Dear editorial office,

Please, find attached the revised version of our manuscript #M140522: "Modeling the process of speciation using a multi-scale framework including a posteriori error estimates". In this revision we have addressed all issues raised by the referees, to whom we are truly thankful for the thorough reading and for the suggestions that helped us in improving the paper. The changes made to the manuscript are colored red for easier comparison with the original version. You can find below detailed answers at reviewers comments and the modifications made to the manuscript. Many thanks,

Mats K. Brun, Elyes Ahmed, Jan M. Nordbotten and Nils Chr. Stenseth.

Referee #1 (Remarks to the Author)

The paper deals with the process of speciation, which is an inportant process within evolutionary biology. The emphasis of the paper is on the posterior error estimation in a multi-scale framework. The paper is well-written and well-organised. The mathematical theories that have been developed are sound. I do not necessarily agree with the modelling framework, but this is not what the current manuscript is about, hence this remark of mine is irrelevant. I recommend minor revisions, and to aid the Authors present their work more clearly, I have the following points that are to be addressed:

1. In general, the abstract may put more emphasis on what the paper is actually about: namely the posterior error analysis. This has only been written in the very last sentences of the abstract. Probably, this is a matter of taste ...

   AUTHOR answer: We agree with this point. The entire abstract has been re-written to better reflect the actual contents of the paper.

2. The diffusion (second spatial derivative) term in the left-hand side of equation (2.2) represents the evolutionary process. I would call g the diffusion tensor instead of the diffusion coefficient (line 180).

   AUTHOR answer: This has been corrected. We searched for other instances of "diffusion coefficient" aswell, but there were none.

3. Since g is the diffusion tensor. It is assumed to be symmetric (Assumption 3.1). Of course, this symmetry facilitates the mathematical treatment with respect to well-posedness and error estimation. It is said that in the manuscript that Assumptions 3.1 follow from biology. Could there be any biological application cases where g is nonsymmetric?

   AUTHOR answer: In thermodynamical systems, whenever sub-scale stochastic processes are modeled as diffusive on a larger scale, one can typically establish that diffusion tensor is symmetric and positive definite due to Onsager's reciprocal relations. To be fair, we are not aware of any analysis establishing the validity of Onsager's reciprocal relations in the current context, and it is conceptually possible that the diffusion tensor might be nonsymmetric. However, we are not aware of any biological system in which this is the case.

   A different perspective on symmmetry is that it implies that there exists a coordinate map of the trait space such that diffusion is isotropic. Conceptually, one could imagine this map being onto the genetic space, and argue that mutations of genes are all equiprobable and thus isotropic. If

one accepts this perspective, then the symmetry of the diffusion tensor is a consequence of the pullback of the mapping.

In other words, we strongly believe that the symmetry of the diffusion tensor is really symmetric, but in absence of a conclusive analysis, we have updated the language in the manuscript to be less absolute with respect to this point.

4. Regarding the boundary condition that n vanishes on the boundary of the trait space (assuming that the trait space is 'large enough'), one could also relax this BC by considering g grad n + A n = 0. Although this BC would relax that n = 0, one has to fiddle around with the A-parameter. Do you think that this alternative BC could do a better job?

   AUTHOR answer: We agree that this is a possible, and most likely the a posteriori analysis would adapt naturally to this approach. However, this introduces a coefficient that most likely is variable along the full boundary of the local domain. As it is anyway a necessary conceptual assumption that a species is relatively local in trait space, we believe that it is conceptually more natural to capture the whole species by introducing a sufficiently large local domain such that zero BCs are satisfactory.

5. Regarding the modelling part. The g-tensor for diffusion is based on 'ordinary' diffusion in which random changes are gradual. In many biological systems, mutations could give large, sudden changes over time, which could be modelled using Levy-jump processes. Then one has to incorporate fractional derivatives. Perhaps the Authors could comment on this. (I am not asking the Authors to carry out the analysis for this case, that would really be too much, the question is just meant to be philosophical).

   AUTHOR answer: Yes, we agree, conceptually fractional derivatives (in place of the gradient operator) can be an attractive modeling option. We have included a comment to this effect.

6. The multi-species model models the transition between species, and in a short subsection (2.3), the relation between the multi-species model and the polultion level model is presented. What I miss om the multi-species model is the actual evolutionary part of the process. There is no random part in this model, which does exist in the population level model through the volatility (diffusion) part. Should not the Authors include a Wiener process (or Levy process to account for jump processes) here (to allow random walk)? Now the approach seems entirely deterministic.

