# EDITOR'S COMMENTS TO AUTHOR

Editor: Chapman, Daniel

Thank you for submitting to JBI. Two expert reviewers have assessed the manuscript. Both found it a potentially valuable contribution. However, they also raised some questions about the statistical analysis, which will require consideration and potentially some reanalysis. I also had some questions about the analysis, as follows.

Dear Editor,

Thank you for this feedback and assessment of our work. We responded to all questions and concerns found by you and both reviewers. You can find below all the comments and our responses directly included and wrote in blue text. To sum-up on major changes, some re-analyses were done to better address important concerns from reviewers. Accordingly, we updated the manuscript statistics and in SI too. We re-wrote part of the manuscript to better clarify some particular points and we expanded the discussion to better express limitations of our work. We finally added a figure (Figure 2) to improve the readability of our results, as advised by both reviewers. We hope all these changes will fulfill all requirements for publication in *Journal of Biogeography*.

I would like clarification of whether transformations to the response variables in your mixed models was applied (e.g. log transform to prevent residual heteroscedasticity and prevent predicted values going below 0) and whether a generalised model (GLMM) might have been more appropriate (e.g. Poisson or negative binomial for richness).

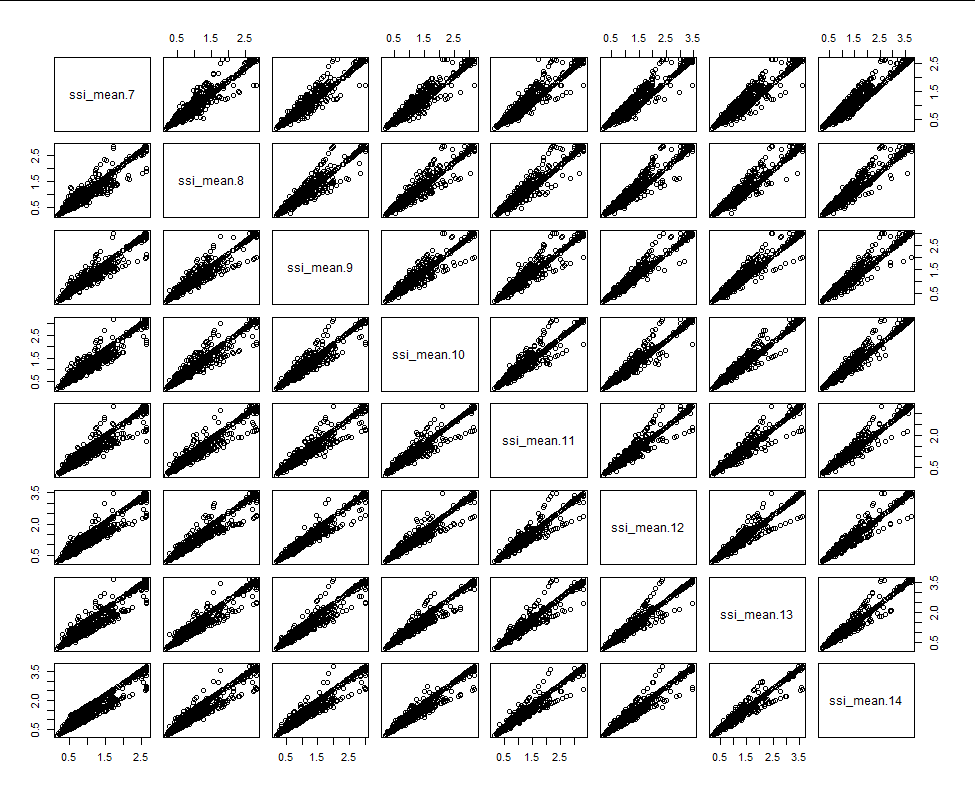
We have added language to clarify that no transformations were applied to either response variable. We also added language to clarify that we visually checked that our models met the assumptions of normality and heteroscedasticity. We opted to use models that relied on a gaussian distribution because the residuals were normally distributed and there was no evidence of heteroscedasticity. Moreover, the distribution of our response variables (i.e. species richness and niche specialization index) were normal with no left-skewed bias that are usually concomitant with the used of a Poisson or negative binomial distribution for e.g. zero-inflated models.

