Dear editor and reviewers,

We thank you all sincerely for your very kind words and equally helpful comments. We are truly glad to be part of such a constructive conversation. As we feel the reviewers have accurately understood and summarized our efforts, we simply address below how we implemented the suggested improvements.

Here is a summary of significant changes:

- We simplified the introductory paragraph on direct and net effects and integrated some of the previous text to the relevant Methods section where the use of equations may help clarify our explanations.

- We improved the clarity of the methods sections, in particular with respect to defining direct and net effects, and with respect to inference (relevant model equations).

- We added a figure in appendix on the relationship between inference and dimensionality of the environment as an illustration to a minor point in Discussion

- We expanded the Discussion, especially regarding the empirical relevance of our findings and connections to the broader context of interaction inference, and partitioned it into subsections for clarity.

Sincerely,

The authors

Review by Hao Ran Lai

The Methods text may need a bit more clarity to guide us through the steps (e.g., at times I was not sure how inference was made / how models were fitted on the simulated data; do the mathematical equations describe either or both the simulation and inference?), but there was nothing major that prevents a reader like me from grasping the full picture.

Thank you for this remark, we have tried to improve the clarity of the methods, especially as pertains to the inference (which equations were used, how was the inference performed).

Here are some specific comments:

For a reader like me who is a frequent user of joint species distribution model (or think in that framework), several of the cited papers in this manuscript keep leaving me a lingering thought "so what does this mean to my JSDM or co-distribution datasets?" The authors did quite a good job outlining JSDMs in paragraph 3, as well as their shortcomings citing Poggiato et al. 2021, but in the Discussion I find myself still hungry for more insights from the authors. From here I learned that dispersal limitation (or the lack thereof) is key to whether the results from JSDMs are reliable in inferring biotic interaction, but which results is it from a JSDM? As a user of JSDMs, it is still unclear to me whether we could reliable interpret the latent associations (or residual components) as biotic interactions / associations not accounted for by the environment (assuming that we have measured all the important abiotic covariates), or do we include neighbour densities as measured covariates. Knowing Leibold et al. (2022) Oikos, I was waiting for a bit more insights into the way

Indeed, while jSDM are a highly active field that is very close in some ways to our issues, and it would seem like an oversight not to refer to these efforts, we are not experts in their use and we were somewhat prudent in attempting prescriptions. Clearly, the challenges outlined here can only relate to the interpretation of latent associations as biotic interactions, and do not imply much about using latent associations for predictive purposes. We now provide some thoughts about JSDM in Discussion section 3.1, indicating that residual co-variation across the whole landscape is probably not the right pattern to look at if we want to infer direct species effects instead, but that is not necessarily the goal of every inference.

For us, these results do imply that attempts at inferring precise pairwise coefficients are not likely to succeed in the environment-tracking scenario, even though inferring overall statistics works, so it is an invitation to judge if empirical knowledge suggests that this scenario is probable or not before interpreting the results of an inference. A statistical interaction term between biotic interactions and environment is certainly an option worth testing in simulated data like ours in the future (it would not be perfect since the model would be mis-specified compared to the ground truth model, which has a more complex structure). I do not know how that would be managed in a JSDM for instance where we are instead looking at residual covariation.

You are correct that the term of community matrix is ambiguous, this usage to mean interaction matrix (or more precisely Jacobian of the dynamical system) is unfortunate, so we have changed it.

Thanks for this remark, we now try to define it summarily after its first mention.

We were referring to the "species sorting" paradigm which is one of the categories in which metacommunity theory classifies prior theories, where species presence or abundance is simply

predicted by how favorable the abiotic and biotic environment is; this typically assumes that the system is close enough to equilibrium to be predictable, and not dominated by other factors such as dispersal. However, we realized that our regime does not exactly match species sorting in the sense that we still allow dispersal limitation, so we removed that reference.

- Does the last summation term (N of y - N of x) imply mass effect?

Indeed it simply indicates immigration from neighboring sites y minus emigration from site x, proportionally to abundance. If this term is larger than others, then we are in a pure mass effect situation.

- Figure 1e is interesting. Do you think it is true / would help to highlight that the dispersal limitation on more superiour competitors is also crucial to the persistence of an inferior focal species in a patch?

Indeed, this is crucial to have a variety of compositions in the same environmental conditions. We now highlight this in the caption of Figure 1.

