Dear editor,

We thank our referee, Dr. J. Ted Mackereth, for his thorough comments on our submission. We greatly appreciate his insight, and we believe our manuscript has improved as a result of his efforts.

Our responses are below, interleaved with comments from Dr. Mackereth in bold. Substantial changes to the revised manuscript are marked in red.

The paper is well put together, if quite long and detailed! Much of the latter portion of the paper is given over to detailed descriptions of comparisons between the model predictions and the APOGEE data. The authors are refreshingly upfront about places where their model fails to correctly predict observations. However, I felt in general the authors might have more strongly emphasised what new predictions the models provide for the Milky Way's formation and evolution. As an example, more than 50% of the abstract is a list of agreement rather than any novel predictions! This is subjective, of course, so I am not explicitly asking for changes here.

We are pleased that the referee appreciates the clarity and transparency of the paper. The comment that we focus on comparison to existing data rather than predictions for future data is a fair one. There are several reasons for this focus: (a) many of the observational constraints we compare to are relatively recent results from APOGEE, and no previous models have been compared to them systematically as a set; (b) our modeling framework still involves multiple parameters/choices, most notably the smooth vs. bursty star formation history, and many predictions depend on those choices; (c) as stated in the paper, within the range of parameters we have investigated, some of the existing observations remain difficult to fit, especially the bimodality of [alpha/Fe].

We have added sentences after the numbered list of conclusions in Section 4 to say that the predictive power of the models can be extended by including elements with large AGB contributions, since these have different timescales and metallicity dependent yields. These sentences are a prospect rather than an actual prediction, but they are a direction we are pursuing, limited mainly by the fact that we don't consider yield predictions reliable enough to be taken at face value without empirical calibration.

Our other main changes to the paper, discussed further below, are the addition of a table of model parameters, greater clarity on the choice of analogues for migration calculations, and further explanation of the SNIa rate fluctuations as a potential explanation for young alpha-rich stars.

Does the imposed SFH match in any quantitative way that from the simulation? I was concerned that this model imposes some star formation beyond R > 13kpc at times as early as 11-12 Gyr ago. While small, I would presume that these annuli try to match at least a few star particles from the simulation at these times. Given the birth radii distribution of the simulations at 8-10Gyr old, then I would expect that the model tags an 'analogue' at a very different radius to the implied birth radius (given the current description of the tagging process). Is this the case? If not, how is this handled? If so, how important is this for the radial migration prescription? Some text describes this, but I felt that the paper would benefit from some more detail here. I would imagine that the dynamics of stars forming at R < 5kpc are very different to those forming at large radii at early times.

The imposed SFH is not taken from the simulation. Instead, it is intended to resemble a simple assumption that one might take for a galaxy like the Milky Way, where "simple" for the sake of chemical evolution models often means neglecting effects like the radial growth of the Galaxy. Dr. Mackereth's understanding of the tagging process is accurate; the model indeed matches these stellar populations born at large radii and early times with star particles from the h277 simulation, but the number of star particles in our sample is sufficiently large that this is not a cause for concern. For each of the stellar population and star particle pairs in our fiducial model, we have recorded the difference in birth radius between the two and analyzed the resulting distributions. We find that even at large radii (R $\gtrsim$ 10 kpc) and early times (age > 10 Gyr), the majority of stellar populations are assigned an analogue that formed within $\sim$2 kpc of its birth radius. Stellar populations born at, e.g., R = 13 kpc are in general not being tagged with analogues that formed as far in as R = 5 kpc, instead finding star particles that formed at $\sim$11 kpc if the search widens that far without finding another analogue first. We have added a few sentences to this effect to paragraph 4 of § 2.2 so that readers can be assured that our models assign outer rather than inner disc star particles to these outer disc stellar populations. The fact that there are occasionally analogue separations of 2 or more kpc is the reason why we assign final particle radii as $R_{\text{final}} = R_{\text{initial}} + \Delta R_{\text{analogue}}$ as opposed to $R_{\text{final}} = R_{\text{final,analogue}}$.

This issue is of little importance to the conclusions presented in this paper. In our exploratory work, we originally used an algorithm in which stellar populations that did not find an analogue within $R \pm 500$ pc and $T \pm 500$ Myr simply stayed at their radius of birth. These models make similar predictions to the ones presented in the paper, which we take as an indication that the fine details of the dynamical history do not impact our results. In general, the conclusions we discuss are impacted more so by the fact *that* stars are migrating and not so much by *how* stars are migrating. We have also added two sentences to this effect to paragraph 4 of § 2.2.

