**Editor comments:**

I have now received two review reports from established experts in this field. Both have agreed on the timely contribution of this work and recommended the necessity and possibility of a substantial revision to address a number of weaknesses in the current manuscript. I also agree with the two reviewers and suggest that the authors give due consideration to each comment raised. In particular, Reviewer #1 raised several issues related to strengthening the Discussion, model selection, and the potential problem of using the z-score as a predictor. Reviewer #2 highlighted the poorly explained/annotated SI material, the limit of quality sampling in addressing false absence problem, the challenge of model validation, and the need to clarify interactions that were predicted to be highly/most likely but not observed in the data. I, therefore, recommend a major revision for the authors to take in these comments.

Thanks for these comments. We addressed all specific reviewer comments as indicated including revising the discussion, model selection, the use of z-score and the annotation on the SI material, and the challenges in model validation. These are all addressed in detail below in response to each reviewer's comment.

**Referee #1**

Comments to the Author

I have read this manuscript with great interest, and (because the file with the appendix was not quite anonymous initially) I have also spent some time reading other papers by the authors. I refer to a few of them in my review. The authors are tackling an important question, but I am a little underwhelmed by the current submission.

I think the question is certainly timely, but the manuscript lacks a strong message, and suffers from ambiguities or lack of clarity in describing the methods, which leave me unsure as to what it contributes. I have no doubt that the authors have the know-how to fix all of these issues, although it will require (I think) a very substantial revision.

Thanks for your careful consideration, interest in the manuscript and confidence in our ability to revise this manuscript. We have revised the writing to help flesh out the main messages and addressed all methodological concerns below. We feel that these comments greatly improved the manuscript.

One of my core concerns with the manuscript is that the discussion is weak. Not in the sense that it is poorly written, but it doesn't really go anywhere. I was looking forward to more actionable ideas, or to circling back on the initial issue that is raised in the introduction. What have we learned now that this manuscript exists? What do we need to do next? What are the consequences of these results?

This is very important, and we have revised the discussion substantially (see details below). In particular, we revisit our hypotheses and try to leave some concrete ideas for future research (e.g., developing model integration of pairwise interactions and network properties).

At times, it really felt like the manuscript was more pessimistic than the results warrant. Overall, the ability to transfer is pretty good. The models work well enough given the low amount of data that goes in, and the fact that the models of interactions do not necessarily predict the structure of the networks is not entirely surprising. It has been discussed in previous papers involving some of these authors. Maybe there is something that eludes me, and it is, actually, a surprise, but the introduction is not lining up a series of expectations for this.

Agree that this is not entirely surprising, but also hasn’t been well tested yet in the literature. We reworked the phrasing (see below).

This would be easier to solve with a series of well laid-out take-home messages at the end of the discussion, and with much tighter writing in the introduction. There is a lot of information given, but not in a way that makes following along intuitive for readers. Even the title is not that informative (I had a different idea going in as to what the manuscript would cover).

Specifically, we addressed the concerns expressed in the previous three paragraphs by (1) better explaining our expectations in the Introduction (e.g., “if trait-matching rules determining interactions are general, then i) we should observe the same traits-interaction relationships across regions and ii) a model trained in a location should be transferable to another one." (L61-63)); (2) completely rework the first paragraph of the Discussion to better highlight the two main results (L380-389); (3) more explicitly explain how they contribute to ecological theory in the Discussion (e.g., “Our ability to predict food web interactions across contrasting environments suggests that spatial food web variation is mainly driven by changes in the distribution of functional traits, and less so by the variation of trait-interaction relationships.” L392-396); (4) give an idea on how future models could use predictions of food web structure to make better predictions of pairwise interactions (“*For example, we could penalize posterior food web predictions resulting into improbable food web structure to update pairwise interaction probabilities*” L447-449); (5) changed the title that now better state the two main results: 1- trait-matching predict pairwise interactions; 2- food web properties are not as well predicted. We believe that the changes in the introduction better present our expectations, put more emphasis on the overarching hypotheses, and how our findings contribute to the literature. This, with the changes made in the Discussion, helps identify the take-home messages and how they stem from the results, and propose actionable ideas. We changed the tone of the Discussion, and by more clearly stating our contribution to theory, the promises of trait-matching models to predict future communities and communities when we know nothing about their food web, and proposing future extensions to improve our models, the Discussion should sound a lot more optimistic.

One thing that surprised me was the lack of consideration for spatial scale, especially because the authors have written so much about network area relationships and networks in space. I was surprised to see networks of different areas compared without nuance. Would the area change the mechanisms involved in species interactions? Would it change the shape of the network, and therefore explain (part of) the results about the lower predictive ability of the networks themselves? In some cases, this bears direct methodological consequences on the variables the authors use in analyses (for example, when they get data from WorldClim).