Did you test for interaction terms between survey type and the human development index, to test for variation in response between the habitats?

We did not originally test for this relationship. Total disturbance differs both between protocols and among wetland classes. Previous work on wetlands in the boreal region of Alberta have found that the type of development surrounding wetlands differs across wetland class (Ficken et al., 2019), likely do to practical considerations on the cost-effectiveness of constructing on different terrain types.

I also wonder whether the U-shaped relationship between human niche specialisation and human development might in some way be an artefact of the specialisation index being calculated in part from the development index, and it being capped at 0 and 100% (e.g. because species found near 0 or 100% can’t exceed those values and so appear more specialised than species found more often at 50%). Is there some way to test whether your procedure would result in this kind of relationship as an artefact (e.g. using simulations of a null model)?

In computing the species specialization index, human development bins were created that artificially imposed upper and lower bounds for each bin. That is, each bin includes a lower and upper boundary, not only the bins that include 0% and 100% human development. We needed to find a difficult trade-off between what each bin represents in the human development gradient (~10% for 10 bins) and that each bin contains the same number of sites. Because SSI values are calculated based on species frequencies in each bin, it was more important to standardized the number of sites per bin and then try to equalized the part of the gradient each bin was representing. That’s why we randomized 1000 times part of the gradient when calculating SSI values, allowing us to obtain bins that represent on average the same amount of the human development gradient (~10%).However, to determine if the number of bins that we chose (10) resulted in an artefact that influenced the observed results, we re-calculated specialization indices using 7 to 14 bins. There were very high correlations (spearman *ρ* > 0.9) among SSIs when calculated with different bin numbers (see figure below). We thus believe that the relationship between human development and niche specialization index are not an artefact of the bin width or the presence of bin boundaries.



As there are only 2 figures in the main text, you might consider presenting NMDS plots, which illustrate the clear difference between development levels (notwithstanding the reviewer’s suggestions to improve the analysis).

Following the reviewer’s suggestions, we have improved the analysis and moved the NMDS plot with two human development levels (low and high) to the main text (now Figure 2).

# REVIEWER COMMENTS TO AUTHOR

## Referee: 1

This paper summarises an interesting dataset from the ABMI, investigating patterns pertaining to human development, species richness, and community structure across wetland community types. I think the approaches taken are generally plausible, and the results are not too surprising. However, the Discussion section seems to be missing a part where the authors critically reflect on whether the combination of their dataset with their chosen modelling strategy has any weaknesses with respect to the ecological inferences drawn. At the moment the Discussion just takes the results and assumes that they are a completely correct representation of reality, with no uncertainty attached to them. This is a missed opportunity to reflect on the modelling approach taken relative to the data, and to therefore make the paper more interesting and useful to other ecologists. Some points that might be considered in such a section of the Discussion are given below in the more detailed comments against manuscript line numbers.

Dear Reviewer 1,

Thank you very much for having carefully reviewed our work. Your comments greatly improved the manuscript and we responded to all your concerns (see below). About the Discussion, we added some text throughout the section and a new subsection to discuss the limitations of our work and to make the discussion of our results broader and more objective.

### Specific comments

L94 Confusing, as introducing NNS can also increase local diversity, as you imply in the first paragraph of the introduction. It’s obviously debatable that NNS always reduce local diversity in terms of always being the proximal cause (as opposed to passengers of change responding to other factors).

You are correct. Thanks for catching this. We have updated this to say that humans “influence” rather than always “reduce” terrestrial diversity. Lines XX.

L123 This sentence should end with “in our study area” to avoid over-reaching.

We have amended the sentence as you suggest. Lines XX.

L154 I think referring to unique visits in time and space as “sites”, which normally implies a location in space only, is the opposite of clear. Why not refer to them as “samples”?

At the suggestion of Reviewer 2, we updated the analyses to include only one sampling year per site. Thus, we believe referring to them as “sites” is now warranted.