   AUTHOR answer: It is correct that the model considered is fully deterministic. We consider this as a question of modeling scales (with the analogy to particle and Boltzmann models for gas dynamics): Deterministic mean-field theories are suitable at sufficiently large scales, which in this context implies that the population densities sufficiently large that the full stochastic process is sampled. Conversely, if small populations are considered (such as rare species, or similar), the particular instances of the stochastic process will be essential, and stochastic equations are a natural modeling framework.

7. Section 3.3.4, regarding uniqueness. I think that the Authors should add that relation (3.7c) (monotone) makes $||R^k||$ convex, which then proves uniqueness. Formally, the Authors only say that $\varphi = n^k$ makes $||R^k||$ zero (which is the minimum), but this formally does not yet exclude any other possible minima.

   AUTHOR answer: Indeed, thanks for pointing this out to us. Section 3.3.4 has now been updated according to the above comment.

8. Line 528 please, replace 'The figure 2 below' with 'Figure 2 below'

AUTHOR answer: This has been corrected.

9. Your analysis is also built on the existence of statistical moments, hence in the case of a Cauchy distribution (which has a very 'normal' appearance), the model may lead to problems. It could be interesting to do a numerical experiment with a Cauchy distribution function. ($f(x) = 1/\pi \cdot 1/(x^2+1)$ is the standard Cauchy distribution, which has no moments (average, standard deviation ...).

   AUTHOR answer: The reviewer raises a subtle point which essentially asks how traits should be measured at the species level.

   The derivation of the species level model from the population level model rests on two fundamental assumptions; (1) that distinct species can be identified at all times over the global population distribution (i.e., as compactly supported, non-overlapping sub-distributions), and (2) that these sub-distributions are of a known statistical quality (in this case, the Gaussian/normal distribution). As discussed in [1] (where the species level model is derived), the choice of the Gaussian was to a certain degree due to convenience (although individual traits are commonly assumed to follow a normal distribution in the biological literature), hence other statistical distributions could also be accommodated (at the cost of an increasingly complex calculation/notation, see e.g., page 5-6 in [1] where the possibility of using the uniform distribution is briefly discussed). However, for the exact reasons given in your comment, the Cauchy distribution is not a possible choice, although Cauchy distributions do not pose a problem in terms of the population level model.

   However, the question about the shape of the distribution is somewhat misleading: The a posteriori error estimates are intended to assess the validity of the (reduced) species level model, be it good or bad. It becomes clear, by applying any monotone transformation of the axis of the trait space (such as e.g., the logarithm), that in the process of setting up the model, one makes a choice as to how to measure traits. Most classically, one can measure the trait "weight" on a linear or logarithmic scale, and even on a cube-root scale (via the proxy variable "length"). Clearly, the trait distribution will not be Gaussian for more than at most one of these choices. With a poor choice of scale for a trait, where e.g., the trait follows a Cauchy distribution, the species level model will in general be poor, and conceptually the a posteriori error estimates will indicate that the population-level model should always be applied, and no multi-scale algorithm is possible.

   To a certain extent, this is completely analogous to what we see during the speciation event. At this point, the distribution becomes far from Gaussian, and the error estimates indicate that the species level model is not suitable.

   In practice, this implies that for any real modeling of a real physical system, one should advice the biologist to choose a measurement scale for which traits are observed to be close to normally distributed, and provide a warning that if this advice is not heeded, then disappointing results can come from the subsequent analysis.

   We have added comments to this effect in the fifth paragraph of the introduction (pages 2-3), where we discuss the species and population level models, we have added a few sentences outlining the above point concerning assumption (2). We have also added in section (2.3) that here n should be smooth enough for these moments to exist.

10. Proof of Proposition 6.1: I would replace 'By construction, we have' with 'By construction, from relations (6.2) and (6.3), we have'

    AUTHOR answer: Indeed, this has been changed.

11. line 617: please, replace 'leads' with 'lead'

AUTHOR answer: This has been corrected.

12. line 664: The Authors note that, indeed, the Gaussian does not satisfy the boundary conditions. Indeed, if the boundaries are far enough, then, the boundary conditions are approximately satisfied. Hence, then there is no problem. The Authors could use the classical bound for a normal distribution (which is much sharper than Chebychev's Inequality) that

$$P(|X| \geq t) \leq \sqrt{\frac{2}{\pi}} \frac{\sigma}{t} exp(-\frac{t^2}{2\sigma^2}),$$

where $X \sim \mathcal{N}(0, \sigma^2)$. Of course, this relation is based on independence. Using the eigenvalues of the covariance (which is symmetric and hence the eigenvectors are orthogonal), the expression could nevertheless provide some estimation of the upper bound of the initial condition on the boundary. Of course, this is not necessary to quantify directly for this manuscript, but the Authors could incorporate this in their future studies on this matter.