We now introduce how spatial scale may influence food web structure in the Introduction (“*These higher-level properties emerge from the joint effect of trait-matching between interacting partners, the trait distribution of species composing the food web (Gravel et al., 2016), and are influenced by the spatial scale of the network (Galiana et al., 2018)*” 77-80), and discuss the potential implication for our result in the Discussion paragraph starting L425. That said, area is expected to constrain food web properties (e.g., degree distribution) and not pairwise interactions (e.g., a wolf could eat a rabbit regardless of the spatial scale of the food web). We, therefore, agree spatial scale is warranted to be introduced and discussed, but we did not make it a central theme of the paper. For the methodological consequences, we addressed those in the later comments.

I was also surprised to see no discussion of a central piece of the manuscript, i.e. the imputation of missing traits. How many traits did you need to impute? Across how many species? What is the performance of the model you used for imputation? It also must be discussed when talking about the results because you are not simply predicting interactions: there is part of the predictors that are, themselves, imputed.

We added appendix S2 with a lot more information on trait coverage and imputation. Only a small proportion of traits are imputed and given the good performance of the imputation (now given in the new appendix), it should not influence the results of the manuscript. We also, specified, for each class and trait, the proportion of missing values in the appendix. Since it should not influence our results, and that they do not relate to our hypotheses and conclusion, we preferred to keep most of the discussion about trait imputation in the Supplemental material.

The training of the interaction model would also benefit from a better description. I think the authors are following some of the recommendations from<https://can01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdoi.org%2F10.32942%2Fosf.io%2Faty7n&data=05%7C01%7Cdominique.caron%40mail.mcgill.ca%7C18b7b712054e4d98180408db5b8ad03c%7Ccd31967152e74a68afa9fcf8f89f09ea%7C0%7C0%7C638204423983683227%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=504k3Wg4ZxbiVztE3rCXQDysTu0EjIMn2Mf%2FTpJA3KM%3D&reserved=0>, but I wasn't sure exactly what they did to select a training and testing set. I would like to see this section extended to make sure that the training is appropriate.

We reworked the section explaining the splitting of the data in the manuscript, and better describe the model (L172-185). We also have Appendix S3 that details the model with a subsection that we extended about the definition of the training and validation sets.

This brings me to my next issue: the authors rely on a single model, but often discuss their results as being very general. Seeing this particular problem as a machine learning problem, it would be inappropriate to stop with the first model, especially because the few papers that used different models have reported that they differ in their ability to predict interactions or networks, which is a core point of this paper. I have no issue with presenting the "best" model in the main text, but I think it would make sense to use as many models as the authors can justify. Otherwise, it would be fair to tone down the conclusions, as there is no way to know what is a genuine ecological challenge, and what is a limitation of GLMs.

We added a second type of model to make sure our findings are not specific to GLMs. We used boosted regression trees, a machine learning algorithm allowing non-linear relationships that have been used before to predict interactions (e.g., [Laigle et al. 2019](https://onlinelibrary.wiley.com/doi/full/10.1111/oik.04712), [Pichler et al. 2022](https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.13329)). We tried to tone down our conclusions by using “our models” or “trait-matching” models throughout the discussion to specify that our results might not apply to models using other types of predictors.

Finally, there is an ambiguity in the methods section about feature engineering. The authors essentially use the z-score of some features as a predictor (modulo the factor of 2 on standard deviation), but is this done within or across networks? In other words, if there is a species shared across the four networks (I know there is probably not, but entertain this idea), would it have the same "corrected" trait values every time? If yes, then I have no issue with the approach (but it must be stated); if no, then this is an issue, as the scale of the predictors varies across each prediction task (regions, or ecosystems). This is potentially a breaking flaw in the paper, in that it might lead to entirely different results (and again, if explored with different models, some may be more robust than others).

For each model, we scaled the training set using the mean and sd in the training food web. When doing the prediction, we scaled the predictors using the mean and sd in the training food web. For example, if we used the Serengeti model to predict the Europe food web, we scaled predictors of European species using the mean and sd in the Serengeti food web. In other words, the units of predictors are always in SD units of the predictors of the training set. This approach has the following two advantages: (1) We keep the advantages of using z-scores for the training process, and (2) for a given model, if a species is present in many food webs, it will have the same set of predictor (i.e., the model will predict the same probability of interaction between the same two species independently of the trait distribution it is present in). We added this information in the Appendix S5.

I have a series of comments on specific parts of the text, most of them minor:

Line 45 - this is true for the number of potential interactions, but not for the number of expected interactions. Even assuming a power law, we expect fewer than S² interactions to actually happen.

We agree that the expected number of interactions is much lower than S². However, in the manuscript, we explicitly specify that it is the number of *possible* interactions that increases quadratically. The total number of interactions and non-interactions to document is therefore exactly equal to S² (including cannibalism).

Line 51-53 - these are rather weak justifications, you can cut this paragraph to make room for more important content.

We completely cut this paragraph. We used the space to focus on the overarching hypotheses of the paper and explain our specific expectations.

Line 64 - this feels like the right place to mention that traits also have phylogenetic signal.

We added that phylogeny can be used as proxy for trait relationships (L56).

