compromised in the complex kinematic system of a disturbed three-dimensional disk galaxy; in fact, the statistical uncertainty and any potential biases related to these assumptions are expected to be amplified for this simulation, which differs from the Milky Way in that it is more strongly perturbed and has a phase-space spiral that inhabits higher vertical energies. We constructed 44 separate data samples from different spatial locations in the simulated host galaxy. Our method produced accurate results for the vertical gravitational potential of these 44 data samples, with an unbiased distribution of errors with a standard deviation of 7%. We also tested our method under severe and unknown spatially dependent selection effects, also with robust results; this sets it apart from traditional dynamical mass measurements that are based on the assumption of a steady state and which are highly sensitive to unknown or poorly modelled incompleteness. Hence, we will be able to make localised mass measurements of distant regions in the Milky Way disk, which would otherwise be compromised by complex and poorly understood selection effects.

Key words. Galaxy: kinematics and dynamics – Galaxy: disk – solar neighborhood – astrometry

1. Introduction

An important avenue for learning about the Milky Way is measuring its gravitational potential and matter density using stellar dynamics (e.g. Dehnen & Binney 1998; Widrow et al. 2008; McMillan 2017; An et al. 2021). This is especially important for the Galaxy's distribution of dark matter (Read 2014; de Salas & Widmark 2021); for example, the local dark matter density is proportional to the signal strength in direct and some indirect dark matter detection experiments (Jungman et al. 1996; Klasen et al. 2015). In recent years, the *Gaia* satellite (Gaia Collaboration 2018) has revolutionised the research field of Galactic dynamics, increasing the astrometric precision and sample size by orders of magnitude compared to previous surveys (Perryman et al. 1997).

Dynamical mass measurements are typically performed under the assumption of a steady state. However, *Gaia* has made it all the more clear that the Galaxy is host to time-varying dynamical structures. One such structure is the phase-space spiral recently discovered by Antoja et al. (2018), seen in the phase-space plane of position and velocity in the direction perpendicular to the Galactic disk, which is present in the solar neighbourhood as well as more distant regions of the Galactic disk (Laporte et al. 2019; Wang et al. 2019; Gaia Collaboration 2021). The phase-space spiral is not in a steady state and thus constitutes a bias to studies that are based on that assumption. However, its presence is not necessarily an obstacle to dynamical mass measurements but it can instead be regarded as an asset because the winding and shape of the phase-space spiral can inform us of the gravitational potential it inhabits.

This work is part of a longer series about weighing the Galactic disk using the phase-space spiral; we have previously published three articles (Widmark et al. 2021a,b,c) which we refer to as Paper I, Paper II, and Paper III. In these articles, we have tested our method on one-dimensional simulations and applied it to actual *Gaia* data, analysing the immediate solar neighbourhood using the radial velocity sample as well as the distant Galactic disk using the proper motion sample.

This method for weighing the Galactic disk is new, and so far it has only been used in the previous articles of this series. That makes it especially important to test and validate our method on simulations where the answer is known. In this work, we have applied it to a billion particle three-dimensional simulation (Hunt et al. 2021), as a test of potential sources of bias that could arise in the complex three-dimensional dynamics of a disk galaxy perturbed by an external satellite. Most importantly, we aim to test our model's fundamental assumptions that the phase-space spiral inhabits a gravitational potential that is vertically separable (commonly known as the "one-dimensional approximation") and static (neglecting the self-gravity of the spiral perturbation). In addition to the main application of our method, we also ran tests in the presence of strong selection effects,

mimicking the incompleteness due to dust extinction and stellar crowding seen in Paper II and Paper III; in this case we included a simple extinction model in our method of inference, similar to the method used in Paper III. Furthermore, we tested our method's sensitivity with respect to a biased height of the disk mid-plane.

This article is structured as follows. The details of the simulation are discussed in Sect. 2. How the data samples are constructed is explained in Sect. 3. In Sect. 4 we discuss our model of inference and in Sect. 5 we present our results. In Sects. 6 and 7, we discuss and conclude.

2. Three-dimensional simulation

We use the pure N-body simulation which is labelled M1 in Hunt et al. (2021), comprised of a Milky Way like host galaxy and a dwarf satellite that merges into it. The initial conditions for the Milky Way like host galaxy was created with the parallelised version of the `galactics` initial condition generator[1] (Kuijken & Dubinski 1995), using the parameters from the Milky Way like model labelled "MWb" in Widrow & Dubinski (2005). This creates a disk which is stable against bar and spiral formation for several billion years (with Toomre parameter $Q = 2.3$). The dwarf galaxy is model L2 of Laporte et al. (2018), which is comprised of two Hernquist spheres (Hernquist 1990). The first represents dark matter and has virial mass $M_{200} = 6 \times 10^{10}\, M_\odot$, concentration parameter $c_{200} = 28$, halo mass $M_h = 8 \times 10^{10}\, M_\odot$, and scale radius $a_h = 8$ kpc. The second represents the stellar component embedded in the dark halo, and has stellar mass $M_* = 6.4 \times 10^8\, M_\odot$ and scale radius $a_h = 0.85$ kpc (see Laporte et al. 2018, for a more thorough description).

The combined model was evolved for 8.3 Gyr with the GPU based N-body tree code `Bonsai` (Bédorf et al. 2012, 2014), using a smoothing length of 50 pc and an opening angle $\theta_o = 0.4$ radians. In this work we analyse the 'present day' snapshot, with $t = 6.87$ Gyr, available on `Flathub`[2] as Model M1, snapshot 703. In this time snapshot, the satellite has a phase-space position of $(x, y, z, u, v, w) = (9.5, -0.4, -6.7, 115.7, 5.9, 311.9)$ kpc and km s^{-1} respectively, in the host galaxy rest frame. It had its most recent pericentre passage 446 Myr before this snapshot, with two disk crossings at 362 and 583 Myr; this is further discussed in Sect. 5.

This simulated galaxy is not intended to be a perfect reproduction of the Milky Way and Sagittarius satellite merger, or even the host or satellites themselves. For example, in the simulation's 'present day' the satellite is both too massive ($8 \times 10^9\, M_\odot$) and too close to the galactic centre compared to the Sgr remnant (e.g. Vasiliev & Belokurov 2020). Instead, it is intended to be a laboratory for studying satellite interaction with an otherwise stable disk, where any non-axisymmetric structure is induced by the satellite interaction.

In terms of phase-space spirals, the simulation has some crucial qualitative differences with respect to the actual Milky Way (as well as the one-dimensional simulations of Paper I). Although the simulation is very high resolution, the number of stellar particles is still roughly a factor of 300 smaller than the number of stars in the Milky Way, making it more difficult to resolve the phase-space spiral in small spatial volumes. Furthermore, the phase-space spiral seen in the three-dimensional simulation is present at greater vertical energies (i.e. greater vertical velocities and greater heights from the mid-plane). The main reason for this is that the simulation is comprised of particles representing stars and dark matter, but lacks a component of cold gas. In the solar neighbourhood, cold gas has a mid-plane matter density that is roughly equal that of stars (see e.g. McKee et al. 2015 and Schutz et al. 2018), although it has a significantly smaller scale height (roughly 100 pc). On large spatial scales, the cold gas contribution is not very significant, but it does matter for the formation of the phase-space spiral at smaller vertical energies. For a perturbation to wind into a phase-space spiral, the vertical oscillation period must vary with vertical energy, which requires an anharmonic gravitational potential. The small scale height of the cold gas component makes the vertical gravitational potential anharmonic at smaller heights, such that the winding of the spiral can occur at lower vertical energies.

The main purpose of this work was to test our method when applied to the complex dynamics of a three-dimensional simulation of a disturbed disk galaxy. Specifically, we sought to test our method's fundamental assumptions that the spiral inhabits a static and vertically separable gravitational potential. Due to the properties of the simulation and its phase-space spiral, as discussed above, we would expect any potential bias that could arise from these fundamental assumptions to be amplified: the strength of the perturbation poses a greater challenge to the assumption that the spiral evolves in a static gravitational potential; the spiral's presence at greater heights is more challenging of the assumption of vertical separability; the lower resolution and weaker statistics makes extracting the shape of the spiral less robust. However, as we shall see in Sect. 5, our method performed well and produced unbiased results despite these additional difficulties.

3. Data sample construction

We constructed the data samples analysed in this work from the simulation particles that represent stars. We divided the galaxy's disk plane into a grid in galactocentric longitude and galactocentric radius, with widths of 15 degrees and 500 pc and a radius range of 6–10 kpc. For each grid point, we selected a spatial volume centred on the grid point, extending 300 pc in the radial direction and 600 pc in the azimuthal direction. For each of these respective volumes, we limited ourselves to azimuthal angular momenta within plus or minus ten per cent of that data samples' mean value, similar to the data construction procedure in Paper II.

We then studied the individual data samples by eye and selected those where a well defined single-armed phase-space spiral was visible. Many disqualified data samples had a phase-space spiral structure which was not very clean, for example with multiple and sometimes fractured arms, likely related to multiple interactions with the orbiting satellite. In a few cases, the phase-space spiral had a clear single arm but with a thwarted shape that could not be reproduced by our fitting algorithm.

After this screening we had a total number of 44 data samples, whose locations in the disk plane are visible in Fig. 1, overlaid on the stellar surface number density. The data samples' locations have a similar disk surface density; the galaxy's more massive inner regions are less prone to being perturbed due to its stronger self-gravity, while the less massive outer regions are influenced by the satellite over longer time-scales which is less conducive to producing a well defined phase-space spiral. The bottom left quadrant of the disk (negative x and y) has almost no useful data samples; this region was most severely affected by the most recent passage of the satellite through the galactic disk.

[1] https://github.com/treecode/galactics.parallel
[2] https://flathub.flatironinstitute.org/jhunt2021

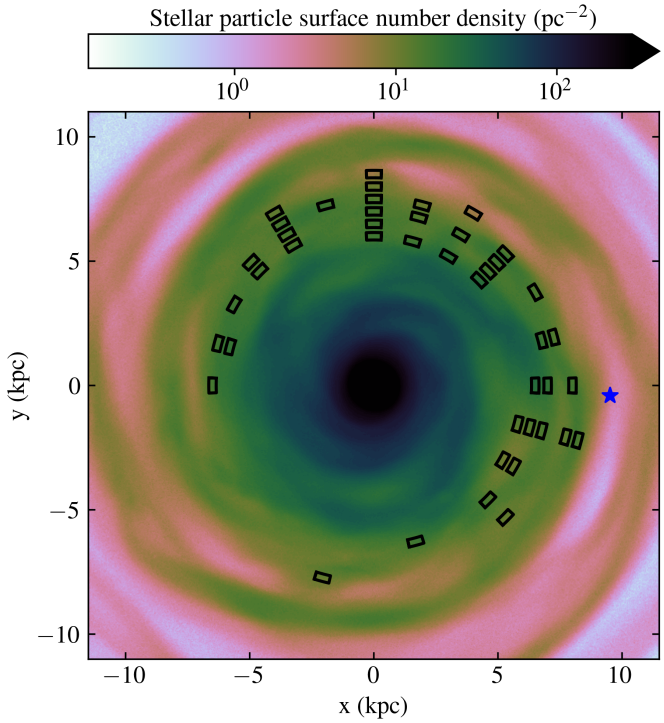

Fig. 1. Stellar surface density of the simulation snapshot, with overlaid black squares corresponding to the areas of our data samples. The star marks the location of the perturbing satellite in the (x, y)-plane.

