

Editor in chief

Dear Olivier Gimenez,

Two external reviewers and I have thoroughly examined your manuscript entitled “*Trade-off between deep learning for species identification and inference about predator-prey co-occurrence: Reproducible R workflow integrating models in computer vision and ecological statistics*”, submitted for publication to *Computo*.

>> Thanks. It is a pleasure to submit our work to *Computo*. The fact that articles are published in diamond open access was key to our decision to give it a try.

The reports of the two reviewers, which are complete and comprehensive, are generally very positive and suggest that your contribution should eventually be published in our journal. However, they point out a number of possible improvements and raise various questions that do not allow the work to be published as it stands.

>> We warmly thank the two reviewers for their feedbacks and constructive comments.

Another important point is that of reproducibility: the two reviewers did not manage to run your study, but point out solutions and we are quite confident that you will be able to address this issue.

>> Apologies for the inconvenience. The main issue was related to a function we use to fit species distribution models that was not available in the CRAN version of the package. This is now fixed in the script by installing the dev version. Note that we did use the dev version in the `r-addons.R` file, which explains why the code compiles smoothly on GitHub. See below our answer to the referees.

In conclusion, we would be delighted if you would propose a new version of your manuscript for resubmission. Please accompany this with a response text justifying how you have incorporated the reviewers' comments into the new version of your manuscript. Please note that in the event of final acceptance, your exchanges with the reviewers will be published with the manuscript.

>> Sorry for being so slow to get back to you, the end of the year has been a little bit hectic.

>> While revising the manuscript, we've been experiencing issues with the `fastai` package which seems to make the GitHub action fail. We've filed an issue <https://github.com/EagerAI/fastai/issues/144> and we're waiting for feedbacks from the package developer. For the time being, the Rmd notebook is available on GitHub at <https://github.com/oliviergimenez/computo-deeplearning-occupancy-lynx/blob/master/dl-occupancy-lynx-paper-revised.Rmd> as well as the HTML <https://github.com/oliviergimenez/computo-deeplearning-occupancy-lynx/blob/master/dl-occupancy-lynx-paper-revised.html>. The PDF we submitted was produced with Chrome.

Thank you for considering our journal to publish your work,

Julien Chiquet, Chief editor of *Computo*

Referee 1

The study is clearly and nicely written with an interesting perspective. The authors made a clear effort to illustrate sophisticated posterior-analysis tasks based on deep learning image classification. I perceive the main points of originality to (i) the integrated and reproducible R pipeline that should drive a wide and fast appropriation of such pipeline, (ii) the “end-to-end automation”: Except coding, there is no human labor from data collection to scientific results (not advertised too much in my opinion, if this is really new), (iii) The aim is not general monitoring, but a the question of species statistical associations, for which such methodology has rarely been used. As developed hereafter, I think the authors might push the discussion of their methodology further, and it doesn't seem much more expensive to include some other classes to avoid biases and obtain richer results.

>> Thank you for your time, and your constructive suggestions. See below our answers and details on how we addressed your comments.

General comments

- I understand sample size may be limiting, but the comparison of classified dataset with ground truth dataset would be more interesting if the classified dataset contained only machine classified images (not used in training) and no human classified images. For instance, you could split the training/test data from jura based on sites. Thus, you could compare the posterior occupancy estimates (classified vs ground truth) on jura test+Ain sites only.

>> This is a neat idea, but as the referee figured out, sample size was definitely one of the reasons why we did not consider this option. Another reason was that we were specifically interested in testing the prediction that generalizing classification algorithms trained in a site (Jura in our case study) might have moderate to poor performances when used in another environment (Ain in our case study).

- An important question is the sensitivity of the co-occurrence signal to confusions of the image classification algorithm. This is not directly addressed here, nor even discussed. First, we could expect that high confusions between prey species would artificially increase the estimated level of co-occurrence, and bias co-occurrence with Lynx. For instance, a part of the surplus in estimated $\Pr(\text{lynx present} | \text{roe deer present and chamois absent})$ and deficit in $\Pr(\text{lynx present} | \text{roe deer present and chamois absent})$ in classified dataset compared to ground truth (Figure 5) could be due to chamois being often classified as roe-deer (Figure 2). Second, observed co-occurrence patterns between focal species may be driven by co-occurrence patterns with other classes hidden among the focal species. As for example, roe deer has 0.67 precision only and include in reality many foxes. A potential competition $\text{Lynx} \leftrightarrow \text{fox}$ for prey could decrease the estimated co-occurrence roe deer / lynx. I invite you to discuss this problem, which might importantly bias your ecological interpretations precisely because of a bias in automatic identification.