Line 72 - I agree with this statement. I was surprised not to see a discussion of the recent preprint by some of the authors of the current manuscript (<https://can01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdoi.org%2F10.32942%2Fosf.io%2Fvyzgr&data=05%7C01%7Cdominique.caron%40mail.mcgill.ca%7C18b7b712054e4d98180408db5b8ad03c%7Ccd31967152e74a68afa9fcf8f89f09ea%7C0%7C0%7C638204423983683227%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=gX1HyNiIzJV4fk3sncjXTUx9%2FzjZTk62Dk%2BoJf%2BmfEc%3D&reserved=0>) which in part addresses this question. More broadly, this paragraph felt strange to read, as it comes up (especially in the last sentence) as critical of previous work, some of which involved the authors of this paper. It is always easy to criticize previous work because the progress of research is incremental. This paragraph would be a lot more interesting if it were phrased in terms of what the authors introduce as the next step.

We rephrased many of the sentences in the paragraph to make it clearer (L60-72), to emphasize the contribution of previous work and to discuss the remaining unknowns about trophic interaction transferability. We believe it comes up as less critical, and more as building upon previous work. As such, regarding the first part of the comment, we decided to mention Strydom et al. (2022; <https://doi.org/10.1111/2041-210X.13835>) instead of the preprint highlighted by the reviewer for three main reasons: first, it tests the transferability of the model proposed in the preprint which is closer to the question we address in the current contribution; second, the paper is now published which is usually preferred because it has gone through peer-review; third, GEB has a limit of 50 references (which we have slightly exceeded to answer the comments for both reviewers), so we decided to cite a paper we already cited and stems from the same research group.

Line 88 - "trained", not "calibrated"

Corrected throughout the manuscript.

Line 90 - use regions here (or ecosystem earlier in the sentence); there are numerous places where the manuscript would benefit from another read-through to make sure the language is homogeneous.

We changed the ecosystems for regions throughout the manuscript. We also re-read the entire paper to identify and correct inconsistencies in the language.

Line 98 - this last sentence was transparently written with full knowledge of the results. I may agree with the prediction, but if you make the prediction here, you need to articulate why you expect this to happen. If I am not mistaken, there is a section in<https://can01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdoi.org%2F10.1098%2Frstb.2021.0063&data=05%7C01%7Cdominique.caron%40mail.mcgill.ca%7C18b7b712054e4d98180408db5b8ad03c%7Ccd31967152e74a68afa9fcf8f89f09ea%7C0%7C0%7C638204423983683227%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=J%2FXjUm9lYNHNP5wQW80O2QT1xuLGPxUqU2uTx8NA9xo%3D&reserved=0> (which shares a few co-authors with this present submission) about the issue of predicting networks and interactions at the same time.

We added explanations with supporting references for both of our predictions. Why we expected these results should be more clear now.

“if trait-matching rules determining interactions are general, then i) we should observe the same traits-interaction relationships across regions and ii) a model trained in a location should be transferable to another one.” (L61-63)

“This discrepancy could arise from fundamental differences in the scale of ecological processes that constrain food web properties (e.g., the number of feasible interactions) from those that drive pairwise interactions, which is commonly the focus of trophic interaction models” (L85-88)

We also think these explanations make the implications of our results clearer in the Discussion.

Line 108 - you cannot know that it is "all" species. Please tone down.

Changed to “all known terrestrial vertebrates in their respective location”. (L113)

Line 111 - this is a potential issue with the integration of this food web; what happens if you remove it from the analysis? Is it changing the results significantly? Is there a lesson to learn about whether trophic groups are appropriate?

We added a sentence in the *Food Web data* section on how we expect false-positive to be more prevalent in the Serengeti food web given species are grouped in trophic groups (L121-122). We also highlight the challenges of predicting interaction between trophic groups in the Results. The challenge is that we use species-level traits whereas the interaction data are at the trophic group resolution. This difference in resolution likely creates discrepancies between the predicted food web and the observed food web.

Line 138 - the "all" is misleading, as in the very next paragraph you mention performing imputation. Please clarify.

We added “available” at the beginning of the sentence and changed “all” by “every”. We also added information about trait coverage in the text and referred to Table 1. Finally, we added an Appendix (Appendix S2) to document further trait coverage and information about the performance of missForest.

Line 147 - I was expecting to see at least a supp. mat. about the performance of this model, since it is a major methodological step.

We added Appendix S2 with information about trait coverage and imputation. It includes information about the coverage of each trait for each missForest run and the performance of the imputation step.

Line 159 - unclear; why would you use co-phenetic distance to talk about functional dissimilarity, when you have information about traits?

We understand the confusion here. When we wrote “functional” we were not only referring to the traits, but also the “function” (interactions) of the species. We changed “functional dissimilarity” to “ecological processes like species interactions” to avoid confusion.

Line 202 - I was expecting to see the figures showing that the chains converge. I would like to see them as a supp. mat.

With respect to model convergence, in Appendix S6, we provide trace rank plots which are similar to trace plots and are used to assess the mixing of the chains. These types of plots are now preferred to trace plots as they are easier to interpret (e.g. McElreath, 2020; p.284). We also provide the rank-normalized potential scale reduction factor on split chains as histograms for simplicity (each parameter has its own Rhat, and with random slopes and intercepts, there are many Rhat for each model) which shows the convergence of all parameters. Finally, we also provided Posterior predictive checks to compare the distribution of the observed data to the distribution of the posterior predictive distribution. A discrepancy in distributions can be caused by a model misfit. We added an explanation of the different diagnostics at the beginning of Appendix S6.

