Draft version September 24, 2015
Preprint typeset using LaTeX style emulateapj v. 12/16/11

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

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

Draft version September 24, 2015

ABSTRACT

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

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

1. RESULTS

We are now in a position to examine the limitations of action-based modelling posed in the introduction (see Section **??**) using our *RoadMapping* machinery. We explore: (i) unbiased estimates, (ii) the role of the survey volume, (iii) imperfect selection functions, (iv) measurement errors and what happens if the actual (v) DF or (vi) Potential are not spanned by the space of models. We do not explore the breakdown of the assumption that the system is axisymmetric and in steady state. With the exception of the test suite on measurement errors in §1.2, we assume that phase-space errors are negligible. All tests are also summarized in Table 1.

[TO DO: Hans-Walter said that there are diagnostic plots in this papers that can be eliminated and their essence summarized in 1-2 sentences in the text. Fine. But which plots does he think can be eliminated? My plots contain either results or are only there to make the paper more readable for others.]

1.1. *Impact of Misjudging the Completeness of the Data Set*

The completeness function (see Section**??**) depends on the characteristics and mode of the survey. It can be very complex and is therefore sometimes not perfectly known. We investigate how much the recovery of the potential can be affected by imperfect knowledge of the selection function. We do this by creating mock data with varying incompleteness (within a maximal survey volume), while assuming constant completeness in the analysis. The mock data comes from a sphere around the sun with an incompleteness function that drops linearly with distance r from the sun (see Test 5, Example 1, in Table 1 and Figure 1). This captures the relevant case of stars being less likely to be observed (than assumed) the further away they are (e.g. due to unknown dust obscuration). We demonstrate that the potential recovery with *RoadMapping* is very robust against somewhat wrong assumptions about the radial completeness of the data (see Figure 2). Apparently, much information about the

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

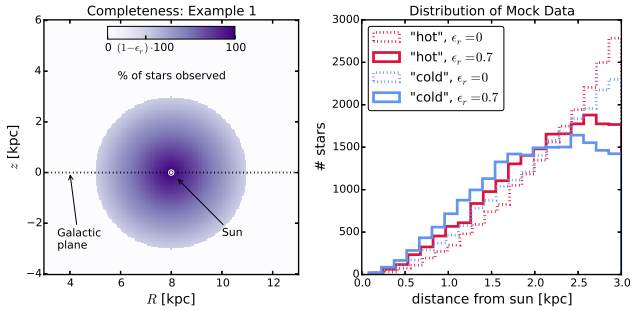

Figure 1. Selection function and mock data distribution for investigating radial incompleteness of the data. All model parameters are summarized as Test 5, Example 1, in Table 1. The survey volume is a sphere around the sun and the percentage of observed stars is decreasing linearly with radius from the sun, as demonstrated in the left panel. How fast this detection/incompleteness rate drops is quantified by the factor ϵ_r. Histograms for four data sets, drawn from two MAPs (`hot` in red and `cool` in blue, see Table **??**) and with two different ϵ_r, 0 and 0.7, are shown in the right panel for illustration purposes. [TO DO: Potential and/or population names in typewriter font]

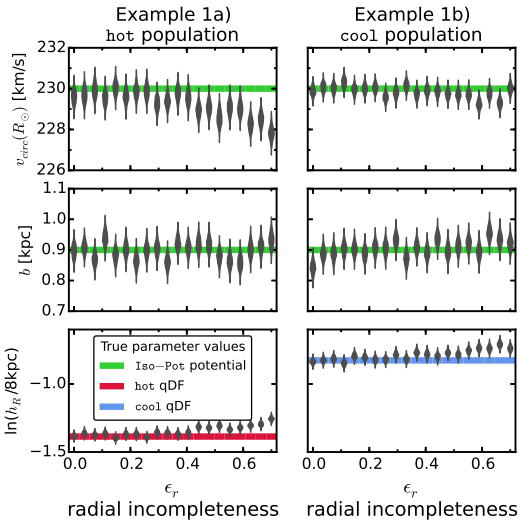

Figure 2. Influence of wrong assumptions about the radial incompleteness of the data on the parameter recovery with *RoadMapping*. Each mock data set was created with different incompleteness parameters ϵ_r (shown on the x-axis and illustrated in Figure 1) and the model parameters are given as Test 5, Example 1, in Table 1. The analysis however did not know about the incompleteness and assumed that all data sets had constant completeness within the survey volume ($\epsilon_r = 0$). The marginalized likelihoods from the fits are shown as violins. The green lines mark the true potential parameters (`Iso-Pot`) and the red and blue lines the true qDF parameters (`hot` MAPin red and `cool` MAPin blue), which we tried to recover. The *RoadMapping* method seems to be very robust against small to intermediate deviations between the true and the assumed data incompleteness. [TO DO: Jo suggested to also remove the h_R panel, but I like, that one can see that it is the spatial tracer distribution that drives the little degradation of the recovery.]

potential comes from the rotation curve measurements in the plane, which is not affected by the incompleteness of the sample. In Appendix 2.1 we also show that the robustness is somewhat less striking but still persists for small misjudgments of the incompleteness in vertical direction, parallel to the disk plane (Figures 6 and 7). This could model the effect of wrong corrections for interstellar extinction in the plane. We also investigate in Appendix 2.1 if indeed most of the information is stored in the rotation curve [TO DO: Comment by HW: I don't have an immediate solution for this, but again, it seems the interesting question of "how much of the information is in the rotation curve" is 'hidden' in the section on selection functions...]. For this we use the same mock data sets as analysed in Figures 2 and 7, but without including the tangential velocities in the modelling (by marginalizing the likelihood over v_T). In this case the potential is much less tightly constrained, even for 20,000 stars. For only small deviations of true and assumed completeness ($\lesssim 10\%$) we can however still incorporate the true potential in our fitting result (see Figure 8).

[TO DO: Mention in text or caption how the panels looked that I removed.]

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

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

In absence of distance uncertainties the error convolved likelihood given in Equation **??** is unbiased. When including distance (modulus) errors, Equation **??** is just an approximation for the true likelihood. The systematic bias thus introduced in the parameter recovery gets larger with the size of the error. This is demonstrated in Figure 6.2. We find however that in case of $\delta(m-M) \lesssim 0.3$ mag (if also $\delta\mu \leq 2$ mas yr^{-1} and a maximum distance of $r_{\rm max} = 3$ kpc, see Test 6.2 in Table 1) the potential parameters can still be recovered within 2 sigma [TO DO: Make sure this is what I claim in abstract and discussion.]. This corresponds to a relative distance error of $\sim 10\%$.