>> These are excellent points indeed, and we added the following paragraph in the Discussion section to account for this comment: “When it comes to the case study, our results should be discussed with regard to the sensitivity of co-occurrence estimates to errors in automatic species classification. In particular, we expected that confusions

between the two prey species might artificially increase the estimated probability of co-occurrence with lynx. This was illustrated by $\Pr(\text{lynx present} \mid \text{roe deer present and chamois absent})$ (resp. $\Pr(\text{lynx present} \mid \text{roe deer absent and chamois present})$) being estimated higher (resp. lower) with the classified than the ground truth dataset (Figure 5). This pattern could be explained by chamois being often classified as (and confused with) roe deer (Figure 2).”. As of the second point, we agree in theory, but there is no competition between lynx and fox for big prey (despite fox preying on fawn in rare occasions). Last, we emphasize that this comment nicely relates to another comment by the other referee who encouraged us to discuss ways to explicitly account for confusions in statistical inference.

- I wonder why authors didn’t include other classes in their occupancy model. I would find very interesting to see estimate of associations between Lynx and humans (regrouped with dogs and hunters), as we might expect a negative effect of frequent human presence on Lynx due to disturbance or even due to competition for prey resource in the case of hunting. Also, what about the association between foxes and Lynx, which might compete, and between foxes and the prey? Further, mentioned earlier, it would be better to integrate these species rather than let them bias your focal estimates. The precisions and recalls in Table 3 don’t prevent the inclusion of these classes compared to Chamois or Roe deer. So what is the reason? Estimation variance reason? Computational limit? A statement about it would be nice.

>> This is indeed a fair question. First, in our experience multi-species occupancy models are very much data-hungry, and this is only by using regularization methods (Clipp et al. 2021) that we can avoid occupancy probabilities to be estimated at the boundary of the parameter space or with large uncertainty. Second, and this is true for any joint species distribution models, models quickly become very complex with many parameters to estimate when the number of species increases and co-occurrence is allowed between all species. Here ecological expertise should be used to consider only meaningful species interactions and apply parsimony when parameterizing models. We added the following paragraph in the Method section to emphasize these limitations: “First, these models are data-hungry and regularization methods [Clipp2021] are needed to avoid occupancy probabilities to be estimated at the boundary of the parameter space or with large uncertainty. Second, and this is true for any joint species distribution models, these models quickly become very complex with many parameters to be estimated when the number of species increases and co-occurrence is allowed between all species. Here, ecological expertise should be used to consider only meaningful species interactions and apply parsimony when parameterizing models.”.

>> With that in mind, we did not include humans as a species in our model because all camera traps were set up in areas with high human activity, so that human occupancy probability is basically one everywhere. We would need sites with moderate and no human activity to be able to assess the effects of human activity on lynx presence. We did include fox and cat in our occupancy analyses (see R Markdown script) but did not consider co-occurrence between these species and our focal species (lynx, chamois and roe deer). Fox do not prey on roe deer, which is why we did not include co-occurrence between these two species as a parameter to be estimated. This being said, there might be competition between lynx and fox for prey like small rodents and birds. To test for this hypothesis, we fitted another occupancy model in which we considered co-occurrence between fox and lynx. This model was better supported by the data than the model

without co-occurrence (AIC was 1544 vs 1557) which opens room for further investigation. We thank the referee for pushing us to investigate this effect, and we added the following sentence in the Discussion section: “Our results are only preliminary and we see several perspectives to our work. First, we focused our analysis on lynx and its main prey, while other species should be included to get a better understanding of the community structure. For example, both lynx and fox prey on small rodents and birds and a model including co-occurrence between these two predators showed better support by the data (AIC was 1544 when co-occurrence was included vs. 1557 when it was not).”.

Point by point:

- The work may be positionned to comparable studies with SDM, earlier work used deep-learning-based image classification to produce occurrence data exploitable for species monitoring and test sensitivity to image classification errors, e.g. Botella et al., 2018.

>> We agree, and we added this sentence “In that spirit, we praise previous work on plants which used deep learning to produce occurrence data and tested the sensitivity of species distribution models to image classification errors [[@botella2018](#)].”