For reference, see:

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P. C. (2021). Rank-normalization, folding, and localization: An improved R ̂ for assessing convergence of MCMC (with discussion). *Bayesian analysis*, *16*(2), 667-718.

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. 2nd edition. Chapman and Hall/CRC.

Line 207 - "food web fitted" is very awkward; what about "the food web predicted by the model trained on ..."?

We changed the sentence to “We used each model to predict the food web on which it was trained and that of the other regions considered” (L218-219)

Line 221 - there is a good discussion of the problem of false negatives in<https://can01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdoi.org%2F10.32942%2FX2DW22&data=05%7C01%7Cdominique.caron%40mail.mcgill.ca%7C18b7b712054e4d98180408db5b8ad03c%7Ccd31967152e74a68afa9fcf8f89f09ea%7C0%7C0%7C638204423983683227%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=2d0vAk2qPwhQTKDIB%2FqsaB6%2F%2BKSaNbDD27Zb8P3F36I%3D&reserved=0> (which shares one author with this submission), I was expecting to see a more nuance discussion of false negatives at this point in the manuscript.

Since each food web is potential, and completed by expert knowledge, we expect them to be less sensitive to sampling problems than in food webs based on observations. We added this nuance to the methods:

“Since all food webs document potential interactions (i.e., the predator could feed on the prey species), and use expert knowledge to define these trophic relations, we assume that they are less sensitive to incomplete sampling than observation-based food webs (i.e., interactions and non-interactions are similarly uncertain).” (L115-118)

We also added a short discussion point about how the predicted interactions could be interaction missed by experts.

“Most food webs, even expert-based, are missing some interactions (e.g., cryptic and opportunistic interactions), so the additional interactions predicted by our models could be candidates for further investigation.” (L412-416)

We agree that the preprint suggested by the reviewer is relevant, but we did not cite it as we already exceeded the 50 references limit, and we consider that the other references are more critical to our manuscript.

Line 225 - the Euclidean distance would be appropriate on a flat Earth. The geographical distance would be far more appropriate here, and would also allow expressing coefficients of the models as X per kilometer (for example).

Good point, we now use the great-circle distance (Haversine formula) instead of Euclidean distance.

Line 228 - this ties to my point about the scale of networks, but for the smallest networks, are you not resampling the same environmental values ad infinitum? Furthermore, please specify the resolution for the data you used.

To make sure the random sampling is not a problem, we changed it to extract all the Worldclim data in each food web. We also added the resolution of the Worldclim data. Because we are simply using the “environmental centroid” to calculate distance, the coarse resolution should not be a problem.

Line 271 - I had trouble understanding what you were doing in this specific instance. I would suggest a rewrite.

We reworked it to make it clearer. We also added a description paragraph and information in the figure captions in appendix S8.

Line 274 - please do discuss the intercept.

Added

“We expect an intercept of zero for unbiased predictions of species role. An intercept greater than zero would suggest an overall overestimation of the role metric, whereas an intercept less than zero would suggest a systematic underestimation of the role metric.” (L278-281)

Line 278 - please write out the full model.

As mentioned in the main text (L277), it is a simple linear regression between the predicted position and the empirical position value of each species in a food web. There are no transformation or additional predictor, so we don’t believe that writing out the full model will help.

Line 292 - please give the PR results as well (every time ROC-AUC is mentioned).

We added the AUPRG results in the text. AUPRG uses the precision to recall curve, but has the advantage of having a common baseline (0 for random prediction). The baseline for the area under the precision-recall curve is the prevalence of positives in the validation set. Therefore, we cannot compare model performance with different predicted food webs. See Siblini et al. 2020 (<https://doi.org/10.1007/978-3-030-44584-3_36>) and Flach and Kull 2015 (https://proceedings.neurips.cc/paper\_files/paper/2015/file/33e8075e9970de0cfea955afd4644bb2-Paper.pdf) for reference.

Line 297 - at this point, you can discuss whether the trophic groups in the Serengeti food web is an issue.

We added a sentence in the methods about how trophic groups in the Serengeti might influence false positive rates:

“Thus, we can expect more false positives (i.e., non-interactions documented as interactions) in the Serengeti than in the other three food webs” (L121-122)

and a discussion about how the transferability might be more challenging because the resolution of the food web is not the same as in other webs:

“The transferability of the Serengeti food web was expected to be more challenging because nodes do not represent species (as is the case for the other food webs), but trophic groups” (L318-320)

and finally, some implication about the interpretation of the prediction of food web properties:

“However, our models overestimated the number of interactions by two-fold even for the Serengeti food web which probably already overestimates interactions since species are grouped into trophic group” (L414-416)

Line 300 - this needs to be a main text table. These results are too important to be in the supp. mat.