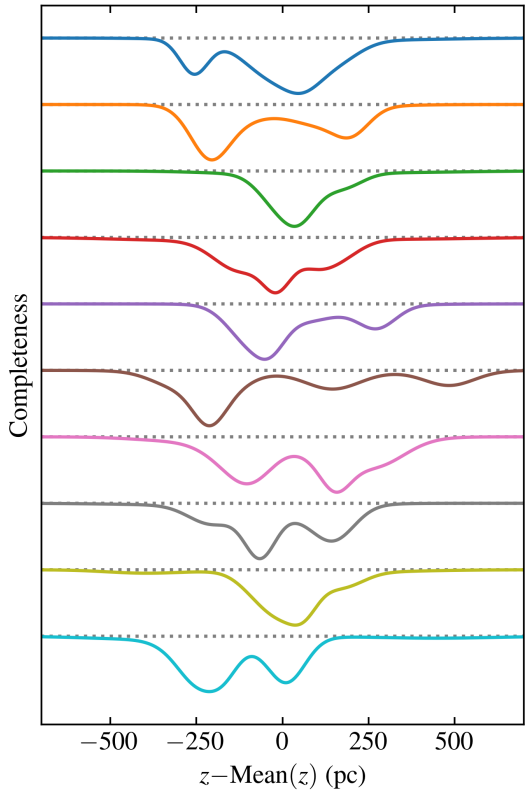

Fig. 2. Ten randomly generated extinction functions. They are shown in terms of their completeness as a function of height, where the dotted lines correspond to 100% and the bottom of the respective curves correspond to 50%.

We also performed additional tests under the influence of severe and spatially dependent selection effects, which can arise mainly due to stellar crowding and dust extinction close to the Galactic mid-plane, similar to what we saw in the *Gaia* data we analysed in Paper III. In order to demonstrate the robustness of our method under unknown selection effects, we perform tests where the simulation data was subjected to extinction. In these tests, our method of inference had no information about the precise form of this extinction; rather, the model of inference includes a simple mask model which was fitted in a data driven manner (also used in Paper III, see Sect. 4 for further details).

We randomly generated ten separate extinction functions. They are one-dimensional Gaussian mixture models proportional to

$$\mathrm{ext.}(z) \propto \sum_{n=1}^{8} a_n \frac{\exp\left[-\frac{(z - z_{\mathrm{ext.},n})^2}{2\sigma_n^2}\right]}{\sqrt{2\pi\sigma_n^2}}. \tag{1}$$

The eight Gaussian components are composed of two subgroups; for the first group of three Gaussian (and the second group of five Gaussians), a_n were generated from a uniform distribution in range 0.2–0.3 (0.6–1), $z_{\mathrm{ext.},n}$ were generated from a normal distribution centred on the respective data sample's mean value of z and a standard deviation of 200 pc (150 pc), and σ_n were generated from a uniform distribution in range 100–200 pc (40–80 pc). After all these parameters have been generated, the extinction function was normalised such that its maximum value corresponds to 50%.

The extinction masks are qualitatively similar to the selection effects observed in Paper III, in the sense of being asymmetric and a mix of multiple narrow and broad bands along the Z-axis. The selection effects present in the *Gaia* data set vary significantly, depending mainly on distance and Galactic longitude; in Paper III, some data samples were not visibly affected while others were rendered completely unusable. The extinction masks constructed in this work correspond to a middle ground, where the extinction masks are by far the dominant feature in the unprocessed data histogram, but the spiral shape can still be robustly extracted in the bulk density fit. The ten randomly generated extinction functions are shown in Fig. 2.

4. Model of inference

The model of inference used in this work is generally the same as was used in previous articles in this series. It is most akin to the version used in Paper III, although with some minor modifications in scale due to the spiral being present only at greater vertical energies in the three-dimensional simulation (see Sect. 2). Although the description in this section is complete, we refer back to previous papers for a more extensive explanation of our method.

Some of the details of our inference, such as the area of the (z, w)-plane where the spiral is fitted, depend on the standard deviations $\mathrm{Std}(z)$ and $\mathrm{Std}(w)$ for the specific data sample's stellar distribution. For our suite of data samples, these standard deviations are distributed according to 427 ± 23 pc and 22 ± 2 km s^{-1}.

When applying our method, we reduce the data to a two-dimensional histogram in height and vertical velocity. The phase-space density in our model of inference is a function of these two phase-space coordinates, equal to

$$f(z, w \,|\, \mathbf{\Psi}) = B(z, w \,|\, \mathbf{\Psi}_{\mathrm{bulk}}) \times \left[1 + m(z, w)\, S(z - \bar{z}, w - \bar{w} \,|\, \mathbf{\Psi}_{\mathrm{spiral}})\right], \tag{2}$$

where $B(z, w\,|\,\mathbf{\Psi}_{\rm bulk})$ is a bulk density distribution, $S(z-\bar{z}, w-\bar{w}\,|\,\mathbf{\Psi}_{\rm spiral})$ is a relative spiral density perturbation, and $m(z, w)$ is an inner mask function which quenches the spiral perturbation at low vertical energies. They depend on the free parameters $\mathbf{\Psi} = \{\mathbf{\Psi}_{\rm bulk}, \mathbf{\Psi}_{\rm spiral}\}$, which are listed in Table 1.

The bulk density distribution is a Gaussian mixture model equal to

$$B(z, w\,|\,\mathbf{\Psi}_{\rm bulk}) = \sum_{k=1}^{6} a_k \frac{\exp\left[-\dfrac{(z-\bar{z})^2}{2\sigma_{z,k}^2}\right]}{\sqrt{2\pi\sigma_{z,k}^2}} \frac{\exp\left[-\dfrac{(w-\bar{w})^2}{2\sigma_{w,k}^2}\right]}{\sqrt{2\pi\sigma_{w,k}^2}}, \quad (3)$$

where the Gaussians are constrained to be centred on the same point in the (z, w)-plane. Using this centre, we define a translated coordinate system according to

$$\begin{aligned} Z &\equiv z - \bar{z}, \\ W &\equiv w - \bar{w}. \end{aligned} \quad (4)$$

The spiral density is equal to

$$S(Z, W\,|\,\mathbf{\Psi}_{\rm spiral}) = \alpha \cos\Big[\varphi(Z, W\,|\,\rho_h) - \tilde{\varphi}(t, E_z\,|\,\rho_h, \tilde{\varphi}_0)\Big]. \quad (5)$$

In this expression, $E_z = \Phi + W^2/2$ is a vertical energy per mass. It depends on the vertical gravitational potential, which is modelled according to the following functional form,

$$\Phi(Z\,|\,\rho_h) = \sum_{h=1}^{4} 4\pi G \rho_h (2^{h-1} \times H)^2 \times \log\left[\cosh\left(\frac{Z}{2^{h-1} \times H}\right)\right], \quad (6)$$

where the ρ_h parameters are free to vary within the range $[0, 0.2]\,M_\odot\,{\rm pc}^{-3}$, and $H = 3/4 \times {\rm Std}(z)$ is proportional to the scale height of the data sample's stellar distribution. The choice of scale heights for the four matter density components differ from previous papers. Here, the scale heights are larger, but due to the lack of a cold gas component in the simulation this function can still faithfully reproduce the shape of the vertical gravitational potential.

The quantity φ in Eq. (5) is an angle of vertical oscillation defined like

$$\varphi(Z, W\,|\,\rho_h) = \begin{cases} 2\pi P^{-1} \displaystyle\int_0^{|Z|} \frac{{\rm d}Z'}{\sqrt{2[E_z - \Phi(Z'\,|\,\rho_h)]}} & \text{if } Z \geq 0 \text{ and } W \geq 0, \\ \pi - 2\pi P^{-1} \displaystyle\int_0^{|Z|} \frac{{\rm d}Z'}{\sqrt{2[E_z - \Phi(Z'\,|\,\rho_h)]}} & \text{if } Z \geq 0 \text{ and } W < 0, \\ \pi + 2\pi P^{-1} \displaystyle\int_0^{|Z|} \frac{{\rm d}Z'}{\sqrt{2[E_z - \Phi(Z'\,|\,\rho_h)]}} & \text{if } Z < 0 \text{ and } W < 0, \\ 2\pi - 2\pi P^{-1} \displaystyle\int_0^{|Z|} \frac{{\rm d}Z'}{\sqrt{2[E_z - \Phi(Z'\,|\,\rho_h)]}} & \text{if } Z < 0 \text{ and } W \geq 0. \end{cases} \quad (7)$$

The quantity $\tilde{\varphi}$ is the angle of the spiral as a function of vertical energy, according to

$$\tilde{\varphi}(t, E_z\,|\,\rho_h, \tilde{\varphi}_0) = \tilde{\varphi}_0 + 2\pi \frac{t}{P(E_z\,|\,\rho_h)}, \quad (8)$$

where P is the period of vertical oscillation, equal to

$$P(E_z\,|\,\rho_h) = \oint \frac{{\rm d}Z}{W}. \quad (9)$$

Table 1. Free parameters of our model.

$\mathbf{\Psi}_{\rm bulk}$	Bulk phase-space density parameters
a_k	Weights of the Gaussian mixture model
$\sigma_{z,k}$, $\sigma_{w,k}$	Dispersions of Gaussian mixture model
$\bar{z}, \bar{w}$	Mean height and vertical velocity
$\mathbf{\Psi}_{\rm spiral}$	Spiral phase-space density parameters
$\rho_{h=\{1,2,3,4\}}$	Mid-plane matter densities
t	Time since the perturbation was produced
$\tilde{\varphi}_0$	Initial angle of the perturbation
α	Relative density amplitude of the spiral
$\mathbf{\Psi}_{z-\rm mask}$	Mask in z (only used for extinction tests)
$\hat{a}_l$	Amplitudes
$\hat{z}_l$	Means
$\hat{\sigma}_{z,l}$	Dispersions

Notes. The third sub-group of parameters, written $\mathbf{\Psi}_{z-\rm mask}$, is only included for the tests where an extinction function is applied to the data.

The inner mask function is equal to

$$m(z, w) = {\rm sigm}\left\{10\left[\frac{(z - {\rm Mean}(z))^2}{(3/2 \times {\rm Std}(z))^2} + \frac{(w - {\rm Mean}(w))^2}{(3/2 \times {\rm Std}(w))^2} - 1\right]\right\}, \quad (10)$$

where

$${\rm sigm}(x) = \frac{1}{1 + \exp(-x)} \quad (11)$$

is a sigmoid function. The inner mask function is centred on the respective data samples' mean values of z and w, which can differ marginally, typically only by a few parsecs, from the fitted $\bar{z}$ and $\bar{w}$ values. The inner mask function covers a larger area of the (z, w)-plane as compared to previous articles in this series (here, denominators include a factor of 3/2). The reason for this choice is that the phase-space spiral is present at greater vertical energies in the three-dimensional simulation.