- “as most, if not all, algorithms are written in the Python language”: Not all, many library have R implementations (e.g. mxnet) or R wrappers (e.g. kerasR, R keras). Mxnet-R has been used for species distribution.

>> True. We deleted “if not all” and added a link to both MXNet for R and the R interface to Keras.

- “wild board” (picturing it was funny though).

>> Well spotted, we corrected the typo.

- Figure 2: Add % across columns and line.

>> Done.

- “an ecologist would probably wonder whether ecological inference about the interactions between lynx and its prey is biased by these average performances, a question we address in the next section.” -> What you are looking at further is co-occurrence, not interaction, so stick with “co-occurrence” or “association” consistently across the manuscript.

>> Agreed, we now use co-occurrence throughout the manuscript.

- Table 2: Is it training or test/validation metrics here? You suggested earlier in the text that test/validation metrics were computed on 20% of the Jura dataset (this is what we want to see here), but in the caption you write “Model training performance”. This is not very interesting to look at training metrics as they don’t inform on the generalising

ability of your algorithm. Test/validation metrics on the Jura dataset would be more relevant and inform on the transferability gap compared to Ain dataset.

>> This table is about test/validation metrics (precision/recall as the titles of the columns suggest). We realize that table caption is confusing, and we now use “Model performance metrics” to make it clear we provide test/validation metrics here.

- “we may also infer potential interactions by calculating conditional probabilities such as for example the probability of a site being occupied by species 2 conditional of species 1” Again, I wouldn’t speak about interaction here, as this probability may also capture a same response to environmental variability, or a same response to another species, etc. Association would be more suited.

>> Agreed, we now use co-occurrence throughout the manuscript.

Reproducibility:

I tried to reproduce article using the Rmarkdown script but without full success. Most of it worked fine, but a chunk (starting 1.644 of dl-occupancy-lynx-paper.Rmd) sends an error and so half of the code couldn’t be tested. The reason is the function “optimizePenalty” is not defined, even though I have installed all dependencies. I guess it is an author’s defined function that they forget to include in the script? Please, include it and test to reproduce the article starting from an empty R environment.

>> Apologies for that. The latest version of the R package *unmarked* we used for occupancy analyses is not on CRAN yet. We added a line of code in the R Markdown script so that the dev version is now installed on the user’s computer via `devtools::install_github("rbchan/unmarked")`

Referee 2

The paper uses Deep learning for species identification to address the question of sites co-occupancy in a predator prey system illustrated with Lynx and two ungulates (Roe deer and chamois). This work highlights that the mediocre classification performances do not have huge impact on the ecological conclusion drawn from the occupancy model. It claims that even if high classification performance is the main goal in computer science, this might not be as crucial while using DL to address ecological questions. This point is interesting and the paper also provides a good tutorial for scientists who seek at exploiting camera traps data for occupancy models, a task which becomes more and more frequent.

I found the paper interesting, well written and useful for the community of data scientist involved in ecology as well as for ecologists that use camera trap data and automatic identification. Although, as mentioned by the authors, one case study does not make a proof on the small impact of missclassification in ecological answer, this is important to recall that when using DL methods in ecology, the purpose is still to address ecological questions.

I would suggest to consider this paper for publication after answering the following comments or suggestions.

>> Thank you for your time, and your constructive suggestions. See below our answers and details on how we addressed your comments.

Title

The notion of trade off mentions in the title is not clear to me. This would be relevant, if the authors explore different DL approach with different computation/implementation time and prove that several methods provide the same ecological conclusion than the one providing the best classification. Therefore, I found the title a bit misleading. I would suggest to dig around the idea that very high classification rate is not absolutely necessary to answer ecological questions.

>> The tradeoff we explore is between the time invested in deep learning vs the time invested in ecological inference. We show that in our case study, we did not need to spend too much time in deep learning trying to achieve perfect classification to get a satisfying answer to an ecological question. This is only one case study though, and we call in the Discussion section for further studies to replicate our results. For this reason, we're reluctant to claim that high classification rate is not necessary to answer ecological questions, and would prefer sticking to the current title. If there is confusion in the title, the abstract should help in clarifying the paper's objectives.

Introduction

I like the presentation of two blocking points for the diffusion of DL methods in the Ecology community. I would appreciate that the authors also question a third blocking point: the computational cost and consequently the environmental cost. This concern is probably stronger in the Ecology community than in other sciences, therefore the idea that we do not need to over train models to answer ecological questions might save hours of computational farm time.