We added AUPRG in Table 1. Because all the other metrics are positively correlated to AUC and AUPRG (shown in Appendix S7), we think it is not necessary to overwhelm readers with a suite of 7 performance metrics. Interested readers can refer to Appendix S7 where we detail each metric and report the results.

Line 305 - when you discuss coefficients, please do provide an interpretation of what they mean, ecologically.

We added a sentence at the end of the paragraph that summarizes the interpretation of the coefficients.

“Overall, these results suggest that model transferability decreases with geographic distance mainly because the environment and phylogeny diverged between the predicted food webs and the ones used to train the model.” (L332-334)

Line 307 - cut the first part of the sentence, if you end up concluding that the effect is removed when the appropriate corrections are applied.

We are reporting both the *direct* and *total* effect of geographic distance because they tell a complementary story: Transferability decreases with geographic distance (total effect), but mainly through the indirect effects (through phylogenetic and environmental differences). In addition, we added a sentence that summarizes the paragraph and should make the nuance clearer (see response to previous comment).

Line 311 - how? This sentence felt disconnected from the methods and from the rest of the paragraph.

We rewrote this sentence to make the connection with the rest of the paragraph clearer:

“We also analyzed transferability at the species level with comparison of species-specific performance as a function of phylogenetic and functional distance between the focal species and the species pool used to train the models (Figure 3).” (L336)

With this modification, we hope it is now clearer that this section is associated with which paragraph in the methods.

Line 336 - if the slope is close to 1, this should be good news, according to the way you talked about this analysis at line 274.

We added information in the methods on the intercepts (see previous comment). A slope of 1 is indeed good news, but an intercept greater than 1 suggests there may be an overestimation of the role values for all species. We removed the mention of the slope to avoid confusion on the coefficient related to our interpretation.

Line 342 - this ("our interaction models") is absolutely the responsible way to frame these results. I stand by my main point that the authors need to experiment with more models, but the discussion should put more emphasis on the fact that this is a single model.

We changed to “our model” throughout the discussion. We also added a second model type (BRTs) to make sure our conclusions are not dependent on the modelling framework.

Line 355 - please quantify "reasonably well".

Added AUC and AUPRG

Line 355 - the sentence starting from "Our results..." does not seem to be justified by something that was explicitly discussed before. Please clarify.

We completely rewrote this paragraph. We now focus solely on stating the main results with a summary of the implication as last sentence. We state the implications, explain how they emerge from the results in the following paragraph. We also believe that we clarify our expectations in the Introduction (see previous response) which should also help with linking the results to the implications.

Line 367 - I do not agree with this sentence; it may be true, but is unsubstantiated. The changes in the distribution of traits can be measured, as you hold the information about the traits.

We agree with the reviewer that our formulation did not have strong foundations. As such, we re-wrote this sentence and added another one before to clarify this discussion point. The new text now reads:

“Theory describes how the variation in food webs arise from two key processes: trait-interaction rules (e.g., body mass relationships between predators and their prey) and the distribution of traits in different systems (Gravel et al., 2016). Our ability to predict food web interactions across contrasting environments suggests that spatial food web variation is mainly driven by changes in the distribution of functional traits, and less so by the variation of trait-interaction relationships.” (L391-396)

These ideas are also now better explained in the Introduction.

We then finish the paragraph with the implication of these generalities in trait-matching:

“These generalities in the trait-interaction relationships across space and within taxonomic groups demonstrate the potential for using data from well-studied areas to predict interactions in areas we know very little about or forecast (and hindcast) food webs given new trait distributions.” (L402-405)

Line 394 - add "our" to "trophic interaction models"

See previous response.

Line 410 - unless you specify how this information can be used, this is speculation.

We detailed what Leung et al. did for SDM and provided an example of how structural food web models could be used to update posterior distribution.

Line 414 - as previously mentioned, there is a similar discussion taking place in<https://can01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdoi.org%2F10.1098%2Frstb.2021.0063&data=05%7C01%7Cdominique.caron%40mail.mcgill.ca%7C18b7b712054e4d98180408db5b8ad03c%7Ccd31967152e74a68afa9fcf8f89f09ea%7C0%7C0%7C638204423983683227%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=J%2FXjUm9lYNHNP5wQW80O2QT1xuLGPxUqU2uTx8NA9xo%3D&reserved=0>; please do clarify the new points that are introduced by this manuscript

Indeed, Strydom et al. (2021) introduce and reviews many of the ideas for predicting ecological interactions and network prediction simultaneously but did not test any of the ideas. Here, we are testing the generalities of trait-matching relationships which is fundamental if we aim to use the trait-matching model to extrapolate in the future (or past) and to new places. We also test the prediction of food web properties by predicting the pairwise interactions (*do parts make the whole?*), which has never been tested before. We believe that the modifications we did to explain and clarify the manuscript’s take-homes and hypotheses (see previous response) should clarify the new points made by our manuscript.

Line 436 - please do cite the other recent research; it was unclear what you were thinking about

We added three papers.

Line 441 - I found the last sentence a little weak; what should we do next? What are actionable next steps?