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

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

Figure 5 now investigates the effect of a systematic *underestimation* of the true proper motion uncertainties $\delta\mu$ by 10% and 50%. We find that this causes a bias in the parameter recovery that grows seemingly linear with

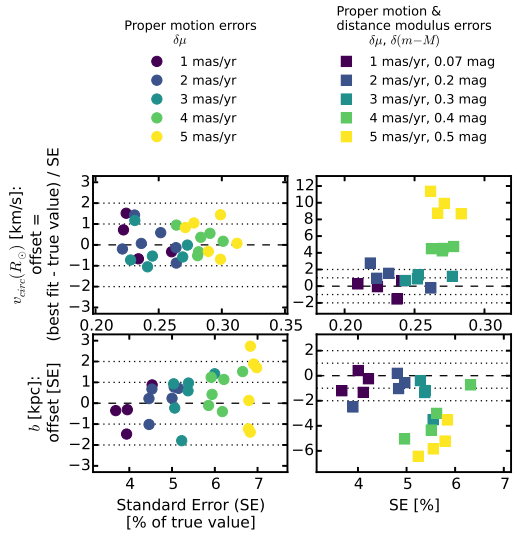

Figure 3. Potential parameter recovery using the approximation for the model probability convolved with measurement uncertainties in Equation **??**. We show *pdf* offset and relative width (i.e., standard error SE) for the potential parameters derived from mock data sets, which were created according to Test 6.2 in Table 1). The data sets in the left panels have only uncertainties in line-of-sight velocity and proper motions, while the data sets in the right panels also have distance (modulus) uncertainties, as indicated in the legends in the first row. For data sets with proper motion error errors $\delta(m-M) \leq 3$ mas yr^{-1} Equation **??** was evaluated with $N_{\rm error} = 800$, for $\delta(m-M) > 3$ mas yr^{-1} we used $N_{\rm error} = 1200$. In absence of distance uncertainties Equation **??** gives unbiased results. For $\delta(m-M) \geq 3$mas yr^{-1} (which corresponds in this test to $\delta v_{\rm max} \gtrsim 43$ km s^{-1}, see Equation **??**) however biases of several sigma are introduced as Equation **??** is only an approximation for the true likelihood in this case. [TO DO: No units at b and vcirc on y axis]

$\delta\mu$. For an underestimation of only 10% however, the bias is still $\lesssim 2$ sigma for 10,000 stars [TO DO: Check] - even for $\delta\mu \sim 3$ mas yr^{-1}.

The size of the bias also depends on the kinematic temperature of the stellar population and the model parameter considered (see Figure 5). The qDF parameters are for example better recovered by hotter populations. This is, because the *relative* difference between the true $\sigma_i(R)$ (with $i \in \{R, z\}$) and measured $\sigma_i(R)$ (which comes from the deconvolution with an underestimated velocity uncertainty) is smaller for hotter populations.

[TO DO: Comment from Jo: Always use 'uncertainty' when describing how ou deal with the errors. 'Error' means the actual error (difference between observed and true).]

1.3. *The Impact of Deviations of the Data from the Idealized qDF*

Our modelling approach assumes that each MAP follows a quasi-isothermal distribution function, qDF. In this Section we explore what happens if this idealization does not hold. We investigate this issue by creating mock data sets (Figure **??**) that are drawn from two distinct qDFs of different temperature, and analyze the composite mock data set by fitting a single qDF to it. These results are illustrated in Figures **??** and **??**. Following the observational evidence, MAPs with cooler qDFs also have longer tracer scale lengths. In the first set of test, we choose qDFs of widely different temperatures and vary

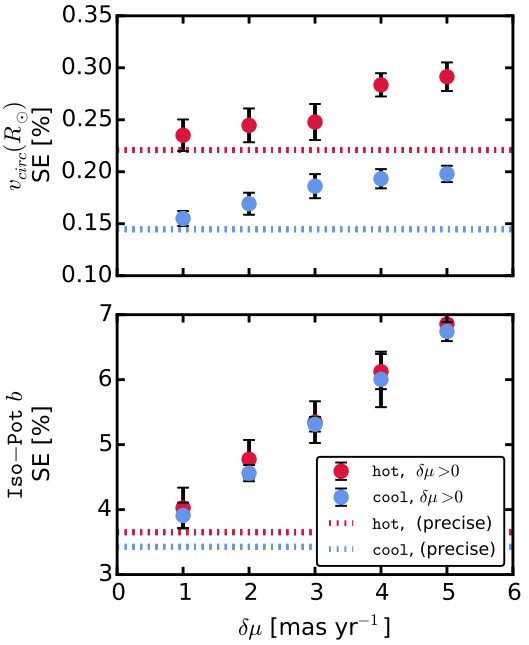

Figure 4. [TO DO: This should be a figure that plots precision (SE) vs. proper motion error for a hot and a cool population (for no distance error). This is to demonstrate the effect of measurement errors in general. Currently running on cluster....]

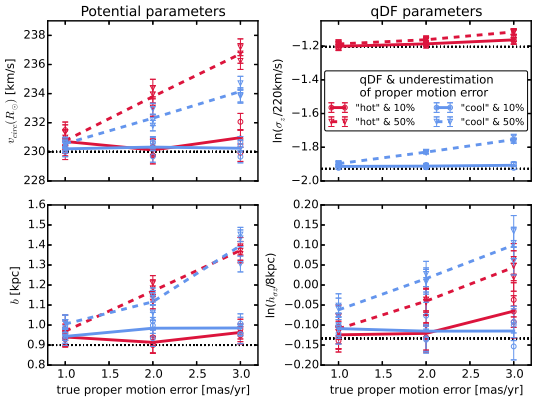

Figure 5. Effect of a systematic underestimation of proper motion errors in the recovery of the model parameters. The true model parameters used to create the mock data are summarized as Test 6.3 in Table 1, four of them are given on the y-axes and the true values are indicated as black dashed lines. The velocities of the mock data were perturbed according to Gaussian errors in the RA and DEC proper motions as indicated on the x-axis. The circles and triangles are the best fit parameters of several mock data sets assuming the proper motion uncertainty, with which the model probability was convolved, was underestimated in the analysis by 10% or 50%, respectively. The error bars correspond to 1 sigma confidence. The lines connect the mean of each two data realisations and are just to guide the eye. [TO DO: rename $h_{\sigma z}$ to $h_{\sigma,z}$, σ_z to $\sigma_{z,0}$] [TO DO: Potential and/or population names in typewriter font] [TO DO: Iso-Pot in Title] [TO DO: Delta mu on x-axis]

their relative fraction (dubbed *Examples 1a/b* in Figure **??** and Test 7 in Table 1); in the second set of tests (*Examples 2a/b* in Figure **??** and Test 7in Table 1), we always mix mock data points from two different qDFs in equal proportion, but vary by how much the qDF's temperatures differ.