In the tests where an extinction function was applied to the data, as described in the end of Sect. 3, we also included a simple mask function as part of our model of inference, similar to how selection effects were modelled in Paper III. In such a case, the model's phase-space density would instead read

$$\begin{aligned} f(z, w\,|\,\mathbf{\Psi}) = {} & B(z, w\,|\,\mathbf{\Psi}_{\rm bulk}) \times \Xi(z\,|\,\mathbf{\Psi}_{z\text{-mask}}) \\ & \times \Big[1 + m(z, w)\, S(z, w - \bar{w}\,|\,\mathbf{\Psi}_{\rm spiral})\Big], \end{aligned} \quad (12)$$

where

$$\Xi(z\,|\,\mathbf{\Psi}_{z\text{-mask}}) = 1 - \sum_{l=1}^{6} \hat{a}_l \exp\left[-\frac{(z - \hat{z}_l)^2}{2\hat{\sigma}_{z,l}^2}\right]. \quad (13)$$

The free parameters of the mask function are written with hats, constrained to lie in the ranges $\hat{a}_l \in [0, 1]$, $\hat{z}_l - \bar{z} \in [-500, 500]$ pc, and $\hat{\sigma}_{z,l} \in [60, 300]$ pc. This mask function is fitted in a data driven manner and has no information about the precise form of the actual extinction function that was applied to the data prior to inference, apart from prior knowledge that extinction has a purely spatial dependence.

As for previous papers in this series, the minimisation algorithm is run in two separate steps, where the bulk density (without any spiral perturbation) is fitted in a first step, and the spiral

density (with a fixed bulk density) is fitted in a second step. The likelihood of the fit is written

$$\ln \mathcal{L}(\boldsymbol{d}_{i,j} \,|\, \boldsymbol{\Psi}) = -\sum_{i,j} M(z_i, w_i) \times \frac{[\boldsymbol{d}_{i,j} - f(z_i, w_i \,|\, \boldsymbol{\Psi})]^2}{2 f(z_i, w_i \,|\, \boldsymbol{\Psi})} + \{\text{constant term}\}, \tag{14}$$

where $\boldsymbol{d}_{i,j}$ is the two-dimensional data histogram with bin lengths of $0.08 \times \mathrm{Std}(z)$ and $0.08 \times \mathrm{Std}(w)$. The quantity

$$M(z, w) = \mathrm{sigm}\left\{ -10 \left[\frac{(z - \mathrm{Mean}(z))^2}{z_{\mathrm{lim.}}^2} + \frac{(w - \mathrm{Mean}(w))^2}{w_{\mathrm{lim.}}^2} - 1 \right] \right\} \tag{15}$$

is an outer mask function, centred on the data sample's mean values of z and w, just like for the inner mask function of Eq. (10). The values for its boundaries are $z_{\mathrm{lim.}} = 3.5 \times \mathrm{Std}(z)$ pc and $w_{\mathrm{lim.}} = 3.5 \times \mathrm{Std}(w)$ in the first step of our minimisation, and $z_{\mathrm{lim.}} = 3 \times 300$ pc and $w_{\mathrm{lim.}} = 3 \times \mathrm{Std}(w)$ in the second step. These values are larger in this work compared to previous papers, because the phase-space spiral is present at greater vertical energies in the three-dimensional simulation.

When running the tests where an extinction function was applied to the data, we fitted the mask function as part of the first step of the minimisation algorithm. For these tests, we fixed the value of $\bar{z}$ to the value that was obtained in the standard fit. The reason for doing so is that the parameter $\bar{z}$ cannot be robustly inferred in the presence of these severe extinction, so its value should be informed by some other measurement, which we in these tests assumed to be readily available information.

When running the tests where the height of the disk mid-plane was biased, we used the bulk density that was inferred in our standard fits. We only recomputed the second minimisation step, where the spatial position of the spiral's centre was shifted by some constant z_{bias}, according to $S(z - \bar{z} + z_{\mathrm{bias}}, w - \bar{w} \,|\, \boldsymbol{\Psi}_{\mathrm{spiral}})$.

5. Results

In this section, we show the results from having applied our method of inference to the 44 data samples we constructed. We also show the results for the cases when the data was subject to severe selection effects and when the height of the disk mid-plane was biased. When comparing our inferred results with the true potential, that true potential is given by the time snapshot in which the inference was performed, calculated separately for each respective data samples as an average over that data sample's area in the directions parallel to the disk plane (as seen in Fig. 1); hence, this true potential is free from any time averaging or large-scale spatial averaging over for example the azimuthal angle.

Due to the lower resolution of the three-dimensional simulation, we actually have less statistical power in this work as compared to when our model was applied to the Milky Way. For the data samples that we constructed from the three-dimensional simulation (as described in Sect. 3), there was on average 1.3×10^4 stellar particles in the region of the (z, w)-plane where the spiral density perturbation was fitted (as defined by the inner and outer mask functions in Sect. 4). In comparison, for the real data samples of Paper III the corresponding mean number of particles was twice as high, despite covering a significantly smaller spatial volume and being subject to severe selection effects.

In Fig. 3, we show the gravitational potential, data histogram, data residual with respect to the fitted bulk, and the fitted spiral for a representative data sample. The corresponding figures of a few more data samples are found in Appendix A. The fitted spirals seen in panel d are in good agreements with the shape seen in the data, most clearly in comparison with panel c. When comparing the inferred gravitational potential with the true potential of the simulation, as seen in panel a, there is an offset for data samples that are located close to the current position of the satellite. This is seen most clearly in Fig. A.3, which is one of the more extreme cases. This external force is not incorporated in our model of inference; in fact, such an external force cannot be inferred from the dynamics in a compact spatial volume alone, as it depends on the assumed boundary conditions. To avoid having to discard the data samples that are in the satellite's vicinity, we added a constant force, corresponding to a constant vertical acceleration field in the data sample's spatial volume, that symmetrises the true potential. Via the Poisson equation, such a constant acceleration field does not affect the underlying matter density field. In this manner, we isolated the gravitational potential that arises from the gravity of the galactic disk. This constant force was chosen such that the true gravitational potential values at $Z = \pm 3 \times \mathrm{Std}(z)$ become equal (although the precise choice of height is not significant for our end result). When we do this, we see a much stronger agreement with the inferred gravitational potential for the few data samples in the satellite's vicinity, as exemplified in Fig. A.3. For other data samples, this correction is negligible, as can be seen in Fig. 3 and other figures in Appendix A. In the remainder of this paper, when we compare the inferred and true gravitational potential, we apply this constant force correction.

For the inferred gravitational potential, our results are most accurate for the approximate height of 1 kpc (in terms of the potential difference with respect to the mid-plane). We show the relative errors for the inferred quantity $\Phi(1\,\mathrm{kpc})$ for our 44 data samples as a histogram in Fig. 4 and in terms of its distribution in the disk plane in Fig. 5. The corresponding plots for $\Phi(800\,\mathrm{pc})$ can be found in Appendix A. For $\Phi(1\,\mathrm{kpc})$, we had a relative accuracy of 7% with no indication of any systematic bias, and the errors did not have any strong spatial dependence. For $\Phi(800\,\mathrm{pc})$, we had a relative accuracy of about 8%, with a 2% bias towards positive errors. This indicates that our method is slightly biased towards gravitational potentials that are too steep close to the disk mid-plane (i.e. biased towards a matter density distribution that is too pinched). We also applied our method using a gravitational potential that was fixed to its true (although momentary) shape in the simulation snapshot. When doing so, the gravitational potential could vary only in terms of its normalisation. In this case we achieved a very similar result, which was non-biased and with an overall accuracy of 7%.

In Fig. 6, we show a histogram of the inferred parameter t, which in our model of inference corresponds to the time since the perturbation was produced. In Figs. A.9 and A.10, we show how these results are distributed in disk. From these figures, it is clear that the strongest outliers are found in the region that is closest to where the perturbing satellite is located. We can compare this with the most recent interactions between the satellite and the inner parts of the Galactic disk; due to its elliptical orbit, the satellite had two recent passages through the disk, at 362 and 583 Myr before the used simulation snapshot, and passed through its pericentre between them at 446 Myr. If we compare these values with the inferred times, they are in rough agreement. As we have discussed in previous papers in this series, see for example Paper I, the inferred time is highly degenerate with the precise shape of the inferred gravitational potential. If the inference is informed by a strong and correct prior for the shape of the gravitational potential, the time can be inferred with