L165 What’s the lower bound on the area sample for the wetland plots?

We have updated the text to reflect the actual number of transects sampled. See lines XX.

L216 The intermediate level here seems rather arbitrary. Looking at Fig 1 there seems to be a good spread of samples between 25 and 75%, and one can well imagine that at the scale of the remote sensing data, a site with 25% human-influenced land-cover could be quite similar to one with 75% at a landscape scale – e.g. perhaps a wetland site is in a similar location to another with respect to human development, but it just happens that the location of your SPOT6 polygon includes a corner of some development in one instance, and most of it in another. At the very least it seems to me that this choice of 45-55% could do with a sensivity analysis, particularly as your choice results in a small intermediate sample, and PERMANOVA is known to be liable to confounding changes in variances with changes in locations (i.e. means). (See David Warton’s papers on mvabund as a more robust alternative to PERMANOVA in some situations).

This range was arbitrarily chosen to represent intermediate human development. We agree that this may not, in an ecological sense, represent the most accurate “intermediate” impact on wetlands. However, we believe that error in this relationship is more likely driven by in characteristics of the human development metric itself. We have discussed this limitation in the discussion. Briefly, human footprint measures only the geographic cover (i.e. area) of non-natural cover. Importantly, it does not include any information on the magnitude of impact different covers have. That is, a wetland surrounded by 50% parkland cover would have the same human development extent as a wetland surrounded by 50% parking lot cover, though these different human development types would likely have very different impacts on wetland ecology. To test whether significant PERMANOVA results might be attributed to different composition variances, rather than means, we included tests of the dispersion of each group. This has been added to the methods on lines XX and in the results on lines XX.

L228 Including sampling methodology as a fixed effect: does this really deal adequately with the different totals and types of areas sampled between the two methods? Terrestrial is 1 ha, wetland is from some unknown lower bound up to 14 plots x 20m2 = 280 m2. Normally the comparison of differently sized areas in terms of richness requires species-area curve adjustments (e.g. see Palmer et al. 2008); now, I can see that a fixed effect allows for some constant difference in richness between method types alongside adjusting for other variables in the model, and maybe this is enough, but on the other hand your wetland sites apparently have transects that vary in length, so is this fixed effect approach sufficient? Something for the Discussion.

We agree that this is an important consideration and have added a discussion of this on lines XX – YY. We opted to use observed richness as our response variable for three reasons:

1. A number of our other models require using observed richness because they are predicated on knowing which species are present. For example, computing the proportion of exotic species at site requires assigning nativity status to individual species and therefore cannot be based on estimated true richness. We therefore thought it would be confusing for readers to understand why we use to “types” of richness (observed and estimated) in the same manuscript.
2. We determined that the protocols both underestimate richness to similar degrees. To do so, we estimated “true” richness using Chao estimators, and compared the number of “missing” species in sites sampled with wetland and terrestrial protocols. On average, sites sampled with the wetland and terrestrial protocols only differed in two missing species (i.e. wetland sites miss 11 species and terrestrial sites miss 9 species). We do not believe that a difference of two species is ecologically concerning.
3. We have clarified in the methods section that the vast majority of sites sampled with the wetland protocol were sampled with the same number of transects (5 transects).

L229 Leading on from the comment against L154, it’s worth noting here that you do not even follow your own convention of calling samples “sites”.

Based on Reviewer 2 comments, we have updated the dataset to include only one sampling time per wetland, and believe that it is now clear and appropriate to refer to them as “sites”.

L230-L235 You seem to approach the question of non-linearity from the point of view of an a priori decision to compare a model with a polynomial to one without, but nowhere do you explicitly say that this was a priori based on some literature evidence or hypothesis. This always makes me suspicious that the decision was actually based on a lot of data dredging and model experimentation, and subsequent lying by omission by not stating that this was actually what happened. If it this modelling strategy was actually decided upon before looking at the data, please state this clearly, otherwise be honest and state that it was a result of experimenting with the data, and discuss what implications this has for your inferences (e.g. invalidity of claimed P-values).