The first set of tests mimics a DF that has wider wings or a sharper core in velocity space than a qDF (Figure **??**). The second test could be understood as mixing neighbouring MAPs due to large bin sizes or abundance measurement errors.

It is worth considering the impact of the DF deviations on the recovery of the potential and of the qDF parameters separately. We find from Example 1 that the potential parameters can be better and more robustly recovered, if a mock-data MAP is polluted by a modest fraction ($\lesssim 30\%$) of stars drawn from a much cooler qDF with a longer scale length, as opposed to the same pollution of stars drawn from a hotter qDF with a shorter scale length.

When considering the case of a 50/50 mix of contributions from different qDFs in Example 2, there is a systematic, but only small, error in recovering the potential parameters, monotonically increasing with the qDF parameter difference; in particular for fractional differences in the qDF parameters of $\lesssim 20\%$ the systematics are insignificant even for samples sizes of 20,000, as used in the mock data.

Overall, mock data drawn from a cooler DF always seem to give tighter constraints on the circular velocity at the Sun [TO DO: Make sure that Sun is written everywhere with a capital S.], because the rotation curve can be constrained easier if more stars are on near-circular orbits. But we found the recovered $v_{\rm circ}(R_\odot)$ not always to be unbiased at the implied precision.

The recovery of the effective qDF parameters, in light of non-qDF mock data is quite intuitive: the effective qDF temperature lies between the two temperatures from which the mixed DF of the mock data was drawn; in all cases the scale length of the velocity dispersion fall-off, $h_{\sigma,R}$ and $h_{\sigma,z}$, is shorter, because the stars drawn form the hotter qDF dominate at small radii, while stars from the cooler qDF (with its longer tracer scale length) dominate at large radii. The recovered tracer scale lengths, h_R vary smoothly between the input values of the two qDFs that entered the mix of mock data, with again the impact of contamination by a hotter qDF (with its shorter scale length in this case) being more important. Overall, we find that the potential inference is quite robust to modest deviations of the data from the assumed DF.

2. DISCUSSION AND SUMMARY

[TO DO: Introduce DF somewhere - use DF wherever we don't need qDF.] [TO DO: Introduce MW somewhere.]

[TO DO: Compare these sections with the results. Points should be made detailed in the results section and short here in the discussion. Says Hans-Walter.]

[TO DO: Abstze mit Indent. Keine Leerzeilen.]

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

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

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

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

Properties of the data set: We could show that *RoadMapping* can provide potential and DF parameter estimates that are very accurate (i.e. unbiased) and precise in the the limit of large datasets, as long as the modelling assumptions are fulfilled. We also found that the *location* of the survey volume within the Galaxy matters little. At given sample size a larger survey *volume* with large coverage in *both radial and vertical* direction will give the tightest constraints on the model parameters. [TO DO: Finished up to here. Continue here.]

Stellar populations of different scale length and temperature probe different regions of the Galaxy (BR13). But there is no easy rule of thumb for which survey volume and stellar population which potential and DF parameter is constrained best.

Surprisingly, (cf. Rix & Bovy 2013) *RoadMapping* seems to be very robust against misjudgments in the selection function of the data. We speculate that this is because missing stars in the data set do not affect the connection between a star's velocity and position, which is given by the potential. Much of the information about the potential profile is stored in the rotation curve, but we find that even when when we do not include measurements of tangential velocities in the analysis, small misjudgments of the incompleteness do not affect the potential recovery.

[TO DO: Comment from HW: Author: rix Subject: This paragraph shouldbe 1 or 2 sentences, following the first paragraph on "Sample/Data Properties". This – at the moments –reads to bequite confusing. I don't quite get whatthe "upshot" is; there is technical detai on N_{error} [enought to say it's expensive]; and, as noted earlier; I don't understand why theerror convolution for a nearbydata point needs to know about $\delta v_{\rm max}$]

Properly convolving the likelihood with measurement errors is computational very expensive. By ignoring positional errors and only including distance errors as part of the velocity error, we can drastically reduce the computational costs. For stars within 3 kpc from the