We rewrote the sentence to be more actionable:

Future research could better link subfields within trait-based ecology for a comprehensive understanding of how species traits and network structure relate to their environment and enable trait-matching approach to inform restoration and re-wilding initiatives on the choice of appropriate species to re-establish former links and potentially ecosystem functions related to these interactions. (L476-480)

Appendix - the figures are unreadable in black and white or grey scale, please pick other colours.

We modified the colour slightly to make them more readable in greyscale. That being said, we are still aware that the readability of the figures in greyscale is still non-optimal. Since appendices are rarely printed, we preferred colours that are accessible to visually-impaired individuals that may read them online. chose the colours of the plots to make them colour-blind friendly.

**Referee #2**

Comments to the Author

The authors use species traits and interactions in a `known' food web to model interactions (and the resulting food web) in other food webs of varying geographic and phylogenetic distance. They find that species interactions are generally predicted well, but the predicted food webs tend to be more connected, less modular, etc. than the original models. Overall the manuscript is very well-written and extremely clear, but I believe that the discussion could be lightly revised to provide a more fulsome interpretation of some of the results and perhaps a more practical take-home message. I enjoyed reading the draft and would be happy to review a revision.

We completely revised the Discussion. We synthesize the main results in the 1st paragraph of the Discussion, with the main implication as the last sentence. We then better explain the implication and interpretation in each of the following paragraph. Finally, we provide an example of how the generalities we found in trait-matching (by being used to predict potential pool of interactions in new ecosystems or future communities without current analogs) and how food web and pairwise predictions could be combined to refine future predictive models.

One major comment is that, while the main-text material is generally clearly explained, there seems to be no such explanation/interpretation of the SI material. Even brief explanations of the summary statistics used would greatly help readers who do not specialise in predicted models.

We added information to many of the appendices that will help interpret the results that are presented. More specifically, we added descriptions and interpretations of the convergence diagnostics at the beginning of Appendix S6, added a description for all performance metrics in Appendix S7, and interpretation of all plots in the associated caption Appendix S8. For the two appendices we added (Appendix S2 and Appendix S5), we also provided interpretation for the results presented.

Before I address more specific points in the manuscript, I would like to offer a potential alternative cause for the differences between model and predicted networks that the authors report. As the authors remark in their introduction, detecting all of the interactions in a system is extremely difficult. Even with expert knowledge to supplement field observations, I (and others) find it rather unlikely that all of the observed absences in an empirical food web are truly interactions that do not occur. A 100% accurate predicted food web would fill in some of these false absences, giving a more connected (probably less modular, etc.) network. In a perfect world, researchers would take these predicted but not observed interactions as the starting point for targeted sampling that could confirm or refute the prediction model - though due to the large amount of effort required both to build a high-quality model and to conduct high-quality sampling (and the different skillsets and interests that are often involved) I am not aware of anyone having attempted this yet. I certainly do not suggest that the authors of the current article should ground-truth predicted interactions all over the globe, but I do think raising the possibility that some of the 'extra' predicted interactions truly occur would enrich the discussion. I also think it would be a valuable addition to provide a list of the most-likely predicted (but not observed) interactions in each web. As these are vertebrate webs, the species involved are much more likely to be further studied in the field than other taxonomic groups and future work may well provide evidence for or against the accuracy of these predicted links.

Thanks for the suggestion. Some of the ideas the reviewer are discussed in [Cirtwill et al. 2019](https://www.google.com/url?q=https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.13180&sa=D&source=docs&ust=1691154753270981&usg=AOvVaw3_ZPZgXWim0IzFGrOyFH7d). Indeed, even food web built from expert knowledge are not expected to be 100% complete, and the false positives (predicted interactions not documented) could be some interaction missed by experts, but not all. We added this nuance in the lines 412-416:

“Most food webs, even expert-based, are missing some interactions (e.g., cryptic and opportunistic interactions), so the additional interactions predicted by our models could be candidate for further investigation. However, our models overestimated the number of interactions by twofold even for the Serengeti food web which probably already overestimates interactions since species are grouped into trophic groups.”

Detailed comments

Figures: In general, the grammar of the figure legends is not up to the same high standard as the rest of the paper. There are several missing 'a', 'the', etc.

We revised all the figure captions.

Fig. 2: Does the total effect of geographic distance include the direct effects of phylogenetic and environmental distances? If so, it would be nice to state that explicitly for (other) graphically-challenged readers.

Yes, it includes the effect of phylogenetic and environmental distances. We added this information in the figure caption.

Fig. 4: The short horizontal lines mentioned are nearly invisible on a standard-size printout. This is interesting information to include, but the design needs some adjustment.

We made the lines thicker for them to be easier to read.

Introduction

Line 51-53: This is a slightly odd first step since food webs are generally defined as predator-prey networks, with some types of predation (parasitism etc.) being rare special cases. Maybe rephrase?

We completely cut this paragraph. We used the space to clarify the overarching hypotheses of the paper and explain our specific expectations.

Lines 61-64: Are there any limits to these existing models you can mention? As written, this paragraph makes it seem like predicting interactions is a solved problem.