Similarly to above, my reading of the current procedure is that it may imply some biases as to where the analogues are selected. Since the search is conducted in cylindrical coordinates, then surely the search volume will be vastly different dependent on the radii at which it is conducted (since delta R is constant)? i.e at R = 1kpc, a ±250pc annulus is ∼10 times smaller than one at 15kpc. Perhaps this has no effect, but to the uninitiated reader this seemed like it could be a concern.

It is true that the search annuli increase in size with increasing radius, but we do not believe that this has any effect on the model. This point is actually closely connected to the previous question; the larger search annuli at large R help ensure that the stellar populations born there are able to find an analogue born at a similar radius without having to extend the search too far inward. Even if we make a large change to our model such that stars stay at their birth radius if they do not find an analogue within $R \pm 500$ pc and $T \pm 500$ Myr, our conclusions are still unaffected. We hope that the aforementioned addition of text to paragraph 4 of § 2.2 helps mitigate this concern. There may be interesting regions of parameter space where the model predictions are sensitive to the fine details of the dynamical history and decisions like this would matter, but in practice we find that the models in the present paper are unaffected.

Finally, the authors use the z dependence in their comparisons but (as far as I could see) did not mention how this was handled. I assumed that the final z of the simulation particles is adopted, but it would be good to state this somewhere explicitly.

We thank Dr. Mackereth for letting us know that this was not made clear enough. His assumption was correct; it is indeed taken from the h277 analogue assigned to a stellar population. To address this, we have modified sentence 5 in paragraph 3 of § 2.2 to say that the stellar population adopts the present-day midplane distance z and the change in radius ΔR_{gal} from its analogue. This sentence previously stated this only for the change in radius. We have also added a table of our model parameters near the end of § 2, and it is stated there that both ΔR_{gal} and z are taken from the h277 analogue.

I think it would be good to explicitly mention that h277 has ∼the same final scale length as the Milky Way, I had to go digging for that information when I was thinking about the above points.

We have added this statement near the end of paragraph 3 of § 2.2.

The 'inside-out' SFH implies that star formation peaked at 10-11 Gyr ago in the very outer disc - is this really a good assumption?

For reference, we are duplicating equation 10 in section 2.5 here, which describes the time-dependence of the SFH in the fiducial model:

$$f_{\text{IO}}(t|R_{\text{gal}}) = (1 - e^{-t/\tau_{\text{rise}}})e^{-t/\tau_{\text{sfh}}}. \tag{1}$$

As mentioned in the text, this equation does have a peak near τ_{rise}, but in detail the time of the peak increases with increasing τ_{sfh} (i.e. with increasing R_{gal}). When τ_{sfh} is as high as ∼30 Gyr, the peak in star formation instead occurs ∼7.5 - 8 Gyr ago. As discussed in the text, our choice of $\tau_{\text{sfh}}(R)$ is motivated by observed radial age gradients of disc galaxies comparable to the Milky Way. Whether or not this is accurate for the Milky Way, it at least to some extent reflects the radial growth of the Galaxy. We have added a couple of sentences to the discussion of equation 10 in § 2.5 to clarify this.

An assumption is made that suggests radial migration does not alter the surface density profile of the Galaxy. Is this assumption backed up by the simulation that is used?

Our star particle sample from h277 indeed supports this assumption. We have added such a statement to the first paragraph of § 2.5 and the third paragraph of § 2.7.

A tabulated version of the free parameters of the model would be a useful reference, as I found myself flicking back and forth quite a lot while reading the paper.

We have constructed such a table, and placed it near the and of our methods section. It is referenced in the text at the beginning of § 2 (in the final sentence before § 2.1) and again at the beginning of § 2.8.

How does the inclusion of bulge-like particles from the simulation change things in the inner disk?

Although they are included in VICE's public code base, bulge-like particles are not included in the sample used in the models for this paper. We state this in the final two sentences of paragraph 4 in § 2.1. As a result, every star particle assigned as an analogue is one with disc-like kinematics. We have added a statement to paragraph 3 of § 2.2 to clarify this at the same time the analogue search process is described.

The paper discusses intermediate-age alpha-enhanced stars at length but does not provide an idea of their [Fe/H], which would of course provide a more detailed constraint as to their origin in the model.

This is an excellent suggestion for further discussion on this prediction of our model. We have added the relevant statements to the end of paragraph 5 of § 3.4.

Fig 10: I wondered how much the noted difference between the model and the observations at 3-5kpc might be due to the presence of low metallicity bar stars at this annulus in the data (for example those noted in Bovy et al. 2019). A flattening of the surface density profile implied in the central region may make differences here too, I think?