>> That is a fair point indeed. We added a sentence about the environmental cost of training and developing deep learning algorithms in the Introduction, and also cite Strubell et al. (2019) which tackle this specific issue: “Second, ecologists may be reluctant to develop deep learning algorithms that require large amounts of computation time and consequently come with an environmental cost due to carbon emissions [@strubell2019energy].”.

Collecting images with camera traps

Could you add some informations on how similar (or on the opposite different) the two sites are ? This would be nice to have clues to understand the difficulty of the classification task.

>> Agreed. We added details on habitat composition in the text: “Animal detections were made using a set of camera traps in the Jura mountains that were deployed in the Jura and Ain counties (see Figure 1). Altitude in the Jura site ranges from 520m to 1150m, and from 400m to 950m for the Ain site. Woodland areas cover 69% of the Ain site, with deciduous forests (63%) followed by coniferous (19.5%) and mixed forest (12.5%). In the Jura site, woodland areas cover 62% of the area, with mixed forests (46.6%), deciduous forests (37.3%) and coniferous (14%). In both sites, the remaining habitat is meadows used by cattle.”. We also added some details on how the traps were set up: “The location of camera traps was chosen to maximise lynx detection. More precisely, camera traps were set up along large paths in the forest, on each side of the path at 50cm high.”. Overall, the difference in habitats we observe between the two sites might explain why transferability from Jura to Ain is not so satisfying. In the near future, we plan to re-train our algorithm using more pictures, and pooling the two sites together to improve the algorithms performances. This is mentioned as a perspective in the Discussion section: “Third, to benefit future camera trap studies of lynx in the Jura mountains, we plan to train a model again using more manually annotated images from both the Jura and the Ain study sites.”.

Deep learning for species identification

The number of classes could be specified from the beginning even so it becomes obvious on page 6. From the ecological perspective, we do not need as many classes but mostly, roe deer, chamois, lynx and other. I wonder how this would affect the classification performance. Did the author try to limit the number of classes? This could be examined or at least discussed. From the reproducible example, it seems that fewer categories were considered.

>> Even though we limited our statistical (occupancy) analyses to lynx and its two main prey, we are eventually interested in considering all species (see also our answers to comments by the other referee) to address further ecological questions, which explains why we used other species in training. Besides, the use of specific species categories instead of just a “other” category should help the algorithm to determine with better confidence when a picture does not contain lynx/chamois/roedeer in situations where there is no doubt that this is another species (think of a car for example), or where a species is detected with which lynx can be confused, e.g. fox. As of the reproducible example, the training dataset contained fewer different species than the whole training dataset, therefore we adjusted the categories accordingly. To address the referee’s comment, we added the following paragraph at the beginning of the “Deep learning for species identification” section: “Even though in the present work we quantified co-occurrence between lynx and its prey, we included other species in the training to investigate the structure and dynamics of the entire community in future work. Also, the use of specific species categories instead of just a “other” category

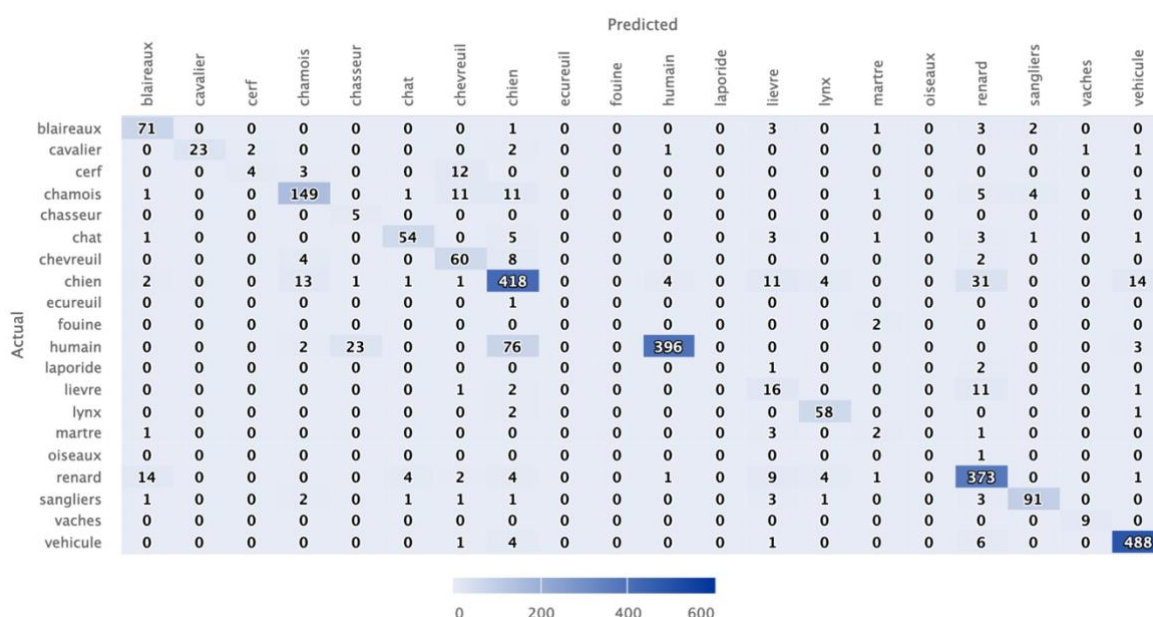
besides the focal species should help the algorithm to determine with better confidence when a picture does not contain a focal species in situations where there is no doubt that this is another species (think of a vehicle for example), or where a species is detected with which a focal species can be confused, e.g. lynx with fox.”. As of the number of classes, it can be inferred from Figure 2.