- Perhaps refer to Equation 2 in Section 1.4, because I found it confusing what the actual inference is.

We apologize for this omission, Section 1.4 was indeed quite incomplete and we have now clarified it significantly, including a reference to equation 2.

Then, throughout the manuscript I constantly thought about multicollinearity, from a data perspective. I am not able to pinpoint this to any place of the manuscript, but let's say the 2nd Discussion paragraph. Here the authors mentioned "...if the same environmental conditions predictably lead to the same species composition..." Under environmental tracking, if one sample communities in the field they would need to face strong autocorrelation / multicollinarity between neighbour densities and the abiotic covariates. Is this true in the authors' simulated data too? Is this the statistical phenomenon that leads to confounding variables / non-indentifiability during the inference process? These are the questions that haunts my sleep whenever I fit a model including both neighbour densities and environment, so I wonder if the authors have any insights. From the top of my head, I also couldn't think of any literature that points out that aggregated statistics may be robust against multicollinearity even when individual coefficients are not, so I wonder if this is another way of highlighting the novel contribution of this work.

Absolutely, we believe that this non-identifiability due to multicolinearity is why the inference of individual coefficients is very wrong in the environment tracking scenario, though it is not obvious at all that we should still get the right mean, and even worse, the right variance (indeed the variance is typically overestimated and we may need to find better ways to estimate it). We now explicitly mention multicolinearity.

The text on equilibrium and "letting indirect effect play out" in the Discussion is really insightful; I hope it was highlighted more clearly / early on.

Thank you, we now highlight it as early as Methods section 1.3, and summarily in introduction.

Review by Malyon Bimler

I was a little lost in the paragraph on net vs direct effects but found my way back in the following paragraph – maybe that section could do with a bit of wordsmithing to make it easier to follow, but that is my only criticism of the introduction.

Thank you very much for your comments. We hope the rewriting makes this section clearer, and we have moved some of the content to discussion.

As someone who studies and encounters a lot of facilitation, I would like to see a simple statement early on that predominantly competitive interactions are considered.

That is good advice and we have now added such a statement, though we expect most of our conclusions points to hold for positive interactions in principle. The main reason for not considering them here is that we feel the kind of very simplified dynamical model we used in our simulations (e.g. generalized Lotka-Volterra equations) fail more to encapsulate the ecological features of facilitation than they do with competition. This is true in some simple senses (e.g. strong enough facilitation causes infinite abundances in L-V models, but that can be corrected easily), and in less simple ones (e.g. the dynamics of purely facilitative systems, without any costs or compromises, tend to boil down to just dynamics of total biomass, see Gao et al Nature 2016) that are not very relevant to our message here and that we did not want to possibly get into.

Maybe remind the reader in caption of Fig 2 that this presents a case where dispersal is present? It is mentionned in the main text but not the caption.

This is now mentioned.

The authors regularly remind the reader that their simulations present a 'best-case' scenario, which I appreciated. I could have also done with a couple of further reminders on what the authors consider net vs direct effects (e.g. at the beginning of the results).

We now do so at the beginning of the results section, as well as clarifying this distinction in introduction.

Unfortunately I did not have access to the code to run the simulations myself.

Our apologies, we put a link to the code in our submission but failed to include it into the submitted version of the main text, as is now done (https://github.com/mrcbarbier/morefromless)

The beginning of the discussion focuses on justifying their choice and approach and re-contextualising it with the issues faced when trying to infer the community interaction matrix from empirical data. Their claims are well-supported by the results and they link their findings to more recent studies. They finish with a paragraph on how their findings may be helpful for parameterising other models. I am not entirely sure what to do with their conclusions in regards to the more empirical aspect of my work: it is reassuring to know that even if the interactions I estimate from empirical data can be incorrect, the overall mean and standard deviation are likely close to correct, but I am not sure if there is anything more I can apply directly to my research. The authors leave the reader to make their own conclusions in that regard, but given how much they ground their study in the problems faced by empiricsts I wouldn't mind hearing more of what the authors think, even if it is purely speculative.