AUTHOR answer: We agree this is not necessary to quantify for the present purposes, but in any case thank you for pointing this out to us, it may very well become useful for future work on this topic.

13. The errors in the figures, as well as the tolerances used in the paper (for instance TOL_{res}) seem to be large to me. Perhaps, they are not large relatively. The Authors should put this into its context, I think.

AUTHOR answer: We chose not to normalize the errors and estimators used in the numerics such that the a posteriori modeling-remainder estimator, $\eta_{rem,i}^k$ (shown in figures 6 and 13) is calculated exactly as given by Theorem 6.3, eq. (6.12d). The reason for such large values is that $\eta_{rem,i}^k$ is scaled by $G^{-\frac{1}{2}}$, where G is the largest eigenvalue of the diffusion tensor g, which is chosen as a rather small constant for both examples (i.e., $5e-6$ and $2e-6$, respectively). We have added a discussion on this in Section 7.2.2, first paragraph (p. 20).

14. Line 768: Please specify which figures you are describing in terms of their labels rather than talking about 'The figures below'.

AUTHOR answer: This paragraph has been rewritten to make it more clear which figures are referred to. Similarly with the first paragraph of section 7.2.2.

15. Please, say something about a comparison of computing times of the heuristic method and the multiscale method in Sectino 7.3.2.

AUTHOR answer: A new remark has been added on p. 23 (remark (7.1)) where we address computation times of the heuristic and multi-scale methods.

16. Figures 11 and 12 seem blurry (or is it my eyes after having read the manuscript).

AUTHOR answer: These figures do not appear blurry to the author (who possess eagle eyed vision, of course). Source files have been double checked, everything should be in order.

17. These were my comments. I appreciate the paper, I enjoyed reading.

AUTHOR answer: Thank you for the positive evaluation of our manuscrupt, we really appreciated your thoughtful comments!

Referee #2 (Remarks to the Author):

This manuscript is concerned with a mathematical modelling for speciation using a general Lotka-Volterra equation modelling the evolution of some ecosystem. This model takes into account species traits and mutations through a diffusion term. This general equation is coupled with a system of ODE modelling the species level: density of species together with the evolution of the mean traits and the covariance matrix.

1. The authors discuss and propose a model for the speciation processes leading to new species. To do so, they make use of a discretized version of the above equations. The discussions on the speciation processes seems to be highly related to the choice of the discrete system and in particular on the reconstruction operator. No connection with the original continuous system is clearly explained neither how it depends on the discretization. Speciation is modeled in term of clustering in the trait space. Does it depend on the traits kept in the model?

   AUTHOR answer: Indeed, there was missing a discussion on in what sense the time-discrete system relates to the original continuous one. To rectify this, we have rewritten the first paragraph of Section 3 (p. 7) where we elaborate on this. Hopefully, this relation is now clear, and the impression that the results are highly related to the time-discrete formulation (which is simply a standard backward Euler approximation to the continuous system), has been rectified.

   As to the clustering in trait space: Yes this depends not just on the traits chosen to model the system, but also on their scale of measurement. We have added some discussion to this effect in the introduction relating also to a comment by the first reviewer.

2. From a modelling view point, how do you choose the selection function $\alpha = \alpha(x, y, t)$? How is it related to the coefficients $A_{i,j}$ in (2.5a)? How the multi-scale parameters are coupled?

   AUTHOR answer: $A_{i,j}$ is related to α through eq. (2.6c), i.e., the effect of species i on species j (as opposed to how individuals with different trait coordinates effect each other, as is the case at the population level). From a modeling point of view, α defines the ecological interactions in the system, and measures potentially competitive, mutualistic, and trophic interactions, and may in general be either symmetric or unsymmetric. Hence, it may take many different functional forms. For the numerical examples in the paper (which are only meant to illustrate the multi-scale framework, not represent any real eco-evolutionary system) we chose in the first case a symmetric competitive ecology, and in the second case an unsymmetric trophic ecology. More details on α can be found in [2]. However it is outside the scope of the paper to address in detail how α should be chosen in order to model different ecologies.

   The strategy for coupling the species and population scales in the multi-scale framework is outlined in section 5 "Speciation", the the main components being the mappings from the species scale to the population scale (reconstruction operator) and vice-versa (compression operator).

References:

[1] Nordbotten, J. M., Bokma, F., Hermansen, J. S., & Stenseth, N. C. (2020). "The dynamics of trait variance in multi-species communities". *Royal Society open science*, 7(8), 200321.

[2] Nordbotten, J. M., & Stenseth, N. C. (2016). "Asymmetric ecological conditions favor Red-Queen type of continued evolution over stasis". *Proceedings of the National Academy of Sciences*, 113(7), 1847-1852.