I assumed that the precision and the recall are computed on the Jury site using only the validation dataset, but this could be precised.

>> We agree, and added “Using the testing dataset” in the relevant paragraph. Also, we modified the caption of Table 2 to avoid confusion (see our answer to a comment by the other referee).

As the reproducible example has to be reduced to limit the computation time, we do not know how many individuals in every classes you have in the Jura dataset. Could you add this information ? Potentially by detailing the computation which lead to every metrics such as precision 0.9 ($9 / (9 + 1)$) so that the reader may have a better understanding on the robustness of the performance metrics.

>> Sure, the information about the number of individuals in every class is given in Table 1. The R code to calculate the various metrics is given in the R Markdown script. The confusion matrix is:



The same comments holds for the Ain study, except that the confusion matrix provide good information on the number of images to be classified in every classes.

>> Indeed, the confusion matrix in Figure 2 holds all the information we need. Note that we also added row and column percentages in response to a comment by the other referee.

Could the author provide some examples of miss classified chamois to provide some insight on the source of confusion. Why are chamois so difficult to identify ?

>> Below are examples of chamois misclassifications (sanglier = wild boar, chevreuil = roe deer, chien = dog, renard = fox, lièvre = hare, blaireau = badger). To be honest, we could not see any pattern that might explain why chamois was so difficult to identify correctly. We are open to any suggestions.



Spatial co-occurrence

Does the for the sampling occasion stand for a month ?

>> Yes, we used monthly occasions. This is clarified in the second paragraph of the “Spatial co-occurrence” section: “We formatted the data by generating *monthly* detection histories”.

I’m not familiar with the package unmarked and I would appreciate some details on the estimation methods. How is the integration over the hidden variable achieved? Are the intervals provided with the point estimates asymptotic confidence intervals of credibility intervals ?

→ from the code it seems that this is the Maximum Likelihood estimation and that the confidence interval is derived using some asymptotic normal approximation, where the standard deviation is a bootstrapped estimation.

>> Let’s consider two species for the sake of simplicity. The hidden variables here are the latent occupancy states 00, 11, 01, and 10 for “neither species present”, “both species present”, “species 1 absent and species 2 present”, and “species 1 present and species 2 absent”. To get the likelihood used in the package unmarked, we sum over all 4 possibilities, see the code at <https://github.com/rbchan/unmarked/blob/master/R/occuMulti.R>, lines 47-86. All details are given in the original publication by Rota et al. (2016), which we cite in the manuscript.

>> As correctly guessed by the referee, standard maximum likelihood estimation is used. We clarified that in the manuscript when we mention the *unmarked* R package. There is also a penalization added to the likelihood to avoid boundary estimates and deal with separation issues – see the paper Clipp et al. (2021) which we cite in the manuscript.

Are the estimated probability of detection given in the document derived from the ground truth or the classified dataset ? This could be clarified, even the authors claim taht they are both quite similar, and actually providing the corresponding results on the second dataset could help the reader to figure out how close the two point estimates are.