We have expanded the discussion to situate it better in the context of empirical issues with estimating interactions, but we readily admit that we are only dealing with a small corner of that problem. It is likely that our results do not yield much advice for, say, an empiricist concerned with inferring any specific biotic interaction between two given species, where presumably ecological knowledge and using other data sources (e.g. direct records of interactions or growth/mortlaity) is more likely to matter. On the other hand, we feel that this is indeed a message of hope for empiricists interested in estimating the overall intensity of interactions in a community as a whole, e.g. as part of figuring out whether community composition is an indicator environmental states or of internal dynamics. It also suggests to develop new methods that strive to estimate community-wide statistics of biotic interactions rather than all pair-wise coefficients.

Supplementary Materials
Section A paragraph 2 – was that done for each individual species?

That is correct, we now clarify this.

Review by Frederik De Laender

1.1: "their abundances are set to zero, and only allowed to vary if their net growth rate $dlogN_i/dt$ becomes positive". At first, confusing to me (I guess the verb "vary" got me sidetracked).

Thank you kindly for your comments. We have tried to make that sentence more legible.

As I understand the simulation protocol, there is no regional equivalence: some species are inherently more fit than others. Intuitively (but I did not think this through properly), I'd say regional equivalence without dispersal (species do differ at the pixel level, but each species persists in the same fraction of pixels) affects inference of species interactions because it affects to what extent the landscape represents the experiment one would need to infer interactions.

There is partial but not perfect regional equivalence, in the sense that no species is best everywhere, but some species are indeed better off on average due to the species-specific mortality constant m_i. This indeed means that we do not explore all possible compositions equally, contrary to what we would get with full regional equivalence. In turn, that would still be just a fraction of a full "biodiversity experiment" trying out every possible species composition. We were actually somewhat surprised by the fact that the dispersal-limited scenario allowed such good inference of interactions despite this (though this is heavily helped by the fact that we know the model from which the data is generated, it has purely pairwise interactions, and there is no measurement error).

I wonder why species in the dispersal limited case were only seeded in 50% of the patches. I'd think that seeding them everywhere (in combination to regional equivalence) can produce patterns that permit more reliable inference.

They were seeded throughout the landscape with a 50% chance of being in any given patch; if seeded everywhere, since there is no stochasticity otherwise, we would simply recover environment tracking (every patch would reach the composition selected by the local environment and interactions from the entire regional pool)

I have the same intuition as the authors that more complex models (displaying more complex dynamics) will not improve inference. However, I'm not sure how model complexity influences

That is a good point, we also believe that the complexity driven by the large number of species largely overrides the intrinsic complexity of the model. In fact, experience with disordered systems suggests that having many species may tend to reduce the important of how complex any given interaction is (since what matters is the aggregate effect of many of them, leading to a kind of central limit theorem), so we might even expect the difference in inference quality between models with complex and simple interactions to shrink with species number. This is speculative, but now briefly evoked in Discussion.

Absolutely, it is now referenced there (in its recently published version).

As this point was clearly lacking, we now include some discussion of consequences from an empiricist's point of view. We hope the following paragraphs quoted from Discussion sections 3.1 (for the choice between direct and net effects) and 3.2 (for "what one gets by asking for less") clarify our thinking in that respect:

> Our work stresses the importance of correctly specifying which concept of biotic interaction one is trying to infer: for instance, estimating context-independent direct effects is sometimes possible even when net effects vary dramatically across the landscape. This is of particular relevance to statistical approaches focusing on the co-distribution of species, e.g. joint Species Distribution Models (Ovaskainen et al, 2017). The residual covariance between species across the whole landscape can be understood as the spatial aggregation of locally varying net effects which we believe (see Appendix: Estimating interactions from residual species co-variation) is not an appropriate path to deduce direct effects.
>
> [...]
>
> From an empirical point of view, our results thus carry encouraging as well as cautionary notes. Species-level questions may require the precise inference of a given pairwise interaction, and our work comes here as a warning to keep in mind possible barriers to that inference when we do not have grounds to assume `natural experiments' such as permitted here by strong dispersal limitation. On the other hand, we feel that for empiricists interested in estimating the intensity of biotic interactions in a community as a whole, e.g. to know whether community composition is a better indicator of environmental states or internal dynamics, our findings bring some home and a suggestion to turn to methods that strive to estimate community-wide statistics rather than individual pairwise species interactions.

Good point, we now indicate the species number in the corner.

We retained them as they are slightly more useful in the extended version of this figure in Appendix, Fig S4, but we can remove them if that is preferable; otherwise we feel they do not detract from legibility.