This is an excellent point. We have added a reference to the metallicity map published in Bovy et al. (2019) to paragraph 3 of § 3.2. However, these models are generally independent of the mass normalization - it matters in determining how far up the $\Sigma_\text{gas} - \dot{\Sigma}_\star$ relation our models go, but this has only a small effect on chemistry. For this reason, we are skeptical that changing the surface density profile would significantly impact the MDFs, but we have not run models with different choices for the surface density gradient, so we cannot say for certain. We refrain from mentioning anything about surface density in our additions to the text.

Sec 3.4, paragraph 4: The description of the SN Ia rate dropping causing an increase in [O/Fe] is confusing. Especially since gas is not mixing between annuli, I would be surprised if CCSNe could raise the [O/Fe] again once any SNIa had occurred? (I think this implies that at some annuli the radial migration is extremely rapid/efficient and mixing stars out of the annulus on timescales far less than the SNIa timescale, right?)

We have expanded and clarified the text on this point. It is true that at some annuli radial migration is efficient and mixes stars on timescales shorter than or comparable to the SN Ia timescale, but that's only one piece of this puzzle. There's also migration on longer timescales coupled with the long tail of the SN Ia DTD - our DTD is nearly a t^{-1} power-law, and in that case only half of the SNe Ia explode between 100 Myr and 1 Gyr, the other half occurring between 1 Gyr and 10 Gyr, leaving plenty of time for radial migration.

It is true that [O/Fe] can increase again once SNe Ia have started to go off; indeed this is what happens during a starburst event (for details, see the Johnson & Weinberg 2020 paper). The reason for this is that it's not the supernovae since $t = 0$ that establish the current chemical composition of the ISM, but rather those that occurred approximately within the previous depletion time (see the Weinberg, Andrews & Freudenburg 2017 paper). In our models, the depletion time is short at large radii due to the substantial mass loading factors there, so the ISM responds quickly to perturbations in the core-collapse to type Ia supernova ratio. We have expanded our discussion in paragraph 4 of § 3.4 in order to communicate this more clearly.

Sec 3.4: I think the Miglio+ 2020 paper suggests that at least all the Kepler over-massive alpha-rich stars are all explained by mass-transfer, but I could be wrong. I think the age-difference scale between this study and the Warfield+ paper are also significantly different, and likely are due to different phenomena, but I may be wrong!

Our understanding of Miglio et al. (2021)'s argument is that at least a substantial fraction of the young α-rich population can be attributed to mass transfer, but a comparison to binary population synthesis models is necessary to make stronger claims. The Hekker & Johnson (2019) paper argues that a portion of the young α-rich stars seen in APOGEE are intrinsically young, because otherwise their data would make sense only if they're the result of mergers on the main sequence. These two papers are consistent with one another if a smaller but still statistically significant portion of the young α-rich population is made of truly young stars not formed via mass transfer. Our interpretation isn't mutually exclusive with the mass transfer scenario, because they would simply explain different sub-populations of young α-rich stars - the model successfully reproduces those which are intrinsically young, but we do not correct our abundances for any binary star interactions. We have added additional discussion to paragraph 5 of § 3.4 to make this distinction a little clearer.

I was confused as to why the 'Late-burst' model was described as having a better agreement to Feuillet et al than the Inside-Out. To an untrained eye they appear to both match the data in some regions better than the other. I think this is further evidence that a more definite bimodality might be achieved by a double burst SFH?

In general, the inside-out model overpredicts ages for solar and sub-solar metallicity stars at $R = 5 - 7$ kpc, an issue which is mitigated by the late-burst model. In the outer Galaxy, the late-burst model better reproduces the C-shaped nature of the AMR as observed by Feuillet et al. (2019). We have revised our discussion in § 3.5 to address this. Perhaps the difference is visualized more clearly in the age-[O/H] relation, which is unaffected by SN Ia enrichment; this is shown in Fig. 18.

I believe that you meant to cite Mackereth et al (2018) in the introduction when referring to the bimodality in simulations. The Mackereth et al (2017) paper which is cited currently concerns modelling of the stellar surface density (and may be of relevance to your discussion of that!).

We have updated the draft to cite the 2018 paper.

Further, the results and discussion in Mackereth et al (2018) may be relevant to your discussion at the tail-end of section 3.3 - since we also concluded that a hybrid two-infall + migration is needed for bimodality, and we suggested the cause of this in the Milky Way was an atypical early mass assembly.

We have added a sentence referencing this.

A couple of typos I spotted along the way:

- **intro: Rejuvinate → rejuvenate**
- **Sec 3.3, para 2: I think the bins are 0.04 dex, not 0.4?**

Good catch - fixed.