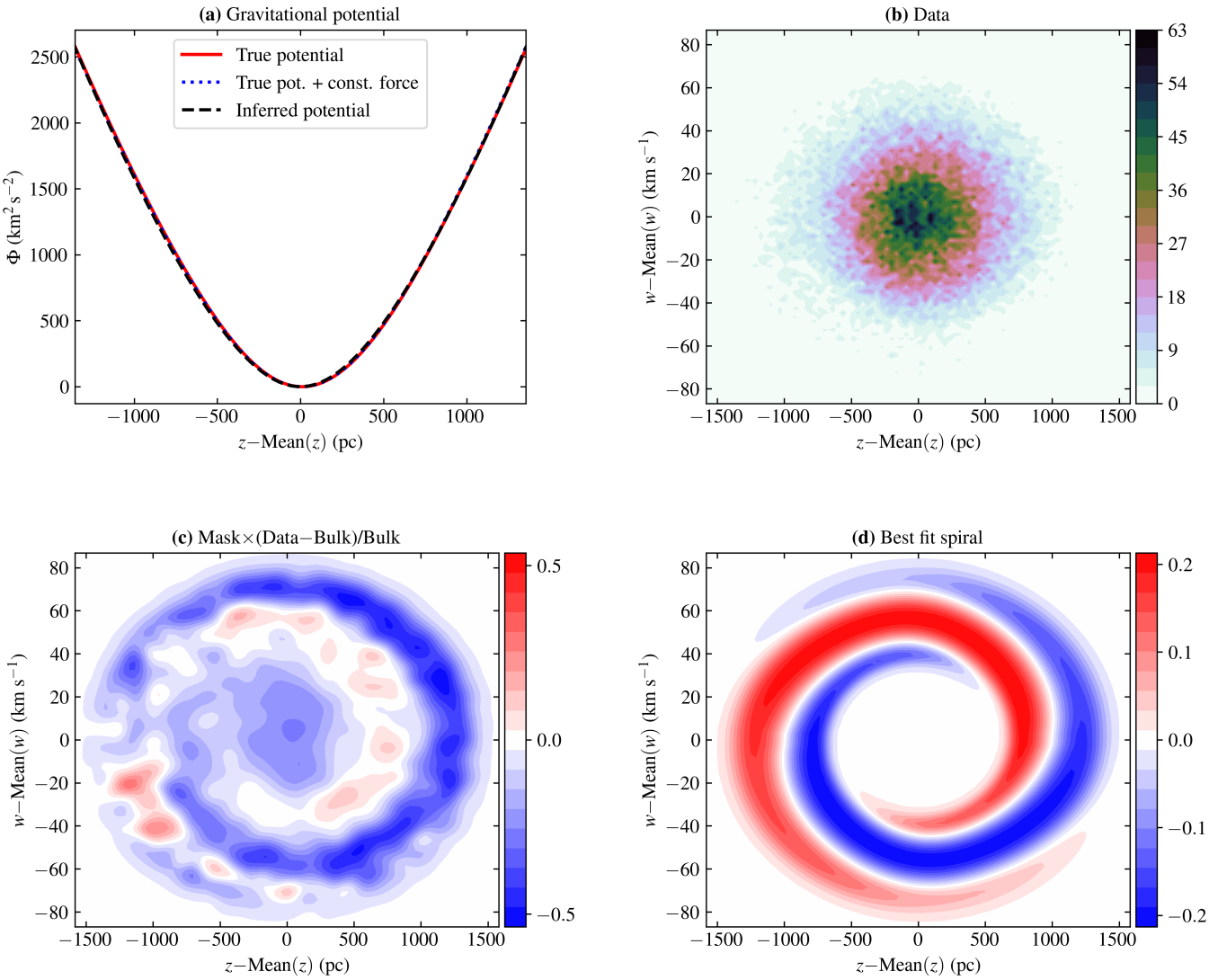

Fig. 3. Data and fitted phase-space density of the data sample with $\bar{l} = 165$ deg and $\bar{R} = 6500$ pc. The four panels show: (*a*) the data histogram; (*b*) the fitted bulk density; (*c*) the residual phase-space density with respect to the fitted bulk; (*d*) the relative phase-space density perturbation of the best fit spiral.

decent accuracy. This is clearly illustrated for the results where the gravitational shape is fixed to its true (although momentary) shape in the simulation snapshot, for which the inferred time is in good agreement with the satellite's most recent close interaction, especially when discounting the few strong outliers which are close at the satellite's current spatial position.

5.1. Extinction

We applied our ten extinction functions to six of our data samples and then applied our model of inference on these 60 separate cases. In Fig. 7, we show a few examples of what the data histogram and data residual with respect to the bulk can look like with extinction.

Our method of inference does not have any prior information about the precise form of the extinction function, apart from the knowledge that it depends solely on spatial position. In fact, the mask function which is part of our fitted model, as expressed in Eq. (13), is too simplistic to precisely model the true extinction function of Eq. (1). In this manner, we test the robustness of our method when incompleteness is severe but unknown and imperfectly modelled. There are degeneracies between the fitted bulk and mask function, which would be detrimental in a traditional dynamical mass measurement based on the assumption of a steady state. However, in our method the inferred gravitational potential is not informed by properties of the bulk density, and depends only on the shape of the extracted spiral. As can be seen in Fig. 7, the shape of the spiral is robustly inferred despite severe extinction.

The resulting biases for $\Phi(1\,\mathrm{kpc})$ are shown in Fig. 8. There is no strong correlation between the biases and the respective extinction functions. Rather, the size and sign of the bias is strongly correlated with the respective data samples. The strongest outlier is the data sample with $\bar{R} = 7500$ pc and $\bar{\phi} = 75$ deg. The data histogram and fitted phase-space distribution of this data sample can be seen in Fig. A.6 (for the standard fit without extinction); the fitted spiral does not perfectly reproduce the shape of the spiral seen in the data, which is somewhat skewed along the diagonal of the (z, w)-plane. Due to this feature, which our model cannot emulate, it seems that this data sample is especially sensitive to spatially localised selection effects. If we consider the biases of all six data samples and ten extinction masks together, we get a mean plus or minus standard deviation of $-0.5 \pm 4.2\%$.

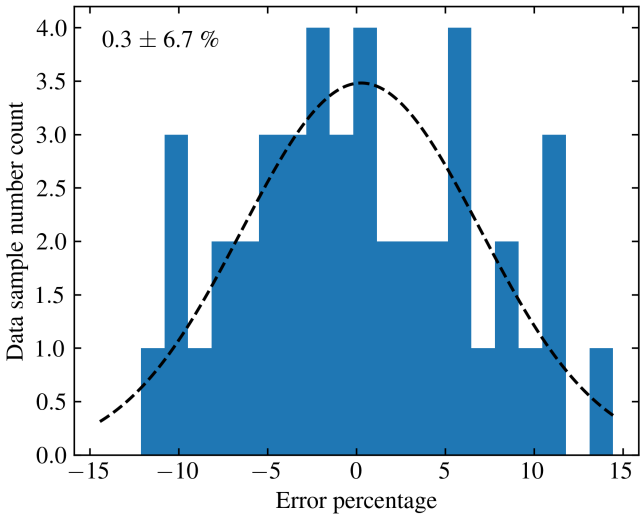

Fig. 4. Histogram of relative errors, for the quantity $\Phi(1\,\mathrm{kpc})$. The dashed black line corresponds to a Gaussian distribution whose mean and standard deviation is given by the distribution of relative errors, which are also written in the top left corner.

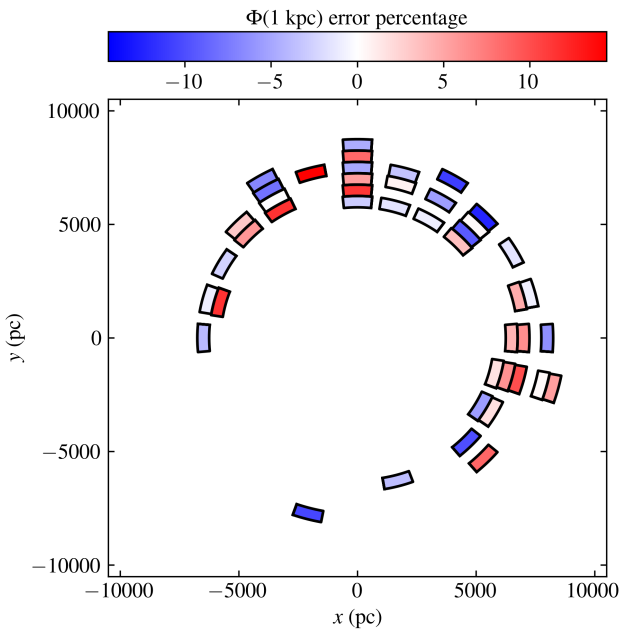

Fig. 5. Relative errors of our spiral method for the quantity $\Phi(1\,\mathrm{kpc})$. The areas of the respective data samples are somewhat inflated for better visibility, see Fig. 1 for their true size.

In summary, our method performs very well under the influence of severe selection effects and the end result is typically only affected by a few per cent, with no clear trend towards high or low values. This demonstrates that spatially dependent selection effects do not need to be precisely modelled; rather, they need to be modelled just well enough for the shape of the phase-space spiral to be extracted.

5.2. Biased height of the disk mid-plane

We tested our method's sensitivity to a biased mid-plane height by imposing a shift in the phase-space spiral's centre, by ±50 pc or ±100 pc. We did so for six data samples (the same six that were used in our extinction tests in Sect. 5.1), giving a total of 24 separate cases.

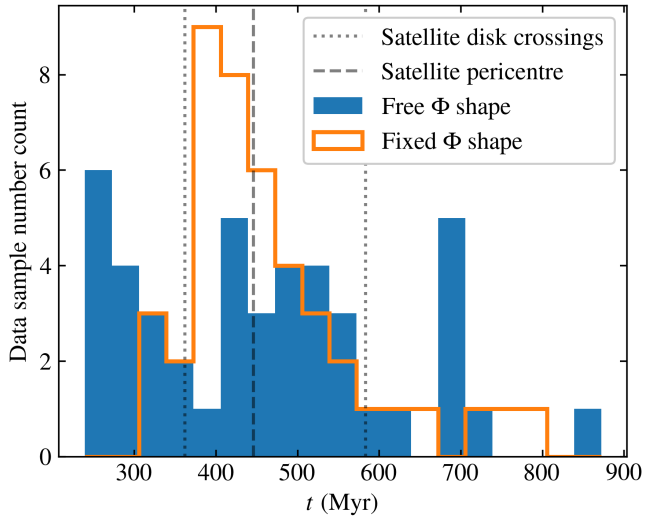

Fig. 6. Two histograms of the inferred parameter t. The two histograms correspond to fits where the shape of the vertical gravitational potential was either free to vary or fixed to its true shape (free only in terms of its normalisation). The dashed and dotted vertical lines mark the recent close passage of the simulation's satellite.

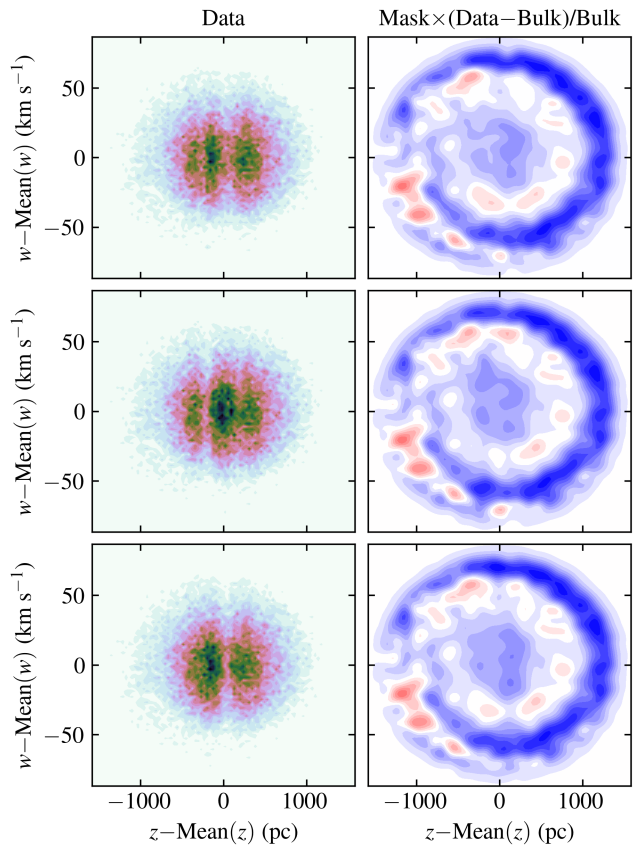

Fig. 7. Data histograms affected by extinction functions (which are described in Sect. 3). The three examples correspond to extinction functions indices 1–3 (from top to bottom), for the data sample with $\bar{l} = 165$ deg and $\bar{R} = 6500$ pc. *Left- and right-hand panels* correspond to panels b and c in Fig. 3 (although the "bulk" also includes the fitted mask function). The axis ranges are shared between all panels.