>> The estimates are truly indistinguishable; below are the results (point estimate ‘mle’, lower ‘lci’ and upper ‘uci’ bounds of the 95% confidence interval) for the ground truth dataset (p_lynx1, p_chamois1, p_deer1) and the classified dataset (p_lynx2, p_chamois2, p_deer2). To make this explicit, we now write “indistinguishable at the third decimal”.

```
> p_lynx1
      mle    lci    uci
lp1 0.5135232 0.450439 0.5761795
> p_lynx2
      mle    lci    uci
lp1 0.5135843 0.4504804 0.5762581
> p_chamois1
      mle    lci    uci
lp3 0.6106873 0.5467317 0.6710483
> p_chamois2
      mle    lci    uci
lp3 0.610839 0.5468376 0.6712357
> p_deer1
      mle    lci    uci
lp2 0.628033 0.5678552 0.684487
> p_deer2
      mle    lci    uci
lp2 0.6281928 0.5679774 0.6846749
```

How is it that the marginal occupancy is higher for chamois with classified dataset while so many chamois are missclassified ? This is quite counterintuitive. Could you provide some insights on this aspect ? With similar probability detection, I would expect that missing chamois identification would lead to a smaller marginal occupancy probabilities.

>> Chamois marginal occupancy seems indeed higher when estimated with the classified dataset, but this is when relying on point estimates only; when looking at 95% confidence intervals, the overlap suggests that there is no difference. Ecologically speaking, marginal occupancy may be still estimated high despite misclassification if the correctly classified data are spread over all camera traps. We added this interpretation in the manuscript: “Note that, despite chamois being often misclassified (Figure 2), its marginal occupancy tends to be higher when estimated with the classified dataset. Ecologically speaking, this might well be the case if the correctly classified detections are spread over all camera traps. The difference in marginal occupancy seems however non-significant judging by the overlap between the two confidence intervals.”.

Why is there no uncertainty associated with the occupancy probability for roe deer ?

>> There is some uncertainty, but it is very small. See the estimates in bold font below. We added some explanation in the caption of Figure 4.

```
> ground truth dataset
param          low2.5    est      up2.5
Pr(lynx present) 0.5974656 0.9305875 0.9830940
Pr(roe deer present) 0.9793906 0.9858191 0.9889957
Pr(chamois present) 0.6637510 0.8869227 0.9704254

> classified dataset
param          low2.5    est      up2.5
Pr(lynx present) 0.4423301 0.9312505 0.9937810
Pr(roe deer present) 0.9862049 0.9917414 0.9942772
Pr(chamois present) 0.7125012 0.8908234 0.9831348
```

Also the results from ground truth and classified dataset are similar, the uncertainty for the occupancy probability of lynx is quite larger. This might lead to less powerful statistical analysis. This could be discussed.

>> This is true indeed. We added a sentence at the end of the “Spatial co-occurrence” section: “Overall, we found similar or higher uncertainty in estimates obtained from the classified dataset (Figures 4 and 5). Sample size being similar for both datasets, we do not have a solid explanation for this pattern.”. At this stage however, we do not have a solid explanation for this pattern. We double checked, and the sample size of both datasets were similar, see below:

```
Ground truth dataset:
29 sites
5 species: lynx deer chamois fox cat
Maximum number of observations per site: 9
Mean number of observations per site:
lynx: 9 deer: 9 chamois: 9 fox: 9 cat: 9
Sites with at least one detection:
lynx: 25 deer: 29 chamois: 23 fox: 29 cat: 12
Tabulation of y observations:
lynx:
  0 1
142 119
deer:
  0 1
101 160
chamois:
  0 1
154 107
fox:
  0 1
86 175
cat:
```

```
0 1
222 39
```

Classified dataset:

```
29 sites
5 species: lynx deer chamois fox cat
Maximum number of observations per site: 9
Mean number of observations per site:
lynx: 9 deer: 9 chamois: 9 fox: 9 cat: 9
Sites with at least one detection:
lynx: 27 deer: 29 chamois: 26 fox: 29 cat: 13
Tabulation of y observations:
lynx:
  0 1
136 125
deer:
  0 1
97 164
chamois:
  0 1
118 143
fox:
  0 1
78 183
cat:
  0 1
220 41
```

I do enjoy the vizualisation choice made to present marginal and conditional occupancy probabilities.

>> Thanks!

Discussion

The perspective in the paper regarding the possibility to account for missclassification within the statistical analysis could be more powerful.

