*Dear Dr. Chapman,*

*Thank you for this feedback and assessment of our work. We responded to the questions and concerns found by you and both reviewers. Below, please find the comments and our responses to them written in blue. We identified three major concerns brought up by reviewers: 1. There were questions about our analytical choices (e.g. how we quantified species richness, which terms were included in statistical models, which models were used). We updated our analyses to incorporate some of these suggestions and accordingly updated the statistics in the main text and SI. These re-analyses did not change any of our original results. 2. We expanded the discussion to include a more thorough discussion of the limitations of our work. We also updated the framing of the paper to clarify that we had a priori chosen to compare linear and quadratic models. 3. As suggested by both reviewers, we included an ordination figure (Figure 2) to show the vascular plant community composition of the focal sites and to round out the story told in this manuscript. This figure was originally included only in the SI. We believe these and the other changes we made greatly improved the clarity and rigor of this manuscript. We hope all these changes will fulfill all requirements for publication in* Journal of Biogeography.

*Best,*

*Martin Jeanmougin on behalf of all the co-authors*

# 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.

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 use of a Poisson or negative binomial distribution for e.g. zero-inflated models.

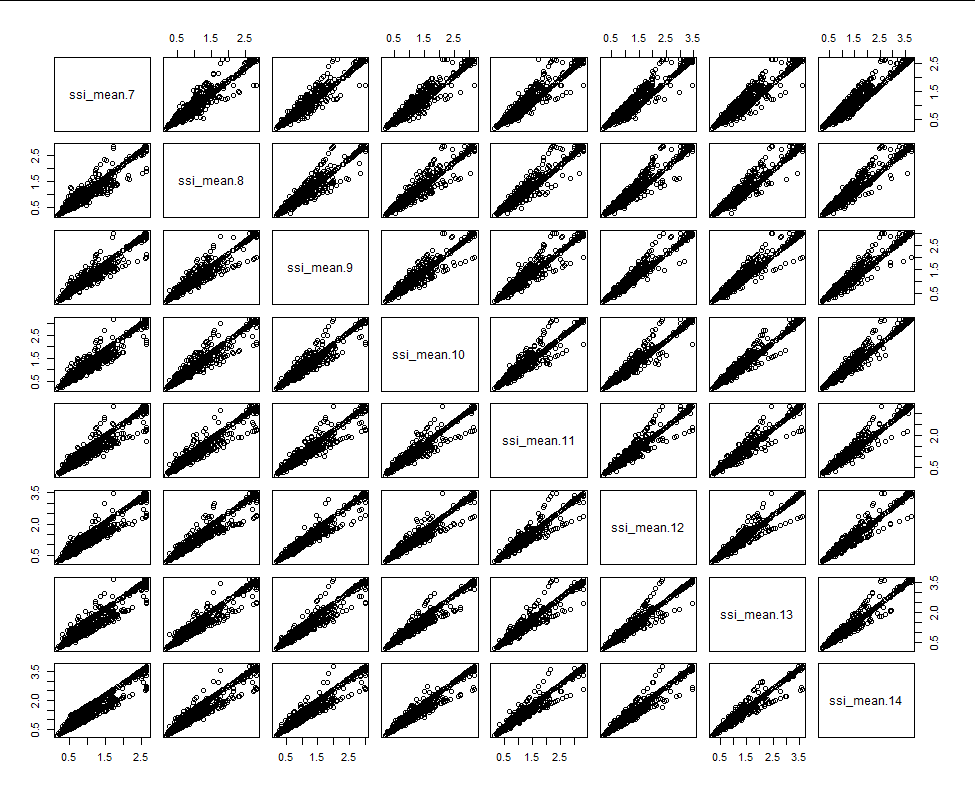
Did you test for interaction terms between survey type and the human development index, to test for variation in response between the habitats?

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.

In creating the bins, we had to balance each bin representing a similar range of ecological variability (i.e. each bin represents a range of 10% human development), and each bin containing the same number of sites. Because SSI values are calculated based on species frequencies in each bin, it was more important to prioritize standardizing the number of sites per bin, rather than equalizing 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).

We moved the NMDS plot with two human development levels (low and high) to the main text (now Figure 2). We also improved the analyses (and updated this and other figures); we respond to the analytical concerns in more detail below.