The resulting bias for the inferred quantity $\Phi(1\,\mathrm{kpc})$, with respect to the case with no added bias in the height of the disk mid-plane, can be seen in Fig. 9. For the height bias of ±50 pc (±100 pc), $\Phi(1\,\mathrm{kpc})$ is affected only minimally, with a mean

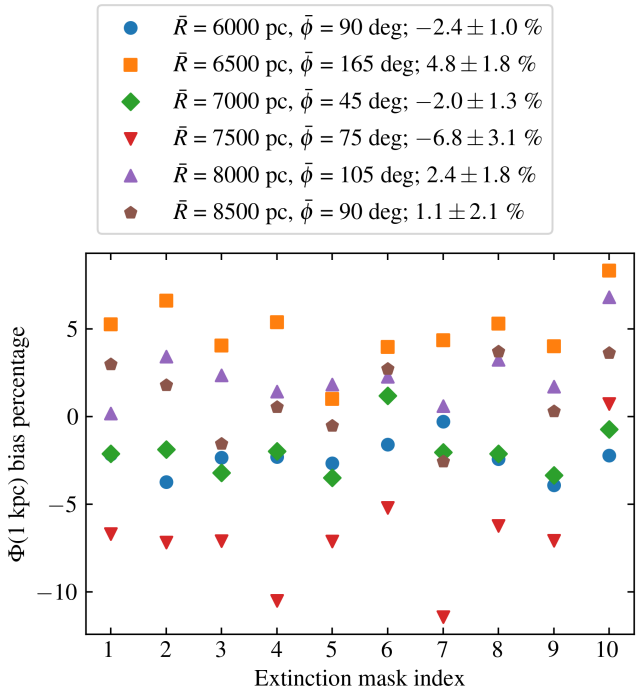

Fig. 8. Biases to the inferred quantity Φ(1 kpc) that arise when we apply strong selection effects to the data. The bias correspond to the change in Φ(1 kpc) with respect to the results with full data completeness. The legend shows the data samples' spatial coordinates as well as the bias mean plus or minus standard deviation.

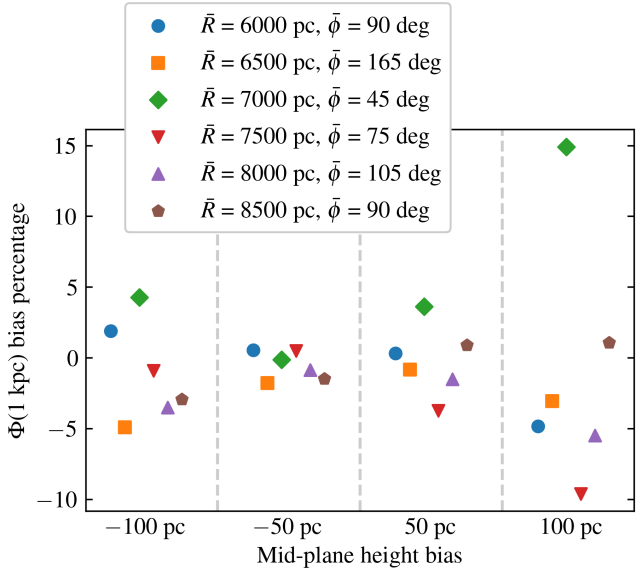

Fig. 9. Biases to the inferred quantity Φ(1 kpc) that arise when the height of the disk mid-plane is biased.

plus or minus standard deviation of $-0.4 \pm 1.7\%$ ($-1.1 \pm 6.0\%$). These values also differ between positive and negative height biases, probably related to the fact that the phase-space spiral itself is an asymmetric structure. The data histograms and fitted spirals of the two strongest outliers ($\bar{R} = 7000$ pc and $\bar{\phi} = 45$ deg, and $\bar{R} = 7500$ pc and $\bar{\phi} = 75$ deg) can be seen in Figs. A.5 and A.6. Both of these data samples have phase-space spirals that are somewhat skewed along the diagonal of the (z, w)-plane, potentially related to their spatial proximity to the perturbing satellite. This feature is not reproduced by our model of inference, which is probably related to why these two data samples are especially sensitive to systematic biases.

Our inferred results are stable, despite the large height biases. However, the scale of these biases does not translate perfectly to the case of the actual Milky Way. As is discussed in detail in Sect. 2, the simulation differs from the Milky Way in some significant regards, mainly due to the simulation's lower resolution and lack of a cold gas disk component, which makes the phase-space spiral present only at higher vertical energies (i.e. greater heights and vertical velocities). For the simulation studied in this work, the most accurate results for the vertical gravitational potential are found at a height of roughly 1 kpc, while for our one-dimensional simulations (Paper I) and the Milky Way (Paper II; Paper III), the most accurate results are found at roughly half that height. Thus the mid-plane height bias applied to the simulation in this work should probably be rescaled by a factor of roughly one half when considering the Milky Way.

6. Discussion

We have tested our new method for weighing the Galactic disk using phase-space spirals on a three-dimensional simulation. As discussed in Sect. 2, the properties of the simulation differ in some significant regards from the Milky Way, mainly in terms of a stronger external perturbation, lower resolution, less statistics, and a spiral that inhabits greater vertical energies. These simulation properties should be more challenging for our method, especially in terms of its fundamental assumptions of a vertically separable and static gravitational potential, likely amplifying statistical uncertainties and any potential biases in our inferred results. Despite these difficulties, we obtain unbiased results for the vertical gravitational potential with a relative accuracy of 7%.

Our fits produced an inferred parameter t, corresponding to the time since the perturbation that gives rise to the phase-space spiral was produced. This time can be compared to the satellite most recent pericentre passage (446 Myr before the simulation snapshot that we used) and its two most recent crossings through the host galaxy's disk (362 and 583 Myr). The inferred times agreed reasonably well with the range in time of the satellite's passages, especially for the case where the gravitational potentials of our model were fixed to their true shapes in the simulation snapshot. This illustrates how the inferred time is highly degenerate with the precise shape of the gravitational potential, which is not very robustly inferred relative to for example Φ(1 kpc). In our analysis, the gravitational potential has been free to vary in shape; however, going forward we might be better off by fixing the shape of the gravitational potential (or at least applying a more constrained prior), since not much additional information is gathered by letting it be free to vary.

We tested our method under severe spatially dependent selection effects, which produced biases of a few per cent. The extinction masks that we applied to the data were randomly generated in order to mimic the asymmetric incompleteness that arises due to stellar crowding and dust extinction close to the Galactic mid-plane. In these tests, we included a simple extinction function as part of our model of inference (as in Paper III), which had no prior knowledge about the true form of extinction mask. The selection effects are modelled jointly by the extinction function and the bulk density distribution, where the latter absorbs any incompleteness component that is smooth and reasonably symmetric across the galactic mid-plane. Therefore, the properties of the bulk is highly degenerate with selection, such that probably neither of them are accurately inferred.

This is not necessarily a problem but rather a strength of our method, because the bulk is not required to fulfil the collisionless Boltzmann equation, and the gravitational potential is extracted only from the shape of the phase-space spiral. This is a major qualitative difference with respect to traditional dynamical mass measurements that are based on the assumption of equilibrium, where the end result can only be as accurate as the modelling of the selection function. For our method using spirals, we should be able to apply it to great distances and depths and make strong cuts in for example data quality without having to carefully understand or model the relevant selection effects.

We also tested our method under a biased height of the disk mid-plane, by shifting the centre of the fitted spiral along the z-axis by ± 50 pc and ± 100 pc. In terms of $\Phi(1\,\mathrm{kpc})$, this produced biases that were mostly contained within a few per cent, with only a few more significant outliers. Overall, our method seems fairly robust with respect to a misunderstood height of the disk mid-plane, at least if the shift is smaller than one tenth of the height where the gravitational potential is measured. As we have seen, selection effects are not a severe obstacle to extracting the phase-space spiral, but could still be detrimental to determining the height of the disk mid-plane. For this reason, this height is better informed by a separate dedicated analysis, which can map the warping of the Galactic disk over a larger spatial area than is covered by our respective data samples.

In terms of the previous articles in this series, this work provides further support of our method and of our results based on *Gaia* data. In the near future, we plan to revisit and expand on the Milky Way analyses that we performed in Paper II and Paper III. In the first half year of 2022, *Gaia* will have its full third data release, which will contain a total of 33 million radial velocity measurements[3], as compared to the current 7.6 million. Furthermore, complementary distance information has been produced using photo-astrometric measurements by for example `StarHorse` (Anders et al. 2022), claiming a distance precision of roughly 3% even for objects with poor parallax measurements; this will allow us to construct data samples with softer cuts in parallax precision, which otherwise induces significant selection effects. With this additional information, we will be able to reach significantly greater depths and distances in the Milky Way's stellar disk, compared to the roughly 3 kpc distance reached in Paper III. Thanks to the robustness of our method, we should be able to weigh the Galactic disk at distances that would otherwise not be reachable with for example Jeans analysis.

7. Conclusion

We have applied our method for weighing the Galactic disk using phase-space spirals to a billion particle three-dimensional simulation of a Milky Way like host galaxy and a merging dwarf satellite. We constructed 44 separate data samples in different locations of the host galaxy's disk, where well defined spirals could be identified by eye. Despite having less statistics and a stronger satellite perturbation than is the case for the actual Milky Way, we obtained non-biased results with a relative accuracy of 7%. This validates the most important assumptions of our model of inference, which is that the spiral inhabits a vertically separable and static gravitational potential.

We also tested our method under severe and unknown spatially dependent selection effects, mimicking the incompleteness that can arise from stellar crowding and dust extinction. We obtained accurate results, demonstrating that our method is robust to such effects as long as the shape of the phase-space spiral is accurately extracted. This is a significant qualitative difference with respect to traditional techniques that are based on the assumption of a steady state (e.g. Jeans analysis), for which an accurate modelling of selection effects is crucial. Hence, our method is not only complementary in the sense that it extracts information from a time-varying dynamical structure and therefore subject to different systematic biases, it also has special merit in that it can be applied to distant regions of the Galactic disk where selection effects are difficult to model.

We will apply our method to future *Gaia* data releases, most imminently its full third data release, also supplemented with photo-astrometric measurements (e.g. `StarHorse`). With greater data depth and precision, we expect to make precise and localised mass measurements of the Galactic disk at even greater distances, further away than what can be reached with other methods.

Acknowledgements. AW acknowledges support from the Carlsberg Foundation via a Semper Ardens grant (CF15-0384). JH is supported by a Flatiron Research Fellowship at the Flatiron institute, which is supported by the Simons Foundation. CL acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 852839). GM acknowledges funding from the Agence Nationale de la Recherche (ANR project ANR-18-CE31-0006 and ANR-19-CE31-0017) and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 834148). This work made use of an HPC facility funded by a grant from VILLUM FONDEN (projectnumber 16599). This research utilised the following open-source Python packages: MATPLOTLIB (Hunter 2007), NUMPY (Harris et al. 2020), SCIPY (Virtanen et al. 2020), PANDAS (McKinney et al. 2010), TENSORFLOW (Abadi et al. 2015).