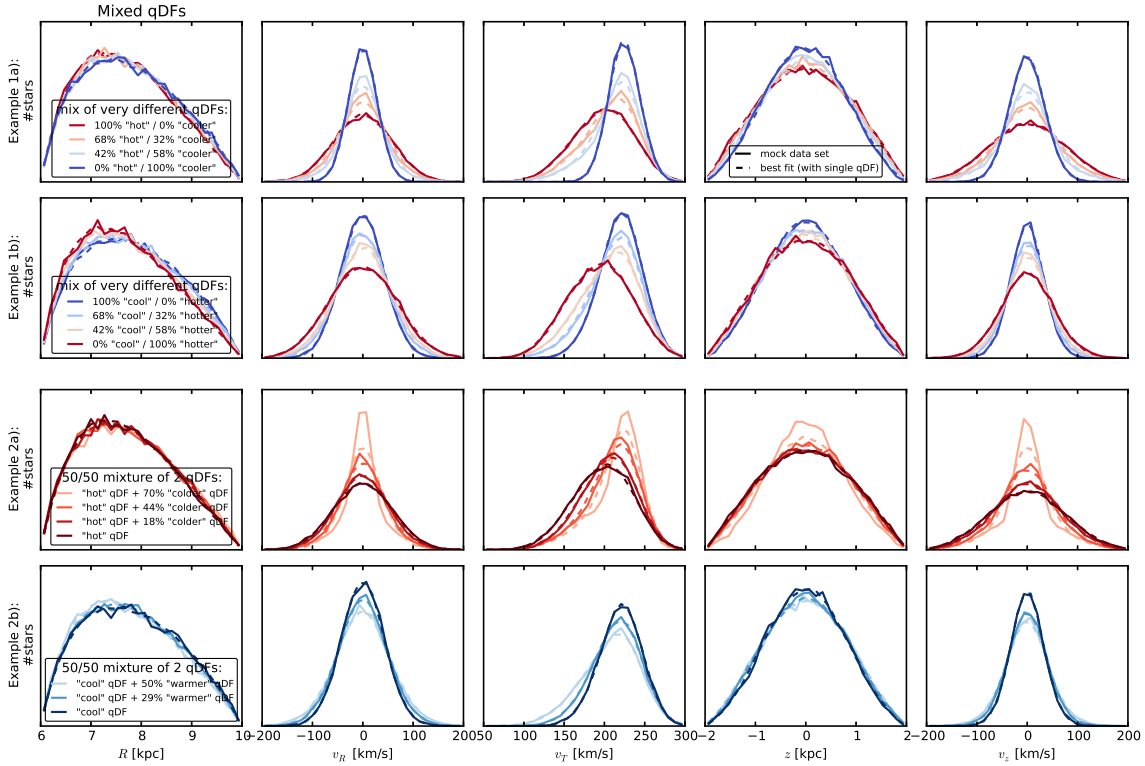

Figure 6. Distribution of mock data, created by mixing stars drawn from two different qDFs (solid lines), and the distribution predicted by the best fit of a single qDF and potential to the data (dashed lines). The model parameters to create the mock data (solid lines) are given in Table 1 as Test 7, and the qDF parameters referenced in the figure's legend are given in Table **??**. The corresponding single qDF best-fit curves (dashed lines) were created by drawing mock data from the best fit parameters found in Figures **??** and **??**. *Example 1:* Distribution of mock data drawn from a superposition of two very different (but fixed) qDFs at varying mixing rates. *Example 2:* Mock data distribution of two MAPs that were mixed at a fixed rate of 50%/50%, but the difference of the qDF parameters of one MAP was varied with respect to the qDF parameters of the other MAP by X% (see Table **??**). The data sets are color coded in the same way as the corresponding analyses in Figures **??** and **??**. This figure demonstrates how mixing two qDFs can be used as a test case for changing the shape of the DF to not follow a pure qDF anymore, e.g. by adding wings or slightly changing the radial density profile. When comparing the mock data and best fit distribution, we see that especially for the most extreme deviations it becomes obvious that a single qDF is a bad assumption for the stars' *true* DF. [TO DO: Potential and/or population names in typewriter font] [TO DO: include X somehow in figure to explain it better. Jo didn't understand what I meant by it in this caption.] [TO DO: These are really many panels. Try to remove some.]

sun this approximation works well for distance errors of $\sim 10\%$ or smaller. The number of MC samples needed for the error convolution using MC integration scales by $N_{\rm error} \propto (\delta v_{\rm max})^2$ with the maximum velocity error at the edge of the sample. If we did not know the true size of the proper motion measurement errors perfectly, we can only reproduce the true model parameters to within $\lesssim 2$ sigma [TO DO: Check???] as long as we do not underestimate it by more than 10% and for proper motion errors $\lesssim 2$ mas yr^{-1}.

2.1. *Influence of wrong assumptions about incompleteness of the data parallel to the Galactic plane*

In §1.1 we found a striking robustness of the *RoadMapping*modelling approach against wrong assumptions about the radial incompleteness of the data set. To further test this result, we investigate a different completeness function that drops with distance from the Galactic plane (see Test 5, Example 2, in Table 1 and Figure 6). We get a similar robust behaviour for small deviations, and only slightly less robustness for larger deviations. That an explanation for this robustness could be, that much of the information about the potential comes from the rotation curve, which is not affected by incompleteness, is demonstrated in Figure 8.

Marginalization over v_T.— The likelihood in Equation **??** is marginalized over the coordinate v_T as follows

$$\begin{aligned}
&\mathscr{L}(p_M \mid D)|_{(v_T \text{ marg.})} \\
=&\prod_i^N P_{(v_T \text{ marg.})}(\boldsymbol{x}_i, v_{R,i}, v_{z,i} \mid p_M) \\
\equiv&\prod_i^N v_0 \cdot \int_0^{1.5 v_{\rm circ}(R_\odot)} \mathrm{d}v_T \; P(\boldsymbol{x}_i, v_{R,i}, v_T, v_{z,i} \mid p_M)
\end{aligned}$$

where $P(\boldsymbol{x}, \boldsymbol{v} \mid p_M)$ is the same as in Equation **??** and the numerical integral over v_T is performed as a 24th order Gauss-Legendre quadrature. The additional factor of v_0 is needed to get the units of $P_{(v_T \text{ marg.})}(\boldsymbol{x}_i, v_{R,i}, v_{z,i} \mid p_M)$ right.

[TO DO: Mention in text or caption how the panels looked that I removed.]

REFERENCES

Batsleer, P., & Dejonghe, H. 1994, A&A, 287, 43
Binney, J. 2010, MNRAS, 401, 2318