Thank you for catching this. This was a major oversight in the framing of our introduction. We have updated the introduction to precise that this *a priori* hypothesis was based on previous work. We also added some clarifications to the methods section stating that a 2nd order polynomial model would be the best fit and repeated that it was based on previous work (Mayor et al 2012). We also state that visual inspection of the relationships did not warrant testing other polynomial models. See line XX in the Introduction section and lines YY in the Methods section.

L232 Leading on from this, if the approach wasn’t a priori, then why not also look at whether a cubic term improves fit in terms of AIC. Or use AIC to choose the number of knots in a spline? Even if the quadratic approach was determined a priori, looking at more flexible forms might reveal other patterns worthy of discussion in an exploratory mode. I think the AIC-based approach to spline knots is available in Frank Harrell’s “Hmisc” R package (which accompanies his Regression Modelling Strategies book). Also, is AIC appropriate, or should AICc have been used?

As it was based on an a priori hypothesis and that visual exploration of relationships did not warrant testing other polynomial models, we did not try to use AIC to choose the number of knots in a spline. About the type of AIC we used, AICc can prevent overfitting and is thus used when sample sizes are small. Since our sample size is >1500, we believe AIC is appropriate.

L283 The first sentence of the Discussion illustrates my point that you are lacking some critical discussion of your own methods and the potential influences of these on your ecological inferences. For example, you say here “we found maximum richness in wetlands surrounded by intermediate extents of human development… etc.”, and then proceed as if this has been proven beyond reasonable doubt. But even your own fitted model in Fig 1a shows that this is barely true overall, or for wetland sites in isolation, and this before even considering uncertainty in these estimates, or in model selection (in terms of whether other plausible models could have been used – note that this is not just about AIC-based model selection, which is itself dependent on certain assumptions about the data and of course on the precise dataset in hand). To be honest, I find the whole discussion rather dull, as it is just trotting out literature-based arguments, without any critical discussion about what was actually done with this dataset in this paper.

We have added a “Limitations” section to the discussion where we discuss the caveats of our findings. See lines XX. We also added text throughout the discussion section to make the discussion of our results broader and more objective (e.g. lines XX and YY). We believe that putting our results into this context has strengthened the discussion and the paper greatly.

### Section B

Given that there are only two figures in the paper, I don’t really see why so much has to be in supplementary material. I would, for example, have thought that the NMDS plots were central enough to actually be in the paper.

We added an updated version of the NMDS plot (now Figure 2) in the paper.

### Ref

Palmer, M.W., McGlinn, D.J. and Fridley, J.D., 2008. Artifacts and artifictions in biodiversity

research. Folia Geobotanica, 43(3), pp.245-257.

## Referee: 2

### Comments to the Author

This concise and very well-written manuscript describes a study using large and valuable environmental monitoring datasets to understand broad patterns in plant diversity in wetlands. The results are intuitive and interesting. However, bringing together different datasets does bring with it some issues, and I therefore have a few comments and concerns regarding how these datasets were treated in the analysis. I think that considering these issues is important for ensuring the robustness of the results.

Dear Reviewer 2,

Thank you very much for having carefully reviewed our work. Your comments greatly improved the manuscript and we agree with your principal concern. The fact we combined different datasets is indeed an important consideration. As you pointed out, we analyzed broad patterns at large scale and to this aims, we were facing some trade-offs between data availability, indices we wanted to compute and their relevance. We opted for an analyses strategy that allowed us to reveal plant diversity patterns we observed when trying to consider potential bias induced by the combination of the different datasets. We responded more precisely below to these comments and explained better our rationale in the main text.

First, there is a very large mismatch between the area covered by the inventories (10 000m2 in the terrestrial and up to 280m2 in the wetland protocols). I would be inclined to use species richness per unit area as the response variables in the models, as well as including protocol as a fixed effect because of other differences in habitat and e.g. the central area vs. transect methodologies.

Although every vegetation sampling method will underestimate true richness to some degree, we agree that the methods used to sample vegetation differs substantially between protocols. For this reason, we retained a Protocol fixed effect term in models regardless of its statistical significance.