References

Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, `https://www.tensorflow.org`
An, J., Naik, A. P., Evans, N. W., & Burrage, C. 2021, MNRAS, 506, 5721
Anders, F., Khalatyan, A., Queiroz, A. B. A., et al. 2022, A&A, 658, A91
Antoja, T., Helmi, A., Romero-Gómez, M., et al. 2018, Nature, 561, 360
Bédorf, J., Gaburov, E., & Portegies Zwart, S. 2012, J. Comput. Phys., 231, 2825
Bédorf, J., Gaburov, E., Fujii, M. S., et al. 2014, Proceedings of the International Conference for High Performance Computing, 54
Dehnen, W., & Binney, J. 1998, MNRAS, 294, 429
de Salas, P. F., & Widmark, A. 2021, Rep. Prog. Phys., 84
Gaia Collaboration (Brown, A. G. A., et al.) 2018, A&A, 616, A1
Gaia Collaboration (Antoja, T., et al.) 2021, A&A, 649, A8
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357
Hernquist, L. 1990, ApJ, 356, 359
Hunt, J. A. S., Stelea, I. A., Johnston, K. V., et al. 2021, MNRAS, 508, 1459
Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90
Jungman, G., Kamionkowski, M., & Griest, K. 1996, Phys. Rep., 267, 195
Klasen, M., Pohl, M., & Sigl, G. 2015, Prog. Part. Nucl. Phys., 85, 1
Kuijken, K., & Dubinski, J. 1995, MNRAS, 277, 1341
Laporte, C. F. P., Johnston, K. V., Gómez, F. A., Garavito-Camargo, N., & Besla, G. 2018, MNRAS, 481, 286
Laporte, C. F. P., Minchev, I., Johnston, K. V., & Gómez, F. A. 2019, MNRAS, 485, 3134
McKee, C. F., Parravano, A., & Hollenbach, D. J. 2015, ApJ, 814, 13
McKinney, W. 2010, in Proceedings of the 9th Python in Science Conference, eds. S. van der Walt, & J. Millman, 56
McMillan, P. J. 2017, MNRAS, 465, 76
Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, A&A, 323, L49
Read, J. I. 2014, J. Phys. G. Nucl. Phys., 41
Schutz, K., Lin, T., Safdi, B. R., & Wu, C.-L. 2018, Phys. Rev. Lett., 121
Vasiliev, E., & Belokurov, V. 2020, MNRAS, 497, 4162
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nat. Methods, 17, 261
Wang, C., Huang, Y., Yuan, H. B., et al. 2019, ApJ, 877, L7
Widmark, A., Laporte, C., & de Salas, P. F. 2021a, A&A, 650, A124 (Paper I)
Widmark, A., Laporte, C., de Salas, P. F., & Monari, G. 2021b, A&A, 653, A86 (Paper II)
Widmark, A., Laporte, C. F. P., & Monari, G. 2021c, A&A, 663, A15 (Paper III)
Widrow, L. M., & Dubinski, J. 2005, ApJ, 631, 838
Widrow, L. M., Pym, B., & Dubinski, J. 2008, ApJ, 679, 1239

[3] `https://www.cosmos.esa.int/web/gaia/dr3`

Appendix A: Supplementary plots

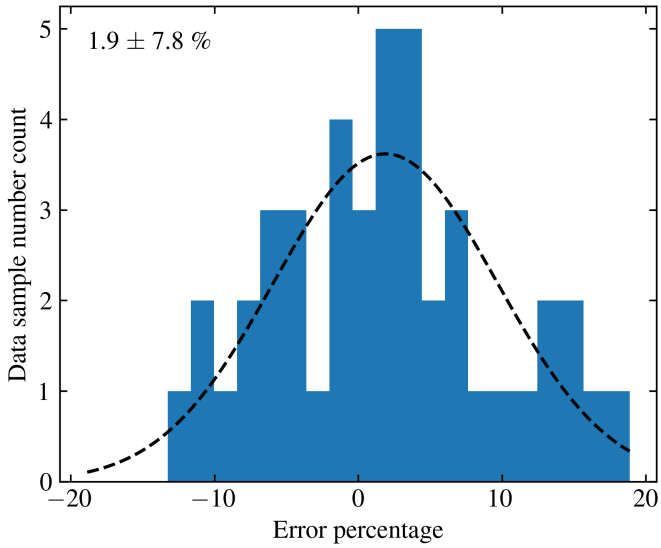

Fig. A.1. Histogram of relative errors, for the quantity Φ(800 pc). The dashed black line corresponds to a Gaussian distribution whose mean and standard deviation is given by the distribution of relative errors, which are also written in the top left corner.

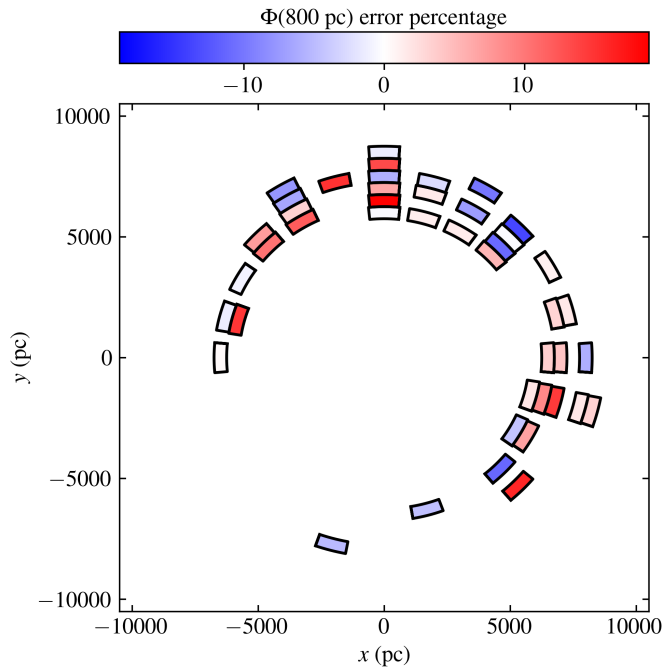

Fig. A.2. Relative errors of our spiral method for the quantity Φ(800 pc).

In this appendix we include a few supplementary plots. In Figs. A.1 and A.1, we show the relative errors for the inferred quantity Φ(800 pc), as a histogram and in terms of its distribution in the disk plane. In Figs. A.3–A.8, we show the gravitational potential, data histogram, data residual with respect to the fitted bulk, and the fitted spiral for six different data samples. The first is included as an example of a highly asymmetric gravitational potential, as a result of its proximity to the perturbing satellite. Finally, in Figs. A.9 and A.10, we show how the inferred time t is distributed in the disk plane, for the case when the shape of the vertical gravitational potential is either free to vary or fixed to its true shape in the simulation snapshot.

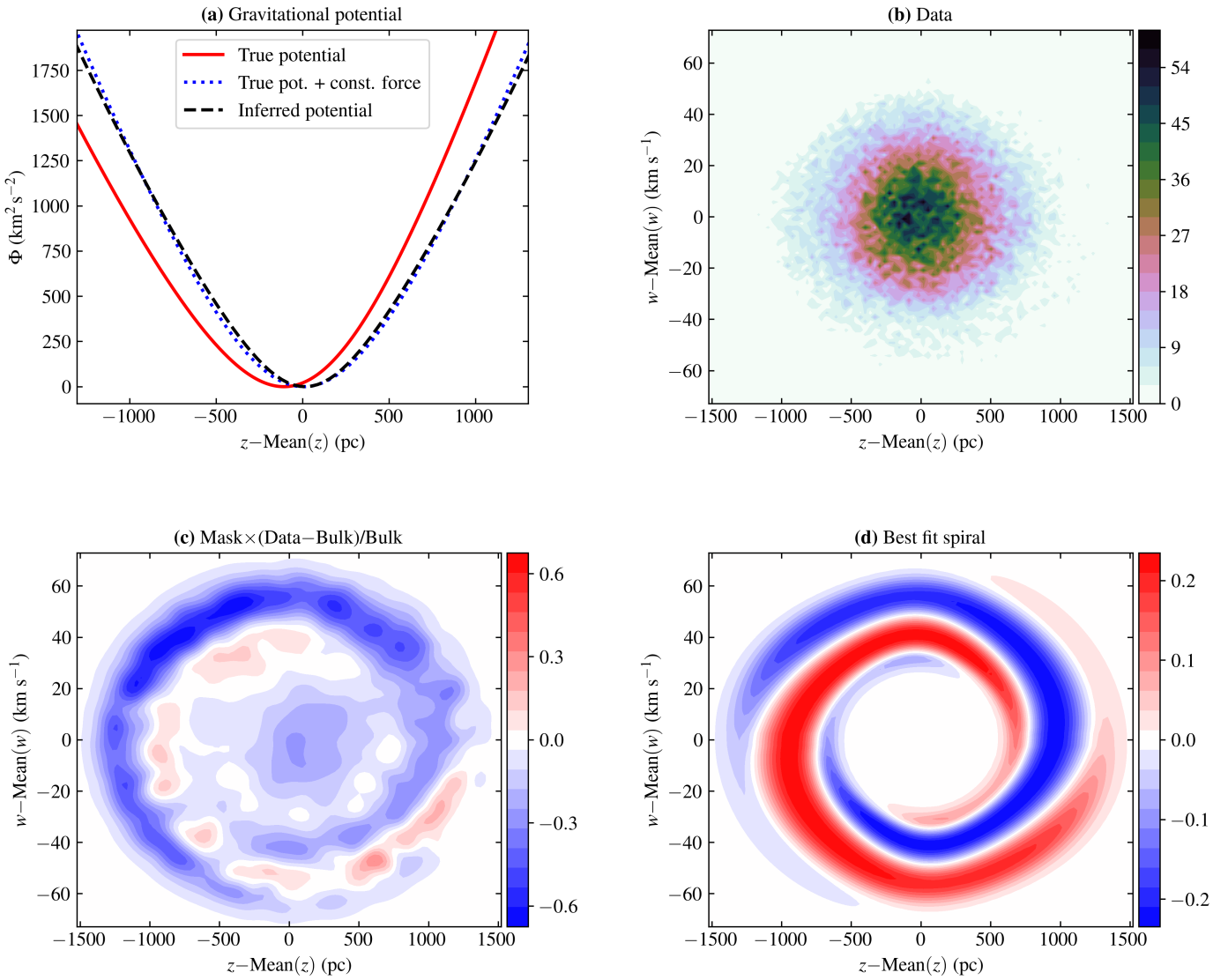

Fig. A.3. Same as for Fig. 3, but for the data sample with $\bar{l} = 30$ deg and $\bar{R} = 7500$ pc.