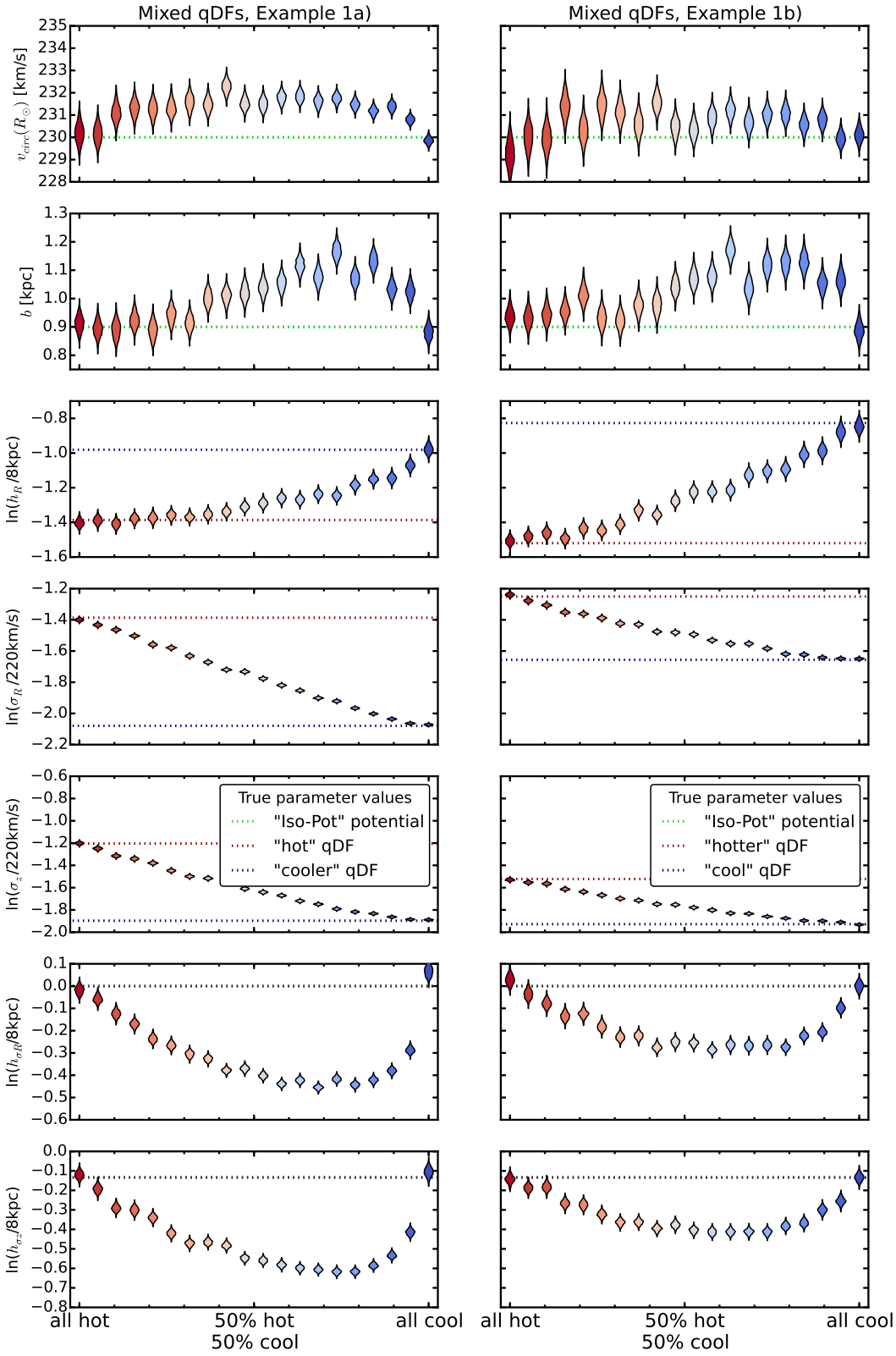

Figure 7. The dependence of the parameter recovery on degree of pollution and temperature of the stellar population. To model the pollution of a hot stellar population by stars coming from a cool population and vice versa, we mix varying amounts of stars from two very different populations, as indicated on the X-axis. The composite mock data set is then fit with one single qDF. The violins represent the marginalized likelihoods found from the MCMC analysis. *Example 1a* (*Example 1b*) in the left (right) panels mixes the `hot` (`cool`) MAP with the `cooler` (`hotter`) MAP in Table **??**. All model parameters used to create the mock data are given in Test 7, *Example 1a) & b)* in Table 1. Some mock data sets are shown in Figure **??**, first two rows, in the same colors as the violins here. We find that a hot population is much less affected by pollution with stars from a cooler population than vice versa. [TO DO: rename $h_{\sigma R}$ to $h_{\sigma,R}$, σ_R to $\sigma_{R,0}$ and analogous for z] [TO DO: Potential and/or population names in typewriter font] [TO DO: Comment from Jo: I feel like just showing one of these examples might be clearer, because they essentially demonstrate the same thing.] [TO DO: Remove σ_R and $h_{\sigma,R}$ panels. Then make two columns with only one expample, potential and DF parameters separately]

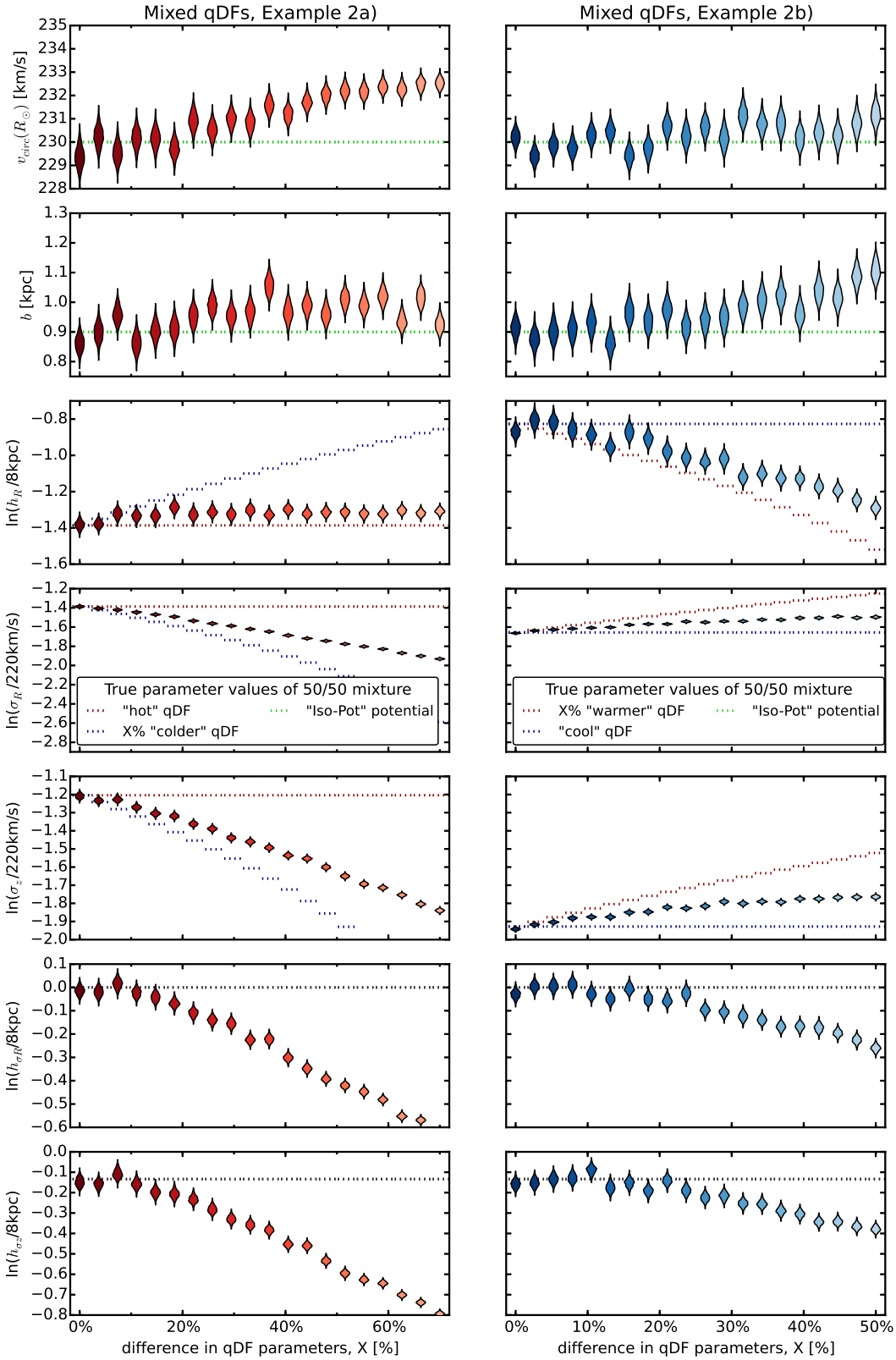

Figure 8. (Caption on next page.)