From the code I have found using ResNet-50, it seems that you have access to the probability of belonging to every class and not only to the hard classification. This uncertainty can therefore be accounting for, for example

- with a Monte Carlo approach, by sampling the class of every individual according to the predicted probabilities and then derived the statistical analysis (but the computational time might be prohibitive)
- with a multinomial latent variable associated to every potential sight integrated with the statistical occupancy model where the probability is given by the ResNet-50 output.

>> Thanks for these clever suggestions. We already suggest that error rates could be added to multispecies occupancy models that account for false positives, and informed by recall and precision metrics obtained during model training – this is what the referee suggests in his/her second item. We added the first suggestions of using a Monte Carlo approach: “An alternative quick and dirty approach would consist in adopting a Monte Carlo approach by sampling the species detected or non-detected in each picture according to its predicted probability of belonging to a given class, then building the corresponding dataset and fitting occupancy models to it for each sample.”. In passing, these perspectives are at the core of a research project currently submitted to CNRS which, if granted, will allow us to develop models accounting for error rates.

Finally the robustness of the result regarding the missclassification might be explained by the aggregation at the month level. A site is declared occupied by species as soon as there is at least one visit during the month. How this robustness to missclassification evolve with more refined model. The same idea holds if we do not want to assume that the occupancy probability is constant overtime.

>> We agree, and we added a sentence to emphasize that robustness might no longer hold if more complex models were fitted to the data, or if detections and non-detections were pooled weekly or daily: “Our demonstration remains however empirical, and ecological inference might no longer be robust to misclassification if detection and non-detections were pooled weekly or daily, or if more complex models, e.g. including time-varying detection probabilities and/or habitat-specific occupancy probabilities, were fitted to the data. Therefore, we encourage others to try and replicate our results.”.

Could we expect to derive more informative data from the camera trap that only a presence/absence data ? What about a monthly count data, would this be as robust to missclassification error ?

>> It probably depends on what the data are used for. If we are to estimate spatial distribution, as soon as there is a detection on a site, this site is estimated as occupied and the species present at this site. Count data are not much useful in this situation. Now if we are after population abundance, the two most popular methods are distance sampling and N-mixture which use counts of unmarked individuals (no individual identification). These models rely on assumptions, like no double-counting of the same individuals. Whether these models provide estimates robust to misclassification remains to be investigated.

Typos and Minor comments

- *To assess the relative contribution of predation and hunting, a predator-prey program was set up jointly by the French O\$ce for Biodiversity, the Federations of Hunters from the Jura, Ain and Haute-Savoie counties and the French National Centre for Scienti!c Research.*

This sentence is not clear for me. Is it the contribution on the predation and hunting of the repartition of prey ? This is not really addressed in the paper.

>> We acknowledge that our wording was confusing. We meant the relative contribution to the community structure and dynamics. We corrected the sentence

which now reads: “To assess the relative contribution of predation and hunting to the community structure and dynamics, ...”

- flipping, brightness and contrast modifications; Shorten and Khoshgoftaar (2019)) the citation does not appear clearly, may be just add “following Shorten and” in place of the “;”

>> Corrected.

Reproducibility

- To run the code, it is important to install the unmarked package from Github and not from the cran, as the function optimizePenalty, was not available on the cran version I installed from cran.

>> Apologies for that. The latest version of the R package *unmarked* we used for occupancy analyses is not on CRAN yet. We added a line of code in the R Markdown script so that the dev version is now installed on the user’s computer via `devtools::install_github("rbchan/unmarked")`.

- On Linux, there was an error on the file extension, please replace the JPG extension by jpg.

>> Done.

- While running the code, I got an error on line 1045, I was not able to identify

Error: \$ operator is invalid for atomic vectors

I have installed the latest version of fastai, so I tried the one specified in the Session Info provided by the authors, but I got the same error. So I have not been able to check the remaining of the code.

>> At line 1045, the function `ImageDataLoaders_from_folder()` is used, which is from the fastai package. To find out the issue, it would help to know whether the referee has succeeded in running any of the examples from the official webpage <https://eagerai.github.io/fastai/index.html>, and in particular that one <https://eagerai.github.io/fastai/#image-data> which uses the function that seems to cause the problem.

>> Note that while resubmitting, we’ve been experiencing issues with the fastai package that might be related to what the referee is describing. We’ve filed an issue <https://github.com/EagerAI/fastai/issues/144> and we’re waiting for feedbacks from the package developer.