We don’t really know the exact limits of these models, and this is what we address in the paper. We rewrote these sections to clarify the limitations:

“It is, however, still unclear how well predictive models of pairwise trophic interactions can transfer knowledge across different regions. If trait-matching rules determining interactions are general, then i) we should observe the same traits-interaction relationships across regions and ii) a model trained in a location should be transferable to another one.” (L60-63)

“It is also unclear how well trophic interaction models can predict the underlying structure of entire food webs emerging from pairwise trophic interactions. Food web structure encompasses system-level properties such as connectivity or number of trophic levels, as well as species traits such as number of prey or their position within the food web. These higher-level properties emerge from the joint effect of trait-matching between interacting partners, the trait distribution of species composing the food web (Gravel et al., 2016), and are influenced by the spatial scale of the network (Galiana et al., 2018).” (L74-80)

Methods

Lines 106-107: These literature review + expert knowledge networks seem especially likely to be missing interactions involving rare, cryptic, and/or poorly-studied species. It is hard for even experts to know where all of their gaps in knowledge are until field studies reveal 'unexpected' feeding relationships. Predation on fledgling birds by ungulates and seed dispersal (+ fruit consumption) by carnivores come to mind as examples of interactions that have been documented in the field but might not be obvious to include in a literature-based food web.

See response to main comments. We agree that the considered food webs are missing some interactions, but we believe these are far less incomplete than food webs based on observations only. We added the following sentence:

“Since all food webs document potential interactions (i.e., the predator could feed on the prey species), and use expert knowledge to define these trophic relations, we assume that they are less sensitive to incomplete sampling than observation-based food webs (i.e., interactions and non-interactions are similarly uncertain).” (L115-118)

That said, even though we may miss cryptic interactions, the networks we use are likely more realistic because they emphasize functional interactions, the ones that really matter for energy flows and community dynamics.

Lines 115-124: It would be helpful to introduce your unusual version of trophic level here (where basal species are animals rather than autotrophs). In the current version it is too easy to interpret 'most species are basal species (e.g., herbivores)' as 'most species ate basal species (e.g., herbivores). Just adding a line emphasizing that these are animal-only food webs so that resources are e.g., herbivores would make the appearance of this caveat in the discussion a lot less surprising.

We added a description of the 1st trophic level in our food webs

“Here, the first trophic level (i.e., basal species) are defined as species not feeding on any other terrestrial vertebrates.” (L132-133)

Line 130: It would be very nice to see what % of each taxonomic group in your food webs has trait data available. The overall numbers listed are impressive but not strictly relevant. This could be a nice table in the SI, especially as a way to gauge how likely missing traits are to affect your results.

We included this information in Appendix S2, along with more details on the missForest runs and performance.

Line 147: Can you cite a paper evaluating the performance of missForest, ideally with some simple summary statistic? Together with info on whether traits are fairly comprehensive or rare (previous comment) this would really help readers to understand the trait data you have.

We added the details of the missForest performance metric in the first paragraph of Appendix S2. [Penone et al. 2014](https://doi.org/10.1111/2041-210X.12232) compared the performance of missForest to other imputation methods for ecological traits.

Line 154: Are snakes included in the squamate tree?

Snakes are included in the tree as they are part of the squamate order.

Line 155: Inconsistent tense are/were

The sentence was confusing, so we rewrote it completely with the correct verb tense.

Line 173: I don't see beta in the equation, only beta(k). In general I had difficulty relating the equation to the text description. Tk could also be described more explicitly.

We slightly modified this section. Now, all the terms in the equation are present and described in the text (L181-185).

Line 177-178: Is the predator or prey trait applied to both beta(k) and beta(k,pred)?

Yes. We added Tk in the text description, and it should be clearer that it links to the trait-based predictors that are described in the following paragraph.

Methods section broadly: I don't have a good intuition for why longevity and clutch size should affect predator-prey interactions. Do you have any (if so, that would be nice to see in the text), or was this a case of using all available traits?

Long-lived animals and low clutch size are associated with a slow pace of life and thus could be less vulnerable to predation because they favour survival over reproduction. “Faster”-lived species can sustain more predation mortality. We added some information in the description of the predictors in Appendix S3.

Line 228: How big are the spatial domains of food webs in km2? Are they similar across webs? How did you draw the boundaries, given that food web data papers are not known for precisely defining the area to which the food web applied? Are these political boundaries, based on biomes, or something else? This information could be another SI table if it's not convenient in the main text.

We added this information in the description of the food webs in Appendix S1 and specify how we defined each area.

Results

Lines 297-300: Your summary of the AUC curve in the methods was quite clear but I didn't notice an explanation for precision-recall curve et al. These other measures are a surprise here and quite hard to interpret as a result.

We added a description of the precision-recall-gain curve in the main text:

“AUPRG is independent of the true negative rate, making it useful for highly imbalanced datasets, like food webs where there are many more absences of than presences of interaction to predict (Saito & Rehmsmeier, 2015). AUPRG takes a value of 0 for random predictions, and a value of 1 for perfect predictions.” (L229-232)

We also described each performance metric in Appendix S7.

Line 312: Missing word? Perhaps 'modelled performance against'?