We do believe this is an important consideration and have added discussion of this on lines XX –YY. However, we opted to use observed richness as our response variable for three main reasons:

1. Since species richness does not scale linearly with area, standardizing observed richness by the area sampled is not appropriate (see Palmer, M. W., McGlinn, D. J., & Fridley, J. D. (2008). Artifacts and Artifictions in Biodiversity Research. Folia Geobotanica, 43(3), 245‑257. doi: 10.1007/s12224-008-9012-y)
2. A number of our other models require using observed richness because they are predicated on knowing which species are present. For example, computing the proportion of exotic species at site requires assigning nativity status to individual species and therefore cannot be based on estimated true richness. We therefore though it would be confusing for readers to understand why we use two “types” of richness (observed and estimated) in the same manuscript.
3. We determined that the protocols both underestimate richness to similar degrees. To do so, we estimated “true” richness using Chao estimators, and compared the number of “missing” species in sites sampled with wetland and terrestrial protocols. On average, sites sampled with the wetland and terrestrial protocols only differed in 2 missing species (i.e. wetland sites miss 11 species and terrestrial sites miss 9 species). We do not believe that a difference of two species is ecologically concerning.

The mismatches in sampling areas probably also make Jaccard distances quite unsuitable for use in the community analyses. Instead, it would be better to use Raup-Crick or Chao distances that try to account for missing species and uneven sampling. These can be implemented in the metaMDS function.

We updated the analyses to use Raup-Crick distances. It did not change interpretation of our results.

Again relating to the central area vs. transect surveys. Coming from a landscape ecology background, I can get annoyed by sometimes lazy referee questions of ‘why choose this size buffer?’, but in this case -and coming from a landscape ecology background- a 250m buffer seems very small to evaluate the effect of human development around a wetland. This is especially true when the terrestrial wetland surveys are already 100m across (if square) – were the 250 metres measured from the centre or the edge of the plot? In the wetland wetlands, the 250m buffer was measured from the edge of the open water zone, while the vegetation transects along the moisture gradient of up to 350m (14 plots \* 25m intervals) or 412 m (including a square plot of 20m2). It therefore seems likely that the transects overlap the buffer in the wetland sites, which would have further implications for including both terrestrial (effect of surrounding human development) and wetland (sometimes including areas of human development) inventories. Is this the case? Either way, I think it should be more clearly described.

Few papers justify the scale/spatial extent of landscape attributes used in their analysis (Jackson and Fahrig 2015). However, a 250 m buffer around a wetland is well supported based on research in Alberta’s prairie parkland wetlands (Rooney et al. 2012; Kraft et al 2019). There has been less research on the optimal scale to assess landscape effects on boreal peatlands. Boreal peatlands in the study area are often large and contiguous complexes that are difficult to identify as a discrete polygon unit. Devito et al. 2005 observe that topography is ineffective at delineating hydrologic response units of peatlands on the Boreal Plains due to a sub humid climate and underlying geology of deep, highly permeable glacial deposits, which result in little surface water runoff and high precipitation uptake by vegetation. Based on this we considered it appropriate to use a 250 m buffer to quantify landscape influence surrounding each site.

That being said, we added additional language to clarify how the human development plots correspond to the vegetation plots. See lines XX.

References: Devito, *et al.* (2005) doi:10.1002/hyp.5881 ; Jackson & Fahrig. (2015) doi:10.1111/geb.12233 ; Kraft *et al.* (2019) doi: 10.1371/journal.pone.0216343 ; Rooney *et al.* (2012) doi: 10.1007/s10980-012-9784-6

Including duplicate and triplicate surveys of some of the sites seems to me to be unnecessary. Were the same plots visited again, or were they new plots/transects? Either way, having more than 1500 separate sites, it just seems unnecessary to keep the duplicates and make more complex models, rather than picking e.g. the first or most species rich survey, or the one that falls closest to the date of the explanatory variable.