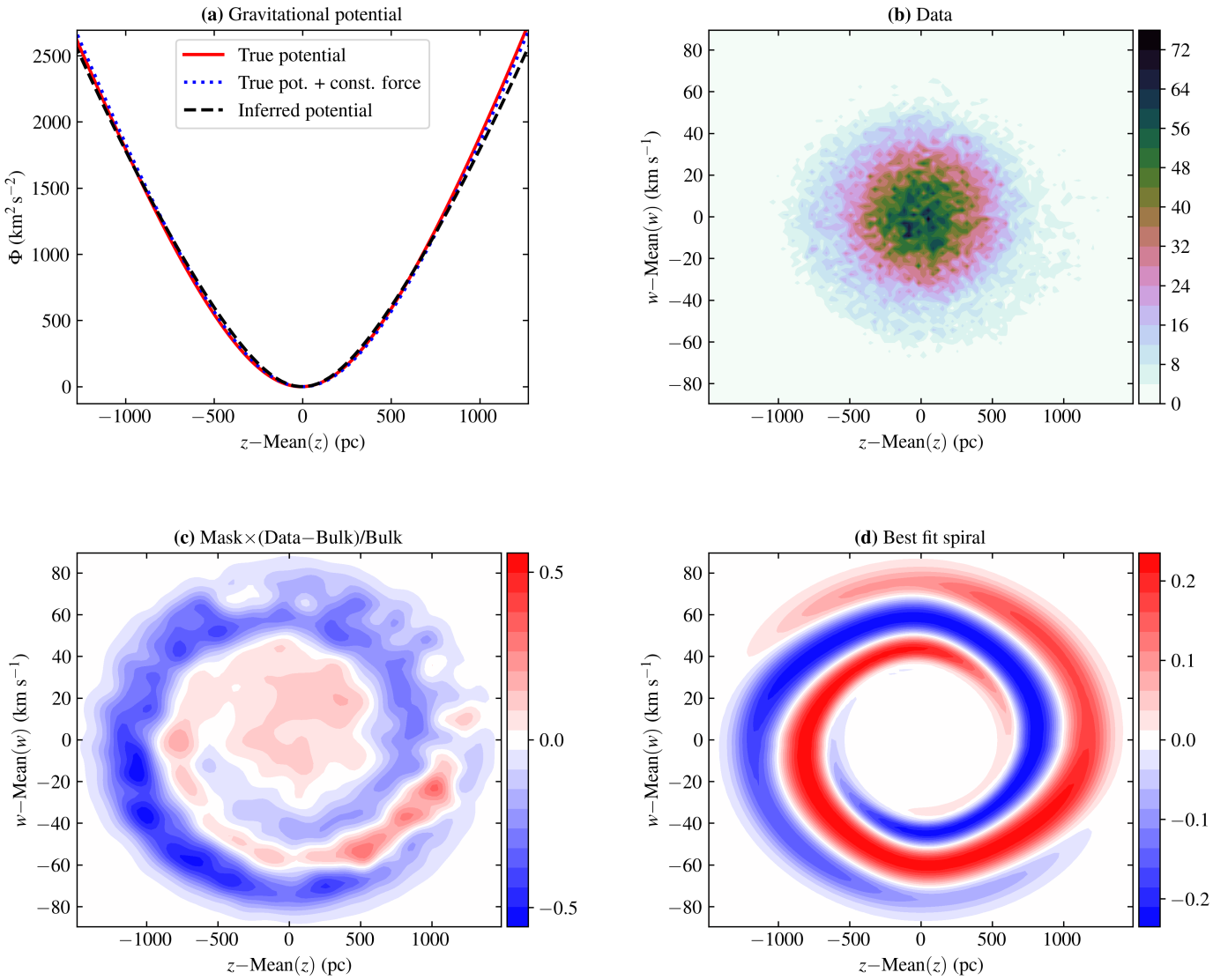

Fig. A.4. Same as for Fig. 3, but for the data sample with $\bar{l} = 90$ deg and $\bar{R} = 6000$ pc.

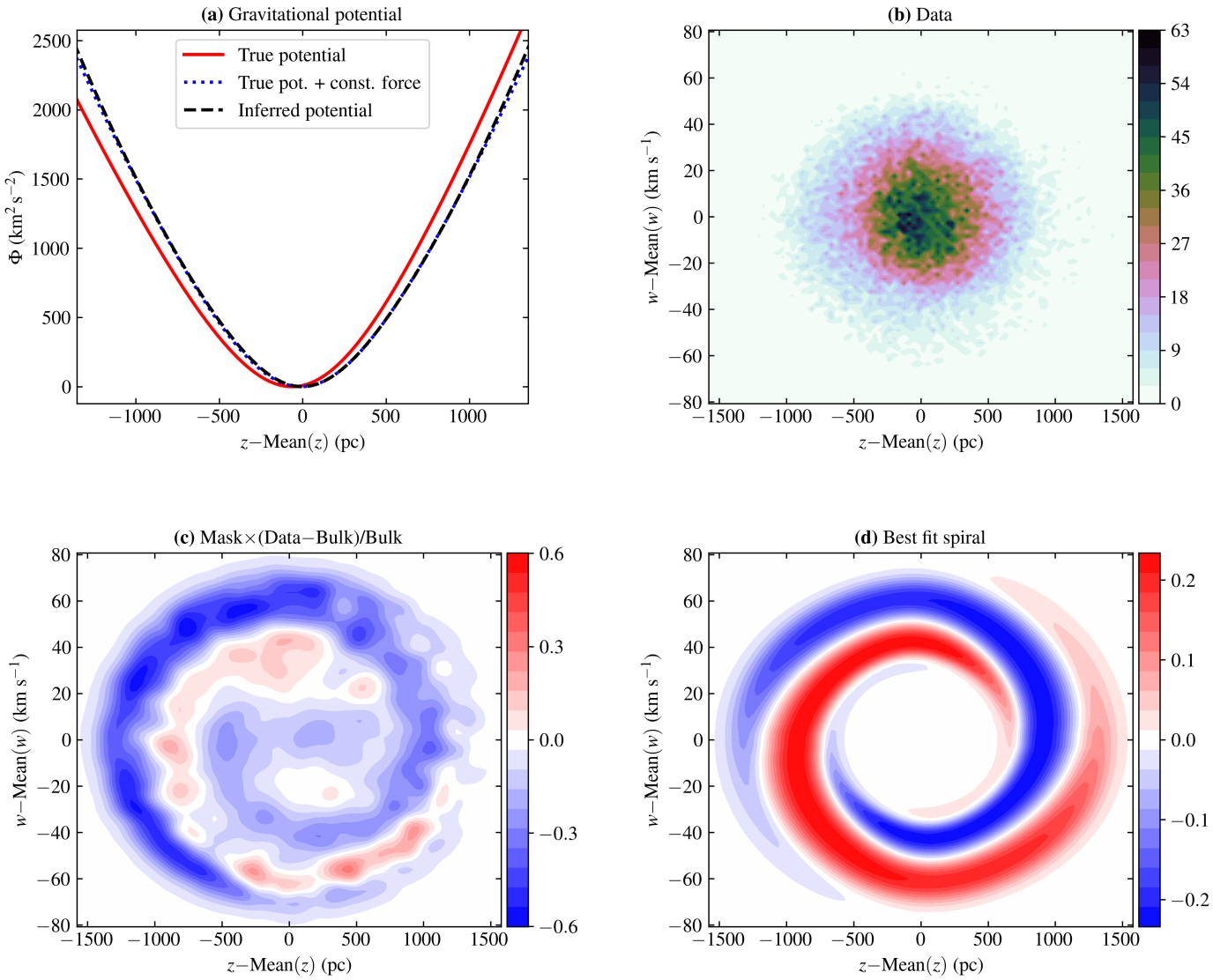

Fig. A.5. Same as for Fig. 3, but for the data sample with $\bar{l} = 45$ deg and $\bar{R} = 7000$ pc.

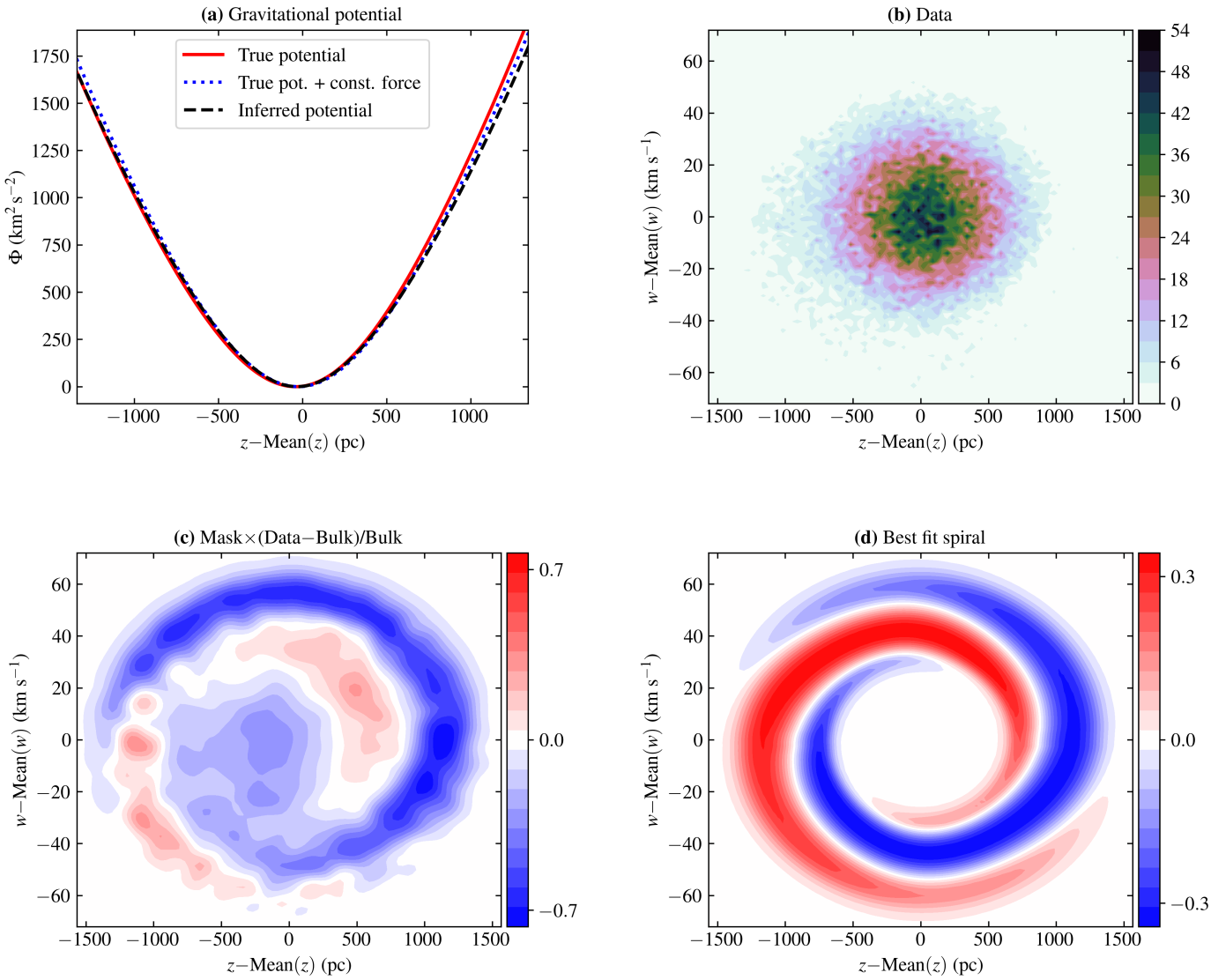

Fig. A.6. Same as for Fig. 3, but for the data sample with $\bar{l} = 75$ deg and $\bar{R} = 7500$ pc.

