

December 18, 2015

Ms. Wilma H. Trick
Max-Planck-Institut fuer Astronomie
Koenigstuhl 17
Heidelberg, Baden-Wuerttemberg 69117
Germany

Title: Action-based Dynamical Modelling for the Milky Way Disk

Dear Ms. Trick,

I have received the referee's report on your submission to The Astrophysical Journal, and append it below.

I hope that you will agree with my assessment that the report is constructive in tone, but you will note that the referee asks that a number of issues be addressed before publication in the ApJ.

Following the referee's report, I have also appended some comments regarding the statistical aspects of your manuscript. ApJ submissions with a statistical component are routinely previewed by Dr. Eric Feigelson, the member of the ApJ editorial board with a broad expertise in statistical matters in an astrophysical context. Perhaps these comments will be of use to you as you prepare your revision. If you have any questions about the comments, however, please direct them to me, not to Dr. Feigelson. Please consider the report carefully. When you resubmit, please include a **detailed cover letter indicating point-by-point your responses to the referee's report**, and also indicating any **other changes you have made** to the text. Reviewers find it helpful if the changes in the text of the manuscript are easily distinguishable from the rest of the text. Therefore we ask you to **print changes in bold face**; this highlighting can be removed easily after the review.

And two minor matters: Please adhere to the ApJ instructions to authors (see http://aas.org/journals/authors/common_instruct#_Toc3.2) regarding the abstract – **"The abstract should be a single paragraph of not more than 250 words ..."** Please also **use the appropriate journal abbreviations in the reference list**: these abbreviations for refereed journals are given on the ADS site http://adsabs.harvard.edu/abs_doc/refereed.html, and on the site http://adsabs.harvard.edu/abs_doc/non_refereed.html for non-refereed publications.

Click this link (it will work one time) to upload your revised manuscript: <https://apj.msubmit.net/cgi-bin/main.plex?el=A7Ew3QPh3A7DACX7F5A9ftdsyk8KtDBL1q74yfVAJnwOA>

Alternatively, you can also log into your account at the EJ Press web site, <http://apj.msubmit.net>. Please use your user's login name: wilmatrick. You can then ask for a new password via the Unknown/Forgotten Password link if you have forgotten your password.

The policy of The Astrophysical Journal is to **view manuscripts as withdrawn if no revised version is received within six months** after the most recent referee's report goes to the authors.

If you have any questions, please contact me.

With best wishes,

Butler Burton

Prof. W. Butler Burton
Associate Editor-in-Chief, The Astrophysical Journal
Professor Emeritus, Leiden University
National Radio Astronomy Observatory

+++++

The text of the review is appended below.

=====

1 Reviewer's Comments:

This paper details and tests a framework for constraining the Milky Way potential using action-based dynamical distribution functions. The work very much builds on previous work from two of the authors as well as the dynamics group in Oxford. However, for the first time the machinery and its assumptions are quite rigorously tested and the paper thoroughly discusses the limitations and downfalls of the methods.

I believe that this work should be published but that there are parts of the paper where the presentation could be clearer and where the discussion could be expanded or reduced. Additionally, I think that this paper will be a good reference paper for those who will fit dynamical models to the Gaia data, but there are a number of places where the paper could be made more 'usable' without much effort.

My main complaint with the paper, however, is that it doesn't tackle the 'real' problem for three reasons:

- 1. The selection volumes considered seem unrealistic. It is unclear how they relate to what one might use with the Gaia data.
[Comment from HW: E.g. how to deal with complex sampling volumes has been demonstrated in Bovy et al 2013 and 2015; the point here is to make a generic exploration of search volume shapes.]
- 2. It is unclear how the error distributions relate to the anticipated errors from the Gaia data.
[TO DO: Good point. I will try to compare the Gaia error regimes with my results. Rene helped me already a little bit and found out that a magnitude cut at $G=15$ should keep the distance errors below 10% at a maximum distance of 3 kpc.]
- 3. The majority of the tests seem to be done with the isochrone potential. This is obviously done for speed reasons. A few cases are tackled with the KKS-Pot and one case with the MW13-Pot. It is then not clear how the precision of the potential parameter recovery for the isochrone relates to