We originally accounted for the repeated sampling by including site ID as a random effect. Even if having more samplings (even for the same sites) could be important to captured the best the way species are distributed along the human development gradient in order to compute relevant species niche specialization indices, we agree that a reduced sample size of >1500 wetlands is sufficiently large and will also simplify the models. Thus, we have taken your suggestion to use only one sampling event per wetland and have updated the text and statistics to reflect this. We chose to use samplings that were the closest to the median sample year (i.e. 2013) to reduce any potential variability in richness due to interannual climatic differences.

I like the idea of including the effect of non-native species on increasing local richness. I just wonder, because of results later in the paper (Fig 2a) that proportion of non-natives is related to human development, if the correlation between the two might affect their suitability for both being included as-is in the same model.

Indeed, proportion of non-natives is related to human development and it was something we expected (based on literature, i.e. by the fact that non-native species introductions are commonly related to human e.g. transport of soils, etc.). Usually, one should avoid to include variables that are correlated because they are not truly independent. Here, since nonnative species and human development are correlated, the estimates and *p*-values of each predictor will be unreliable. However, the overall fit of the model should still be reliable and is actually improved by inclusion of non-native species proportion. This allows us to make the hypothesis we discussed that non-native species could be an important overlooked factor that could also explain the observed patterns, though we are not able to disentangle which one with human development is significant or the most important based on estimates and *p*-values.

Finally, I wonder about potential spatial autocorrelation. I do not know Alberta, but it seems large enough to have quite large variation in topography, land use, land-use history, soil, bedrock and climate. Therefore I would expect 1585 wetlands across the state to have a moderately large chance of having more similar vegetation to those that are relatively nearby than those that are far away, with differences controlled by more than just very local measures of human development. I strongly recommend that geographic location is controlled for in some way in the modelling.

We tested the spatial autocorrelation of models by calculating Moran’s I and found very little evidence of autocorrelation. Supporting information also show correlograms of the residuals of best models. We omitted this result in the original submission - thank you for catching it. We have updated the methods (lines XX) and results (YY) to include this information.

### Smaller comments.

Abstract. Perhaps a short sentence or short parenthesis about what is meant in this study by niche specialisation?

We have added information on this to the abstract. LXX

Introduction: I like the introduction as it is, but if it were possible to slot in something about nonnative species and niches, it would round off the scope of the study nicely. Also, it would be good here to define fragmentation, as it every-increasingly seems to mean different things to different people (and different things to the same people, over time).

We have added a definition of “fragmentation” on lines XX. We also added some information on the niches of nonnative species on lines YYY.

101- myriad ways? I don’t think the ‘of’ is needed.

Good catch. We’ve fixed this.

Methods (or SI)

It would be nice with a figure showing wetlands with high, intermediate and low human development. I think these things can mean very different things in different regions.

We have included an NMDS ordination figure (now Figure 2) depicting the composition of wetlands in the high versus low human development group. We opted to display high vs low development wetlands (rather than high, intermediate, and low) for two reasons. First, showing the comparisons of high/low development wetlands is better aligned with our goal to differentiate wetland groups which have similar richness and niche specialization levels. Second, we believe it allows us to tell a clearer story. The range of human development %s was arbitrarily chosen to intermediate human development level. However, this intermediate range may not represent the most an intermediate ecological impact due to the characteristics of the human development metric itself. That is, human footprint measures only the geographic cover (i.e. area) of non-natural cover. Importantly, it does not include any information on the magnitude of impact different covers have. That is, a wetland surrounded by 50% parkland cover would have the same human development extent as a wetland surrounded by 50% parking lot cover, though these different human development types would likely have very different impacts on wetland ecology. We discuss this limitation on lines XX of the discussion section.

Results.

For models that don’t include polynomial fits (and because it doesn’t say that variables were scaled), it would be nice to include effect sizes in the results, with plain English explanations, for example ‘for each 10% of human development surrounding the wetland, the proportion of non-native species increased by NN%.

SI line 61: It is very difficult to see the overlap between terrestrial and wetland wetlands, because clusters of wetland sites make it impossible to see if there are any terrestrial sites there.

We updated the figure to enhance its readability. The main NMDS plot is now the Figure 2 in the main text.