# REVIEWER COMMENTS TO AUTHOR

## Referee: 1

*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. In particular, we believed that we have improved and clarified the analysis and statistics in response to your concerns. We have also expanded the Discussion to include a discussion of the limitations of our work and to make the discussion of our results broader and more objective. Below, you will find individual responses (written in blue) to your specific comments.*

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.

### 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.

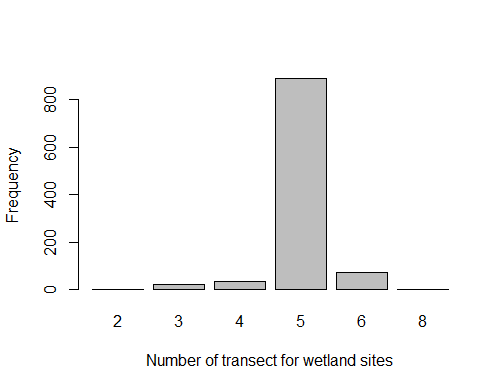
We have amended the sentence as you suggest. Lines XX.

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. Note that the vast majority of sites used 5 transects to assess vegetation (see next below).



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 to wetlands. However, we believe that error in this relationship is more likely driven by characteristics of the human development metric itself, rather than the chosen range of human development extents. We have discussed this limitation in the discussion (lines xx). 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 cover types (i.e. human development types) have. That is, a wetland surrounded by 50% parkland cover would have the same human development extent (50%) 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 the use of different methods in each protocol and the resulting potential impact on richness is an important consideration; we have have added a discussion of this on lines XX – YY.

Sampling effort (i.e. number of transects) is relatively consistent within wetland protocol sites. We clarified this in the methods section (lines XX): the vast majority of sites sampled with the wetland protocol were sampled with the same number of transects (5 transects). We therefore believe that a protocol fixed effect is sufficient for differentiating the methods used in the wetland and terrestrial protocols. Additionally, separating the protocols would prohibit calculating SSI (species specialization index; used to calculate niche specialization) values because neither protocol spanned the full gradient of human development. Together, both protocols included sites that spanned the entire human development gradient, allowing us to calculate accurate SSI values for species. We also believe that analyzing both protocols together allows us to present a clearer story than repeating the analyses and results for two protocols, when the results are qualitatively similar.

We opted to use observed richness as our response variable for two reasons:

1. A number of our other analyses 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 used two “types” of richness (observed and estimated) in the same manuscript.
2. We determined that both protocols 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.

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 clarify 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

*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 analysis 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.*

### 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.

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. 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 this was a relationship 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.). Including correlated variables in the same model will make the estimates and *p*-values of each predictor unreliable, but the overall fit of the model will still be reliable. In our case, the fit is actually improved by including the proportion of non-native species, and this fit was assessed using a metric (AIC) that accounts for over-fitting by penalizing adding additional terms.. Including nonnative species in our final models allows us to state a resulting hypothesis derived from our findings - that non-native species could also explain the observed relationships between richness and human development. Since, as mentioned above, the p-values and estimates are not necessarily reliable (since human development and nonnative species are correlated and both included in final models), though, we are not able to disentangle which variable is the most important driver of richness .

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 accidentally omitted this result in the original submission - thank you for catching it. We tested the spatial autocorrelation of models by calculating Moran’s I and found very little evidence of autocorrelation. We show correlograms of the residuals of best models in the SI. 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 groups. 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 of differentiating wetlands which have similar richness and niche specialization levels. Second, we believe it allows us to tell a clearer story. The range of human development extents used to define “intermediate” human development was arbitrarily chosen. 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 (50%) 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%.

The best model predicting nonnative species is also polynomial; the polynomial model had an AIC value 7 lower than the linear model (see lines XXX). For this reason, we did not include an effect size or “plain language” interpretation of the model. We compared the linear and polynomial models for statistical similarity with the other models in the paper and for overall methodological clarity. However, unlike for richness and niche specialization, we did not have any a priori expectation that a polynomial model would be the best fit for predicting nonnative species. We thus will defer to the editor and reviewer; if you prefer that we change this model to linear we are happy to do this.

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.