Figure 8. The dependence of the parameter recovery on the difference in qDF parameters of a 50%/50% mixture of two stellar populations and their temperature. Half of the star in each mock data set in *Example 2a* (*Example 2b*) was drawn from the `hot` (`cool`) qDF in Table **??**, and the other half drawn from a `colder` (`warmer`) population that has X% smaller (larger) $\sigma_{R,0}$ and $\sigma_{z,0}$ and X% larger (smaller) h_R. Each composite mock data set is then fitted by a single qDF and the marginalized MCMC likelihoods for the best fit parameters are shown as violins. The model parameters used for the mock data creation are given as Test 7, *Example 2a) & b)* in Table 1. Some mock data sets are shown in figure **??**, last two rows, where the distributions have the same colors as the corresponding best fit violins here. By mixing MAPs with varying difference in their qDF parameters, we model the effect of bin size in the [Fe/H]-[α/Fe] plane when sorting stars into different MAPs: The smaller the bin size, the smaller the difference in qDF parameters of stars in the same bin. We find that the bin sizes should be chosen such that the difference in qDF parameters between neighbouring MAPs is less than 20%. [TO DO: rename $h_{\sigma R}$ to $h_{\sigma,R}$, σ_R to $\sigma_{R,0}$ and analogous for z] [TO DO: Potential and/or population names in typewriter font]

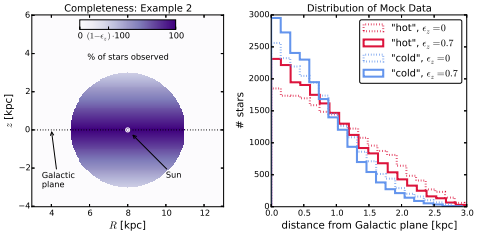

Figure 9. Selection function and mock data distribution for investigating vertical incompleteness of the data. All model parameters are summarized as Test 5, Example 2, in Table 1. The survey volume is a sphere around the sun and the percentage of observed stars is decreasing linearly with distance from the Galactic plane, as demonstrated in the left panel. How fast this detection/incompleteness rate drops is quantized by the factor ϵ_z. Histograms for four data sets, drawn from two MAPs (`hot` in red and `cool` in blue, see Table **??**) and with two different ϵ_z, 0 and 0.7, are shown in the right panel for illustration purposes. [TO DO: Potential and/or population names in typewriter font]

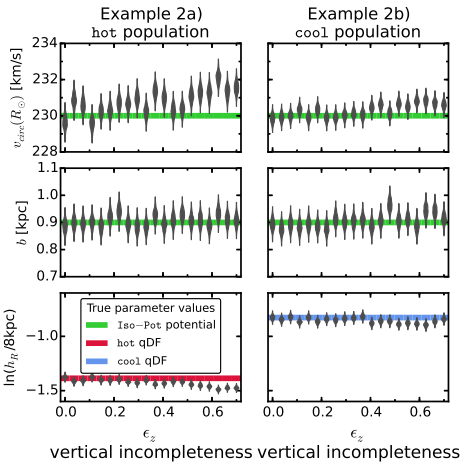

Figure 10. Influence of wrong assumptions about the incompleteness parallel to the Galactic plane of the data on the parameter reocovery with *RoadMapping*. Each mock data set was created having different incompleteness parameters ϵ_z (shown on the x-axis and illustrated in Figure 6) and the model parameters are given as Test 5, Example 2, in Table 1. The analysis however didn't know about the incompleteness and assumed that all data sets had constant completeness within the survey volume ($\epsilon_z = 0$). The marginalized likelihoods from the fits are shown as violins. The green lines mark the true potential parameters (`Iso-Pot`) and the red and blue lines the true qDF parameters (`hot` MAP in red and `cool` MAP in blue), which we tried to recover. The *RoadMapping* method seems to be robust against small to intermediate deviations between the true and the assumed vertical data incompleteness, as well as the radial incompleteness in Figure 7.

Binney, J., & McMillan, P. 2011, MNRAS, 413, 1889
Binney, J. 2011, Pramana, 77, 39
Binney, J. 2012, MNRAS, 426, 1324
Binney, J. 2012, MNRAS, 426, 1328
Binney, J. 2013, NAR [TO DO: emulateapj doesn't know NAR], 57, 29
Binney, J., & Tremaine, S. 2008, Galactic Dynamics: Second Edition, by James Binney and Scott Tremaine. ISBN 978-0-691-13026-2 (HB). Published by Princeton University Press, Princeton, NJ USA, 2008.
Bovy, J., & Tremaine, S. 2012, ApJ, 756, 89
Bovy, J., Rix, H.-W., & Hogg, D. W. 2012b, ApJ, 751, 131
Bovy, J., Rix, H.-W., Hogg, D. W. et al., 2012c, ApJ, 755,115
Bovy, J., Rix, H.-W., Liu, C., et al. 2012, ApJ, 753, 148
Bovy, J., & Rix, H.-W. 2013, ApJ, 779, 115 (BR13)
Bovy, J. 2015, ApJS, 216, 29
Büdenbender, A., van de Ven, G., & Watkins, L. L. 2015, MNRAS, 452, 956
Dehnen, W., & Binney, J. 1998, MNRAS, 294, 429
de Lorenzi, F., Debattista, V. P., Gerhard, O., & Sambhus, N. 2007, MNRAS, 376, 71
Famaey, B., & Dejonghe, H. 2003, MNRAS, 340, 752
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Garbari, S., Liu, C., Read, J. I., & Lake, G. 2012, MNRAS, 425, 1445
Gilmore, G., & Reid, N. 1983, MNRAS, 202, 1025
Henon, M. 1959, Annales d'Astrophysique, 22, 126
Holmberg, J., Nordström, B., & Andersen, J. 2009, A&A, 501, 941
Hunt, J. A. S., & Kawata, D. 2014, MNRAS, 443, 2112
Hunt, J. A. S., & Kawata, D. 2014, MNRAS, 443, 2112
Jurić, M., Ivezić, Ž., Brooks, A., et al. 2008, ApJ, 673, 864
Kawata, D., Hunt, J. A. S., Grand, R. J. J., Pasetto, S., & Cropper, M. 2014, MNRAS, 443, 2757
Klement, R., Fuchs, B., & Rix, H.-W. 2008, ApJ, 685, 261
Kuijken, K., & Gilmore, G. 1989, MNRAS, 239, 605
McMillan, P. J. 2011, MNRAS, 414, 2446
McMillan, P. J. 2012, European Physical Journal Web of Conferences, 19, 10002
McMillan, P. J., & Binney, J. J. 2008, MNRAS, 390, 429
McMillan, P. J., & Binney, J. 2012, MNRAS, 419, 2251
McMillan, P. J., & Binney, J. J. 2013, MNRAS, 433, 1411
Nordström, B., Mayor, M., Andersen, J., et al. 2004, A&A, 418, 989
Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, A&A, 369, 339
Piffl, T., Binney, J., McMillan, P. J., et al. 2014, MNRAS, 445, 3133
Read, J. I. 2014, Journal of Physics G Nuclear Physics, 41, 063101
Rix, H.-W., & Bovy, J. 2013, A&A Rev., 21, 61
Sanders, J. L., & Binney, J. 2015, MNRAS, 449, 3479
Sellwood, J. A. 2010, MNRAS, 409, 145
Steinmetz, M., Zwitter, T., Siebert, A., et al. 2006, AJ, 132, 1645
Strigari, L. E. 2013, Phys. Rep., 531, 1
Syer D., Tremaine S. 1996, MNRAS, 282, 223
Ting, Y.-S., Rix, H.-W., Bovy, J., & van de Ven, G. 2013, MNRAS, 434, 652
Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, AJ, 137, 4377
Zhang, L., Rix, H.-W., van de Ven, G., et al. 2013, ApJ, 772, 108