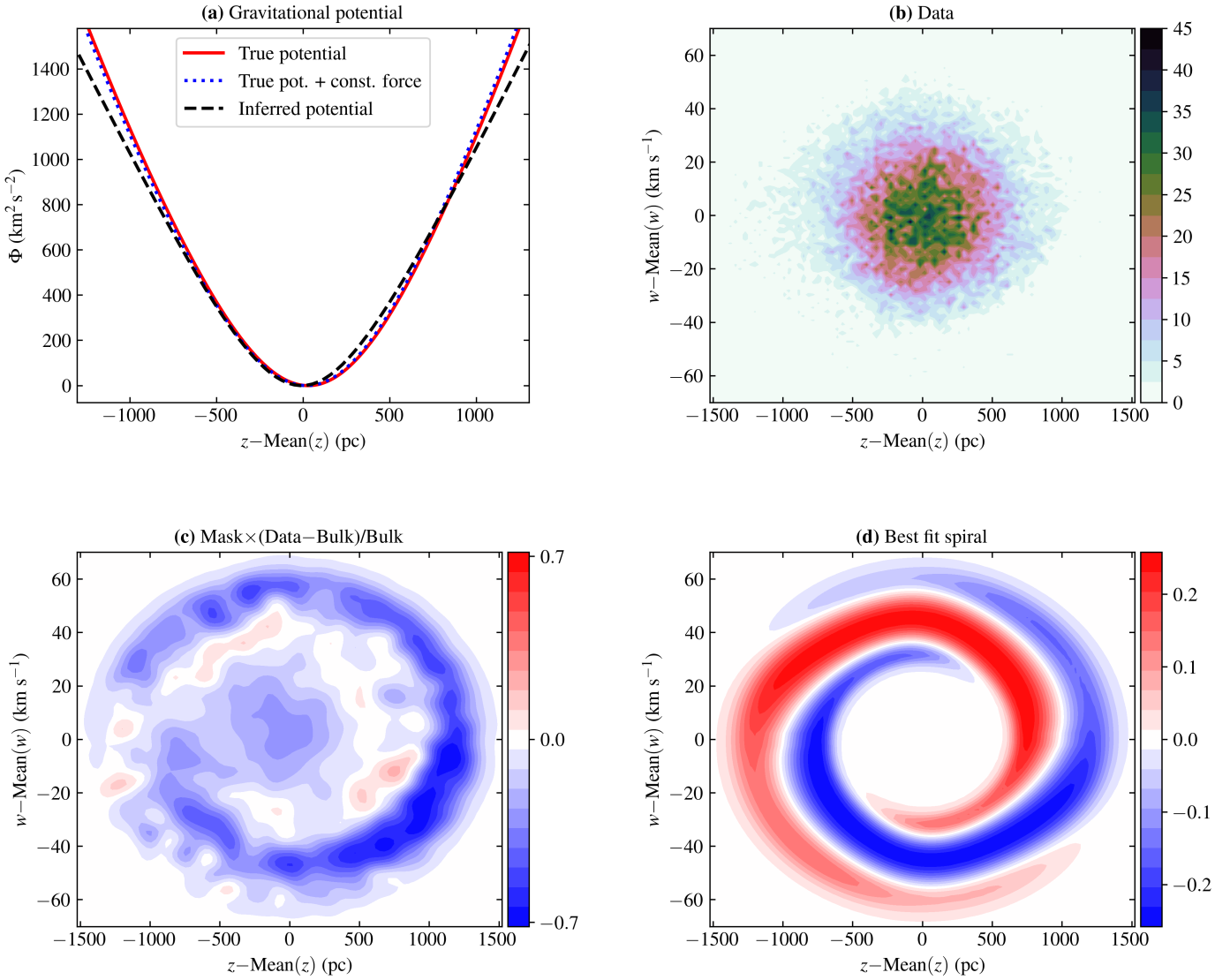

Fig. A.7. Same as for Fig. A.7, but for the data sample with $\bar{l} = 105$ deg and $\bar{R} = 8000$ pc.

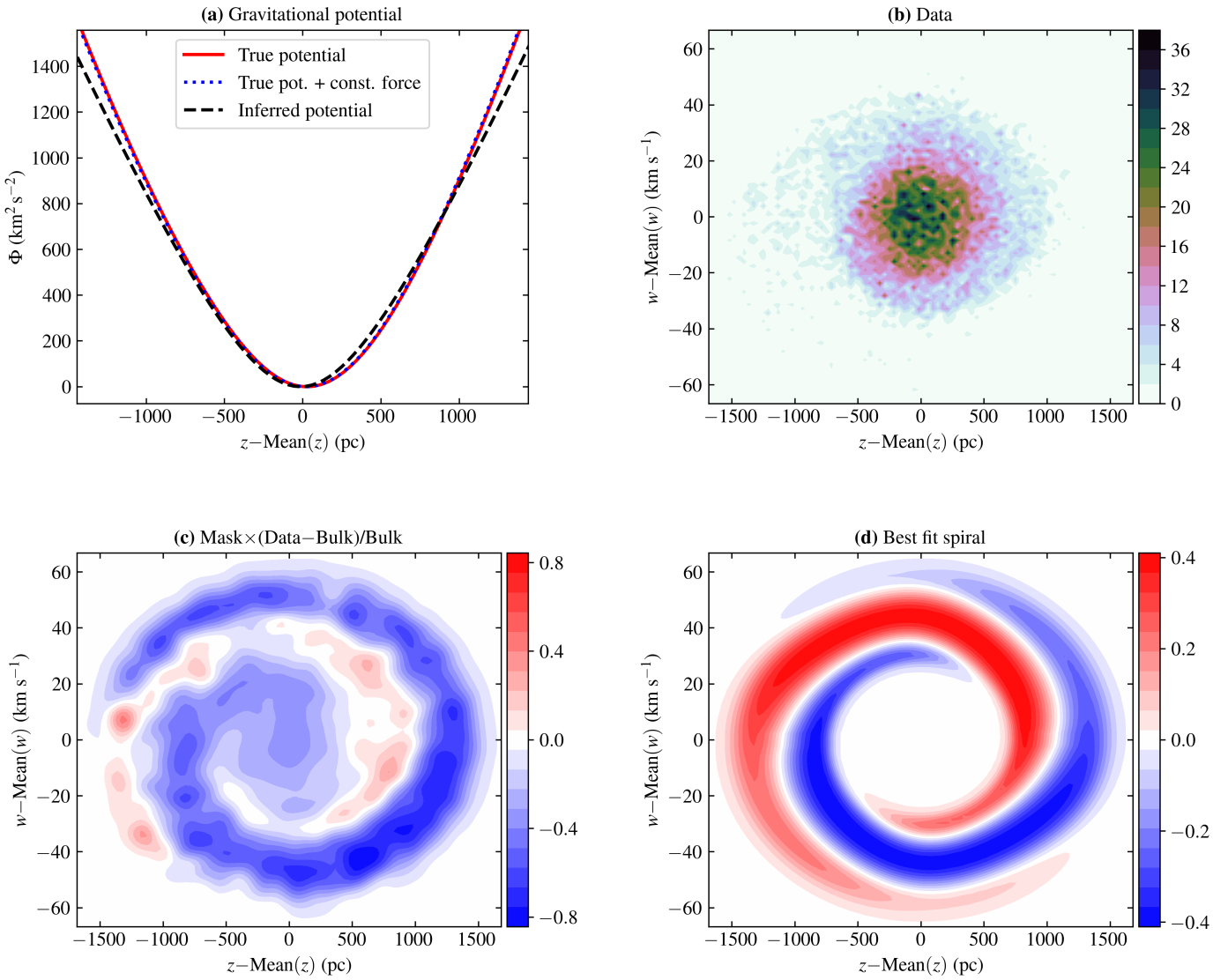

Fig. A.8. Same as for Fig. 3, but for the data sample with $\bar{l} = 90$ deg and $\bar{R} = 8500$ pc.

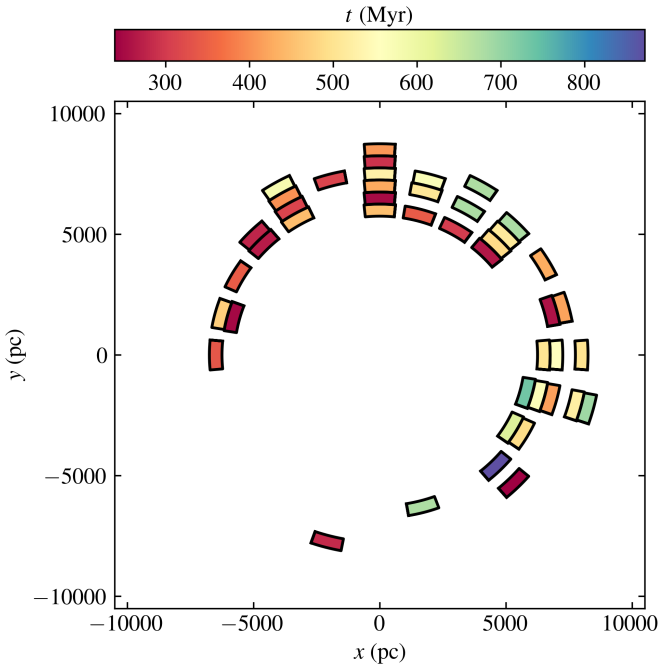

Fig. A.9. Spatial distribution of the inferred parameter t, for the case when the gravitational potential is free to vary in shape.

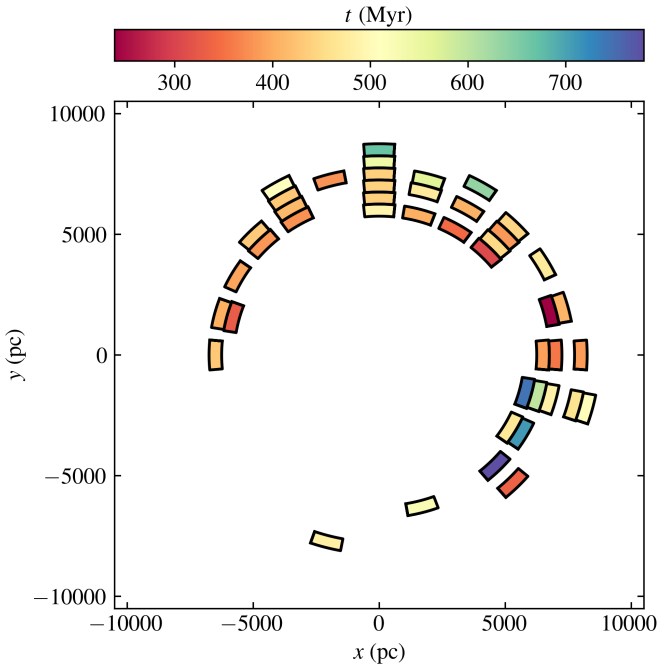

Fig. A.10. Spatial distribution of the inferred parameter t, for the case when the gravitational potential is fixed to its true shape and free to vary only in terms of its normalisation.