Corrected

Lines 321-323: This is a very interesting result, and I was hoping to see it developed further in the discussion. Why might species with intermediate specialisation be harder to predict? Perhaps they are more influenced by abundances of potential prey than trait matching? Though in that case many are likely to be truly more generalist than they appear (would eat more prey species if they were locally abundant). How does this break down across your role measures? Is degree more affected than motif roles, for example, or are intermediate species the hardest to predict in all regards?

We completely agree that this is an interesting result. We would have liked to developed further, but we are already over the word limit. Because we can only speculate on why we found this, and the implications of the biases, we decided to focus the Discussion on the three main findings we highlight in the first paragraph. We added the following section that still discuss this result:

“The prey and predators of species of intermediate specialization were also less well predicted, which suggest that their trophic niches are harder to define by functional traits and could be influenced by other factors like abundances and co-occurrences. The overall overestimation of the number of prey and predators, and the bias against species of intermediate specialization, probably propagated through the food webs, explaining why the centrality of species was more evenly distributed, and why the predicted food webs were more connected, less modular, and with higher trophic levels.” (L417-423)

Discussion

Line 379: This paragraph seems like an ideal place to raise the possibility that some of the predicted interactions corresponding to observed absences might be true, but undocumented, interactions. The true network probably does contain more interactions than we know, and a good predictive model should be able to find likely candidates for overlooked interactions.

We added the following sentence:

“Most food webs, even expert-based, are missing some interactions (e.g., cryptic and opportunistic interactions), so the additional interactions predicted by our models could be candidate for further investigation” L412-414

Line 388: This paragraph feels quite repetitive and could be condensed to more clearly focus on the idea of constant bias in predicted network properties (if that is indeed the main point).

We completely cut this paragraph to make space to explain our main findings and their implication. We put more emphasis on (1) the generalities in trait-interaction relationships within taxa and across space show the potential of trait-matching model to predict pool of interactions in new or future ecosystems and (2) the fact that we could not predict food web properties suggest that the constraints (like spatial scale) acting on the properties cannot be taken into account by pairwise interaction models.

Line 403: I found this paragraph quite vague without having read the Leung et al. paper cited. Practically speaking, how would this work? Some sort of sub-sampling to reduce a predicted network to one compatible with the niche model or some other approach that captures structural properties of observed networks?

We briefly detailed what the approach suggested by Leung et al. is. We also gave a more practical example of how predictions from structural models can be used to update posterior predictions made by Bayesian interaction models:

“For example, Leung et al. (2019) combined predicted species distributions and expected species richness to quantify species and spatial biases. They correlated biases to species traits and environmental co-variates to improve predicted species distribution. Similar methods combining predictions of interactions and networks have the potential to provide better food web predictions. For example, we could penalize posterior food web predictions resulting into improbable food web structure to update pairwise interaction probabilities.“ (L443-449)

Line 424: This discussion of potential vs. locally-realised interactions is another place where I think the possibility of missing some real interactions in the published webs is likely to contribute to the differences you observe. If it comes to that, the empirical webs also undoubtedly over-predict the interactions that occur at any particular place or time within their area of coverage.

See previous response about predicting missed interactions.

Line 438-439: This traits vs. trait matching theme didn't emerge clearly from the results but it is interesting. It would be nice to expand upon this point earlier in the discussion, especially since it is easier to manage one species' traits than the trait match between two species.

We made many revisions in the Introduction to better explain how (1) transferability of trait-based models depends on the generalities of trait-interaction relationships (L61-63), (2) theory suggest that food web vary through either variation in trait-interaction relationships or through change in trait distribution (L77-80). Along with the changes in the Discussion, it should be clear the link between the result (i.e., good transferability) to the implication to theory (trait-matching are general, therefore, food web must vary more through trait distribution).

Appendices

Fig. S5.1 and similar: Can you provide some interpretation of what these plots mean for a non-specialist? I have no idea what these are telling me.

We added a description of each type of plot in the description paragraph, with the main interpretation.

Fig. S5.2 and similar: Check scaling - some taxonomic labels are getting cut off. I also do not know how to interpret these plots - some explanation in the caption would be very helpful.

The labels and titles have been corrected. We gave some explanation in the description paragraph at the beginning of the appendix.

Fig. S5.3 and similar: Can you provide some assistance interpreting these Rhat values? What does it mean for a non-specialist?

We have now provided some interpretation for the Rhat values to assist the understanding of these results.

Figure S6.1 As mentioned in the main text, I found your description of the AUC curve clear and likely to be helpful for someone unfamiliar with this measure. Unfortunately, while I have used AUC before I don't know the other statistics described in this figure and therefore cannot interpret it without some similar description of what these measures might mean.

We have added a complete description of what the statistics and measures included in this figure mean to help the interpretation of the results.

Figure S7.2: Based on the main text, I would expect some kind of non-linear relationship here where there is a closer match for specialists and generalists than intermediate species. Did you try fitting any non-linear models? Ditto for the other plots in this section.

We only fitted linear models. The basic expectation is that the relationship between the observed and predicted values would be linear. Having non-linear models would complexify a lot the interpretation of the biases.