more realistic potentials (e.g. will the circular velocity always be so well constrained irrespective of the potential form?). [Comment from Wilma: This is a valid point. Relating the parameters isochrone \leftrightarrow realistic potential is not possible. But at least we could maybe have one example that shows that v_{circ} is well recovered also for a realistic potential? Idea 1: I tried to make Figure 19 with v_{circ} not being fixed. While v_{circ} was indeed tightly constrained for the cool population, unfortunately I did not manage to make the analysis of the hot population converge. I probably need higher numerical accuracy. I could try again. I hope that 7 days of computation time on 50 cores suffice. Idea 2: I also could make Figure 6 for the KKS-Potential instead of the isochrone potential. That should be a bit quicker, I think.] The issue of computational speed is mentioned in the discussion but I think this issue is slightly glossed over. [Comment from Wilma: No idea what he wants...] Using more realistic potentials in all cases would obviously be much better and would also test the ‘Staeckel fudge’ apparatus more fully. The improvements to this apparatus are highlighted in the paper but I don’t think that it is shown that it has been fully tested. [Comment from Wilma: The Staeckel fudge was tested by other people excessively. I think we do not have to do that. Concerning the use of realistic potentials in all my tests: Just not possible. Sorry. But maybe we could do the following: I could have a potential with a fixed Hernquist halo+bulge and a Miyamoto-Nagai disk with scale length $a=0.9\text{kpc}$ (or would that make the potential too unrealistic?). Then isochrone and this more realistic potential have the same free parameters. I could for a few selected test cases in 3.3 (incomplete data) and 3.5 (deviations from qDF) do the analysis with this more realistic potential and overplot these violins with my existing violin plots. Let’s hope that we get the same trend...] I think that the heavy use of the isochrone potential should be flagged in the abstract.

It would be good if some of these concerns were brought to the fore and highlighted in the paper. I have detailed places in the paper where things could be changed. There is also a short list of typos at the end.

2 Abstract

- 1. 4th sentence – explain what ‘slightly wrong’ means more quantitatively.
- 2. 5th sentence – Are the constraints of high precision on the potential or DF? – clarify
- 3. The tests referred to in the abstract have been performed independently but the lists suggest that you have shown that when all the listed conditions are satisfied the constraints are of high precision. This should be clarified.

3 Introduction

- 1. Magorrian (2014) has provided a framework for constraining the potential without assuming a particular parametrized form for the DF. Whilst Magorrian's method is computationally intensive, it should be referenced in the introduction as it relates to the later discussion of choosing a particular DF parametrization.
- 2. At the end of the introduction I think that you should refer people more strongly to the results section as much of section 2 is presenting a framework that appears elsewhere.

4 Dynamical Modelling

4.1 Section 2.2

- 1. A slightly fuller introduction of the actions is merited. Mention why the actions are introduced. What advantages do they present? Also state that the action-angles are canonical – this is important later for transformation of the pdf. \Rightarrow *I completely agree. I added a corresponding paragraph.*
- 2. The third sentence isn't quite right – the most general are the triaxial Staeckel potentials of which the axisymmetric Staeckel potentials are special cases and all spherical potentials are special cases of these. The isochrone potential is the most general potential in which the actions are not computed as a quadrature.
- 3. The potential discussion could be put into a separate section. Also mention that the circular speed at the Sun is the same for all three potentials. Re-iterate that the reason for using the isochrone and Staeckel potential is the ease with which the actions can be computed.

4.2 Section 2.3

- 1. Guiding-center should not be in brackets as this is important. \Rightarrow *Done.*
- 2. The final sentence of the left-hand column of Fig 3 should read $L \ll L_0$. [TO DO: Ask Jo, if this is correct. I think so, but make sure.]
- 3. Top of page 4 – explain what X is in the text.
- 4. Do you interpolate in log density?
- 5. The footnote says 'should be chosen as' – add a forward reference to Fig. 4

4.3 Section 2.4

- 1. An entire section dedicated to this topic seems unnecessary. See below.

4.4 Section 2.5

- 1. As this section only explains technical details rather than testing the apparatus I think that this section can be put in an appendix along with Fig. 2 and 3. Fig 2 and 3 are illustrative but I think that they are similar to BR13 Fig2 and 3 so do not need to appear in the main body. I don't think it is a 'test' so should be removed from Table 3. \Rightarrow I agree. I moved the mock data section and figures to the appendix and removed the parameters belonging to Figure 2 + 3 from the 'tests' Table 3.
- 2. The discussion of selection on very erroneous x coordinates is interesting but surely this isn't the way the data will actually be handled? [Comment by Wilma: No idea what he means.]