[TO DO: In which order should I give the references????]
[TO DO: replace the references which I typed myself with the ones from ADS.]
[TO DO: Check if all references are actually used in paper. ???]

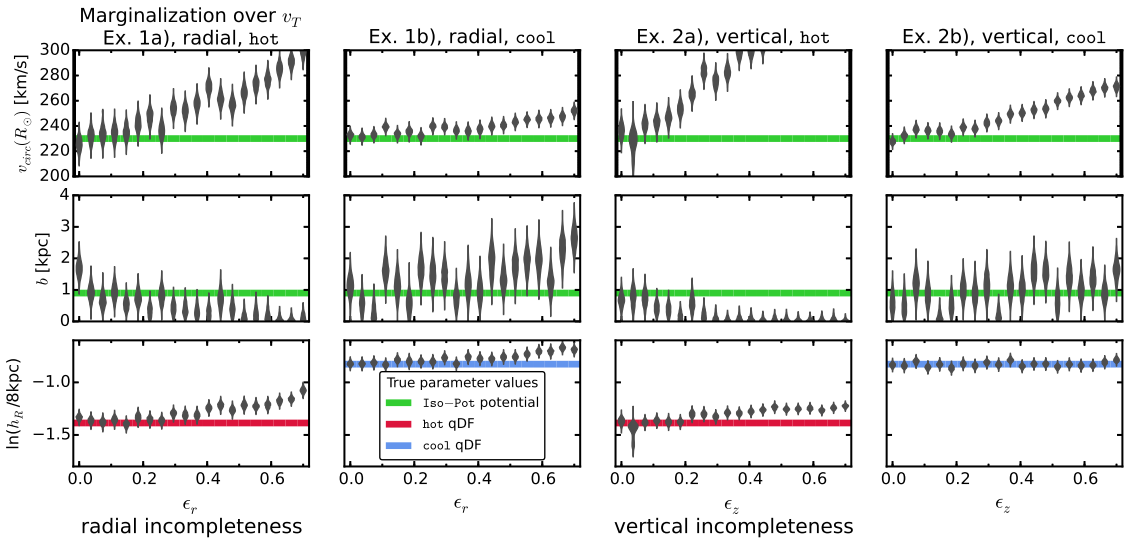

Figure 11. Influence of wrong assumptions about radial and vertical incompleteness on the parameter recovery, when *not* including information about the tangential velocities in the analysis. The mock data sets are the same as in Figure 2 and 7, but this time we did not include the data coordinates v_T in the analysis and therefore marginalized the likelihood over v_T instead (see §2.1). This demonstrates that much of the information about the potential is actually stored in the rotation curve, i.e. $v_T(R)$, which is not affected by removing stars from the data set. But even if we do not include v_T we can still recover the potential within the errors, at least for small ($\epsilon_z \lesssim 10\%$).

Table 1
Summary of test suites in this work: The first column indicates the test suite, the second column the potential, DF and selection function model etc. used for the mock data creation, the third model the corresponding model assumed in the analysis, and the last column lists the figures belonging to the test suite. Parameters that are not left free in the analyis, are always fixed to their true value. Unless otherwise stated we calculate the likelihood by the nested-grid and MCMC approach outlined in §**??** and use $N_x = 16$, $N_v = 24$, $n_\sigma = 5$ as numerical accuracy for the likelihood normalisation in Equations (**??**) and (**??**).

Test		Model for Mock Data	Model in Analysis	Figures
Test 1 : Influence of survey volume on mock data distribution, also in action space	*Potential:*	`KKS-Pot`	-	Mock data: Figure **??**
	DF:	`hot` or `cool` qDF		
	Survey volume:	a) $R \in [4, 12]$ kpc,$z \in [-4, 4]$ kpc, $\phi \in [-20°, 20°]$. b) $R \in [6, 10]$ kpc,$z \in [1, 5]$ kpc, $\phi \in [-20°, 20°]$.		
	# stars per data set:	20,000		
	# data sets:	4 (= 2 × 2 models)		
Test 2 : Numerical accuracy in calculation of the likelihood normalisation	*Potential:*	`KKS-Pot`	-	Convergence of normalisation: Figure **??**
	DF:	`hot` qDF		
	Survey volume:	sphere around Sun, $r_{\max} = 0.2, 1, 2, 3$ or 4 kpc		
	Numerical accuracy:	$N_x \in [5, 20]$, $N_v \in [6, 40]$, $n_\sigma \in [3.5, 7]$		
Test 3.1 : *pdf* is a multivariate Gaussian for large data sets.	*Potential:*	`Iso-Pot`	`Iso-Pot`, all parameters free	Figure **??**
	DF:	`hot` qDF	qDF, all parameters free	
	Survey Volume:	sphere around Sun, $r_{\max} = 2$ kpc	(fixed & known)	
	# stars per data set:	20,000		
	# data sets:	5 (only one is shown)		
	Numerical accuracy:		$N_v = 20$ and $n_\sigma = 4$	
Test 3.2 : Width of the likelihood scales with number of stars by $\propto 1/\sqrt{N}$.	*Potential:*	`Iso-Pot`	`Iso-Pot`, free parameter: b	Figure **??**
	DF:	`hot` qDF	`hot` qDF, free parameters: $\ln h_R, \ln \sigma_{R,0}, \ln h_{\sigma,R}$	
	Survey volume:	sphere around Sun, $r_{\max} = 3$ kpc	(fixed & known)	
	# stars per data set:	between 100 and 40,000		
	# data sets:	132		
	Analysis method:		likelihood on grid	
	Numerical accuracy:		$N_v = 20$ and $n_\sigma = 4$ (for speed)	
Test 3.3 : Parameter estimates are unbiased.	*Potential:*	`Iso-Pot`	`Iso-Pot`, free parameter: b	Figure **??**
	DF:	`hot` or `cool` qDF	`hot/cool` qDF, free parameters: $\ln h_R, \ln \sigma_{R,0}, \ln h_{\sigma,R}$	
	Survey volume:	5 spheres around Sun, $r_{\max} = 0.2, 1, 2, 3$ or 4 kpc	(fixed & known)	
	# stars per data set:	20,000		
	# data sets:	640 (= 2 × 2 × 5 models ×32 realisations)		
	Analysis method:		likelihood on grid	
	Numerical accuracy:		$N_v = 20$ and $N_\sigma = 4$ (for speed)	
Test 4 : Influence of position & shape of survey volume on parameter recovery	*Potential:*	i) `Iso-Pot` or ii) `MW13-Pot`	i) `Iso-Pot`, all parameters free ii) `MW13-Pot`, R_d and f_h free	Figure **??**
	DF:	`hot` qDF	i) qDF, all parameters free ii) qDF, only h_R, $\sigma_{z,0}$ and $h_{\sigma,R}$ free	
	Survey volume:	4 different wedges, see Figure **??**, upper right panel	(fixed & known)	
	# of stars per data set:	20,000		
	Analysis method:		i) MCMC, ii) likelihood on grid	
Test 5 : Influence of wrong assumptions about the data set (in-)completeness on parameter recovery	*Potential:*	`Iso-Pot`	`Iso-Pot`, all parameters free	Illustration & mock data: Figures 1 & 6 Analysis results: Figures 2 & 7 Analysis results: when not using v_T data: Figure 8
	DF:	a) `hot` or b) `cool` qDF	qDF, all parameters free	
	Survey volume:	sphere around Sun, $r_{\max} = 3$ kpc	(fixed & known)	
	Completeness:	*Example 1:* radial incompleteness, completeness$(r) = 1 - \epsilon_r \frac{r}{r_{\max}}$, twenty $\epsilon_r \in [0, 0.7]$ $r \equiv$ distance from Sun,	data set complete, completeness$(r) = 1$, $\epsilon_r = 0$	
		Example 2: planar incompleteness,	data set complete,	

Table 1 — *Continued*

Test		Model for Mock Data	Model in Analysis	Figures
		completeness$(z) = 1 - \epsilon_z \frac{\lvert z\rvert}{r_{\max}}$, $\epsilon_r \in [0, 0.7]$, $z \equiv$ distance from Gal. plane.	completeness$(r) = 1$, twenty $\epsilon_z = 0$	
	# stars per data set:	20,000		
	# data sets:	40 $(= 2 \times 2 \times 20)$		
Test 6.1 : Numerical convergence of convolution with measurement uncertainties	*Potential:*	`Iso-Pot`	"Iso-Pot, all parameters free"	Figure **??**
	DF:	`hot` qDF	qDF, all parameters free	
	Survey Volume:	sphere around Sun, $r_{\max} = 3$ kpc	(fixed & known)	
	Uncertainties:	$\delta\text{RA} = \delta\text{DEC} = \delta(m - M) = 0$ $\delta v_{\text{los}} = 2$ km/s $\delta\mu_{\text{RA}} = \delta\mu_{\text{DEC}} =$ 2,3,4 or 5 mas/yr	Convolution with perfectly known measurement uncertainties	
	Numerical Accuracy:		$N_{\text{error}} \in [25, 1200]$	
	# stars per data set:	10,000		
Test 6.2 : Testing the convolution with measurement & without uncertainties with distance errors	*Potential:*	`Iso-Pot`	`Iso-Pot`, all parameters free	Figure 3
	DF:	`hot` qDF	qDF, all parameters free	
	Survey Volume:	sphere around Sun, $r_{\max} = 3$ kpc	(fixed & known)	
	Uncertainties:	$\delta\text{RA} = \delta\text{DEC} = 0$, $\delta v_{\text{los}} = 2$ km s^{-1}, $\delta\mu_{\text{RA}} = \delta\mu_{\text{DEC}} =$ 1, 2,3,4 or 5 mas/yr, a) $\delta(m - M) = 0$ or b) $\delta(m - M) \neq 0$. (see Figure 3)	Convolution with measurement uncertainties, ignoring distance errors in position (see Section **??**)	
	Numerical Accuracy:		$N_{\text{error}} = 800$ for $\delta\mu \leq 3$mas yr^{-1}, $N_{\text{error}} = 1200$ for $\delta\mu > 3$mas yr^{-1}	
	# stars per data set:	10,000		
Test 6.3 : Underestimation of proper motion errors	*Potential:*	`Iso-Pot`	`Iso-Pot`, all parameters free	Figure 5
	DF:	`hot` or `cool` qDF	qDF, all parameters free	
	Survey volume:	sphere around Sun, $r_{\max} = 3$ kpc [TO DO: CHECK]	(fixed & known)	
	Uncertainties:	only proper motion errors 1, 2 or 3 mas/yr	Convolution with proper motion errors 10% or 50% underestimated	
	# stars per data set:	10,000		
	# data sets:	24 $(= 2 \times 2 \times 3 \times 3$ realisations)		
Test 7 : Deviations in the assumed DF from the star's true DF	*Potential:*	`Iso-Pot`	`Iso-Pot`, all parameters free	mock data: Figure **??** Analysis results: Figures **??** & **??**
	DF:	mix of two qDFs *Example 1:* with fixed qDF parameters, but 20 different mixing rates: a) `hot` & `cooler` qDF or b) `cool` & `hotter` qDF *Example 2:* 20 fixed 50/50 mixtures, with varying qDF parameters (by X%): a) `hot` & `colder` qDF or b) `cool` & `warmer` qDF	single qDF, all parameters free	
	Survey volume:	sphere around Sun, $r_{\max} = 2$ kpc	(fixed & known)	
	# stars per data set:	20,000		
	# data sets:	40 $(= 2 \times 2 \times 20)$		
Test 8 : Deviations of the assumed potential model from the star's true potential	*Potential:*	`MW14-Pot`	`KKS-Pot`, all parameters free, only $v_{\text{circ}}(R_\odot) = 230$km s^{-1} fixed	potential contours: Figure **??**
	DF:	`hot` or `cool` qDF	qDF, all parameters free	qDF recovery: Figure **??**
	Survey volume:	sphere around Sun, $r_{\max} = 4$ kpc	(fixed & known)	
	# stars per data set:	20,000		

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