4.5 Section 2.6

- 1. The selection function can be briefly mentioned at the beginning of this section and stated that you assume here for simplicity it is a function of \vec{x} .
- 2. The Jacobian from J,theta to x,v should be mentioned here. [Comment by Wilma: I don't see how the Jacobian is relevant here.]
- 3. pdf should be defined in a separate equation.
- 4. Figure 4. – it wasn't clear to me that the 'truth' normalization used a high enough set of parameters. $N_x = 20$, $N_v = 56$ and $n_\sigma = 7$ only seem slightly larger than the values actually compared to.
Comment by Wilma: I can re-run it. What would be large enough values?
- On the other hand I'm not sure, if we maybe should tackle the question of numerical accuracy in a different way. Let's talk about this...
- 5. Make the normalization discussion a separate section.
- 6. The discussion of the likelihood normalization should reference and compare with McMillan and Binney (2013) as the discussion is very similar.
- 7. Is there any general advice on how to choose N_x , N_v and N_σ ? The authors have shown it is OK for the mock datasets but do I have to redo the authors' exercise when I have a real dataset? [Comment by Wilma: My honest answer is "If you have 20,000 stars, you can start with the same parameters as me. But yes, ultimately you would have to do the full exercise to find out, if for your special case this accuracy is fine.". Is it okay to say it like this? I have no idea what other general advice I should give...]
- 8. Put error discussion in separate section.
- 9. Reduce size of caption for Fig 5. More of the details could go in the text.

- 10. Equation (15) is a novelty. It is troubling that the tests that use this approximation all seem to use the isochrone but the approximation is still necessary. Is that because it is computationally awkward to calculate this integral or just very slow? [Comment by Wilma: MC integration/convolution over the errors is incredibly slow (but the only idea we had so far). To do some statistics as for Figure 5 I HAD to use the isochrone potential. If we would not use the cheat to ignore the errors in the normalisation, the computation time to do one single analysis would explode. So I think the computation time is the problem. What does he mean by "computationally awkward"?]
- 11. The penultimate sentence of this section contradicts the previous sentence without validation. Why is this?

4.6 Section 2.7

- 1. I liked this section – it was well thought out and informative. \Rightarrow Thanks. :-)
- 2. Here a fixed sampling is used for the error samples. I think again you should reference McMillan & Binney (2013) as they discussed the numerical stability of this method.

5 Section 3

- 1. It is stated that the breakdown of axisymmetry and steady state assumptions is not explored. I wonder as well about the impact of resonances, particularly when the data are very high quality. This cannot be explored in the current setup as the data are generated from an action-based DF but perhaps should be mentioned as a potential limitation of the approach.

5.1 Section 3.1

- 1. This seems a good sanity check but should it be published? Fig 6. seems sufficient to me to demonstrate that your code works. I don't think the paper would miss this section.
[Comment by Wilma: I see the point, but I would like to keep it, please... It was a lot of work...]

5.2 Section 3.2

- 1. I understand that the selections used in Fig 9 are illustrative but the pink selection just doesn't seem realistic. I think Fig 8. is a sufficient demonstration of the difference between different selections. Fig. 9 doesn't add anything and is barely discussed in the text. Also, without observational uncertainties (which will be greater for the more distant boxes) the discussion seems superficial. I would consider removing this.
[Comment by Wilma: Maybe I should make the point of this figure clearer. It does consider unrealistic selection functions, but I just wanted to figure

out if there are regions within the Galaxy where stars are in general on more informative orbits. Or if we, for example, have a natural disadvantage or advantage due to the sun's position within the MW. I vote for keeping the figure.]

5.3 Section 3.3

- 1. Isn't the reason for the cold population being more robust that it doesn't have as many stars at large distance as the hot population so it is less affected by the cuts? I suppose this not necessarily true for lines-of-sight in the plane.
[Comment by Wilma: Good point. Any ideas?]

5.4 Section 3.4

- 1. It would be nice to state how the considered errors are related to the anticipated Gaia errors or other surveys. [Comment by HW: The proper motion errors which we choose cover very coarsely the proper motion error regions of ground based surveys. TO DO: Look it up in more detail. Rene helped me with the distance errors.]

5.5 Section 3.5

- 1. I think this and section 3.6 are the most valuable in the paper as they really explore potential systematics. In my opinion, these are the key results.
- 2. Fig 15. – it would be interesting to see the difference between the fits and the truth. Do the fits break down in particular places?
[Comment by Wilma: Not sure what he wants. I think the differences become already visible in the plot, right? What does he mean by "particular places"?]

5.6 Section 3.6

- 1. Fig 19 is difficult to interpret. Is it possible to display the difference?
[Comment by Wilma: I don't think that it is difficult to interpret. Showing the difference would let me include much less information into the plot + I personally think reading residual plots is even more difficult. What I COULD do, is adding two panels that compare cuts through the density along $z @ R=8$ and along $R @ z=0$ for all models.]
- 2. The fact the density is not well recovered seems interesting as it points to possible biases in the surface density of the disc/dark matter measurements if one uses the wrong potential. It would be good to have the discrepancy quantized in the text.
[Comment by Wilma: How to quantize the discrepancy?]
- 3. I think Fig. 20 could be removed. As mentioned it doesn't make sense to compare the DF parameters between different potentials so I am not sure what Fig 20 is telling us.

[Comment by Wilma: I admit it is not a very important figure, but it makes the point in the text clearer that the qDF parameters in themselves are not that informative depending on which potential is used. I think that is easily forgotten and people should be reminded of this.]

5.7 Section 3.7

- 1. Should this section be moved to the discussion section?

6 Summary & Discussion

- 1. Perhaps add statements comparing the errors explored to those anticipated from Gaia.
- 2. The two approaches mentioned at the end of 'Gravitational potential beyond the...' are stated as formally similar but I think it is clear that one is better than the other. The true Staackel approach limits you to potentials with the same foci. This is an obvious limitation and has been discussed before.
- 3. The definition of X in $f(J, [X/H])$ doesn't seem to make sense. [Comment by Wilma: Why not?]
- 4. The first section in future work is very interesting. Use of different DFs and potentials as explored in this paper is interesting but a true test of the apparatus on a more realistic galaxy would make the 'RoadMapping' tool much more attractive.
- 5. I think that the final two questions of the future work section are weak. Clearly the rotation curve is only describing the in-plane force not the force everywhere. Parametrizations will naturally convince you that the rotation curve is well measured but I think there is a lot more flexibility. Also, the advantage of using the approximate actions is that more realistic potentials can be considered. [Comment by Wilma: I think he is right. Do we have any other questions that have popped up that we wanted to pose in the discussion section?]

6.1 Table 3

- 1. Can you add a summary column that summarizes the result? This would make the paper much more 'usable'.

7 Typos \Rightarrow *Done*

Abstract. 1. 3rd sentence – 'rules of thumb' for how data, model and machinery most affect ... and DF. \Rightarrow *Done*.

Introduction. 1. Start of 3rd para: 'to constrain' \rightarrow 'constraining' \Rightarrow *Done*.
2. Start of penultimate para: 'to restrict' \rightarrow 'restricting' \Rightarrow *Done*.

Table 1. 1. 'troughout' in caption. \Rightarrow *Done*.

Section 2.3. 1. Second sentence: 'about'→'on' \Rightarrow *Done*. 2. 'the circular orbit' to 'near-circular orbit'? \Rightarrow *Done*. 3. Top of page 4 two 'in's \Rightarrow *Done*.

Section 2.6 1. Top of right column page 7 – replace 'besides' with 'not only... but also' \Rightarrow *Done*.

Section 2.7 1. Need '(MCMC)' after MCMC \Rightarrow *Done*.

Fig 14. 1. 'pest'→'best' \Rightarrow *Done*.

Fig 15. 1. 'refereed'→'referred' \Rightarrow *Done*.

Section 3.5. 1. Second paragraph right column page 13 – 'sun'→'Sun' \Rightarrow *Done*.

Table 3. 'analys'→'analysis' \Rightarrow *Done*.

8 Comment from Dr. Eric Feigelson

=====

Below are comments on statistical aspects of the manuscript: ApJ submissions with a statistical component are previewed by Dr. Eric Feigelson, the member of the ApJ editorial board with a broad expertise in statistical matters in an astrophysical context.

+++++

An elaborate Bayesian inferential procedure is described in sec 2.6-2.7 for parameter estimation with results shown in Fig 6. But with uninformative uniform priors and a simply unimodal likelihood with a nearly multivariate normal distribution, this effort is unnecessary. The same result would be obtained with maximum likelihood estimation via the EM Algorithm (probably ~ 100 iterations with convergence guaranteed by theorem) and parameter uncertainties estimated from the Fisher Information Matrix. The confluence of Bayesian and MLE procedures in this case should be presented. [Comment by Wilma: Jo said, we could ignore that... Can we really? What should I answer the referee?]