

We thank the recommender and the two reviewers for their comments and inputs on our manuscript. In accordance with their feedback, we have revised our model, workflow, and paper. We implemented a small change in the way the model estimates initial population densities, which got rid of an estimation bias in first-year parameters and increased the model's efficiency in data use. We have also increased the robustness of our results and added some nuance to their discussion by implementing additional tests for sampling correlations, and calculating generation time in addition to vital rates and population densities/growth rates.

We would also like to note that when attempting to re-run our “fully reproducible” workflow on our own server after a few software and package updates, as well as a slight change in (the size of) the user base of the server, we – ironically – ran into some reproducibility issues. After several weeks of detective work, we identified poor memory and error handling as the culprits. We subsequently recruited Francesco Frassinelli – who is now listed as an author on the paper for his contribution – to take reproducibility to a new level. In addition to well-documented and semi-automated R workflows, we now also present an implementation that uses fully reproducible software environments using Nix and GNUparallel. Not only did this allow us to re-run our workflow and make it truly reproducible, but it also serves as an example for others struggling with similar issues and/or simply looking for setups to easily run their analyses on e.g. high performance computing clusters. Our repository now also contains vignettes that offer step-by-step walkthroughs for others to run and learn from our workflow.

In the following, we will respond point-by-point to specific queries raised by the recommender and the reviewers.

### **RECOMMENDER (Mark Hewison)**

This is an ambitious and rich manuscript which attempts to build a comprehensive overview of the population dynamics of willow ptarmigan across space and time using data collected through a large citizen science programme in Norway. The data are impressive, the analysis appear detailed and the conclusions will have a strong impact on both our fundamental understanding of this system and its management.

This work has been seen by two experts in the field who provide a consensus of opinion on its merit and weaknesses. While both reviewers appreciated the amount of work involved in accumulating and analyzing the data set, they also highlighted some issues that need to be addressed.

Indeed, from my own reading, I would like to emphasize the issue of integrating data from known fate animals. While this considerably enriches the information available for analysis, there is an obvious question of representativity in extrapolating these data from a single location to the entire spatial extent of the study. Given the strong spatial variation in environmental conditions across the wide latitudinal gradient, we need to be reassured that this is not a problem for model estimates and the authors' interpretations.

>> Our model does not directly extrapolate telemetry data from a single location to the entire spatial extent of the study. Rather, it assumes that there is an underlying distribution of area-specific survival values, one of which is partially informed by telemetry data. This

means that information does indeed flow from the telemetry data via survival probability in the Lierne area to the hyperparameters that define the distribution of survival values across all areas, but the amount of information contributed that way is very small compared to amount of information provided by the line transect data. We have now visualized this by comparing posterior distributions obtained from the full integrated model to a second model that does not use telemetry data but is otherwise equivalent (folder “Comparison\_noTelemetry” in our OSF repository). This shows that the use of telemetry data is crucial for estimating the seasonal decomposition of survival within the Lierne area (where telemetry data are collected), but has negligible effects on other parameter estimates. We now also refer to this comparison and its results in the Methods and Results sections of the manuscript.

Concerning the impact of rodent abundance on ptarmigan recruitment, would it be possible to modify Fig. 5C, for example, explicitly plotting the X-Y relationship with all data points visible, to get a better handle on the effect size?

>> We have now visualized the actual observational data. We have not done this in Fig. 5C as the result would have been quite messy, but instead added the observational data for each area in the supplementary figure “Rep\_betaR.R. This now illustrates well both the effect size and some of the potential weaknesses of the rodent data (e.g. zero inflation in some areas).

The results of the models appear to reveal that birds survived best in areas where recruitment was lowest which the authors interpret as evidence for an expected trade-off between survival and recruitment. I found this result striking and rather surprising, as at the population level, one would expect environmental conditions to drive positive covariation between vital rates. The interpretation of this result requires some attention, as life history trade-offs are expected at the individual level, and even then, are often confounded by individual variation in quality. I need convincing that these results can be seen as evidence for a life history trade-off.

>> This is a species for which quite different processes determine survival and recruitment, not least because it is a hunted species and hunting limits (i.e. mortality risk) are influenced by observed recruitment/population density, at least in some areas. Many hunted species also exhibit some form of compensatory recruitment (or immigration), which again is likely to result in negative survival-recruitment relationships. Additionally, we know that, in a historical perspective, there are now less ptarmigan than there used to be. It would appear likely that areas in which ptarmigans historically had both poor survival and recruitment may no longer support populations today (and hence not be included in these analyses). Based on these considerations, we do not find the lack of positive co-variation of survival and recruitment all that surprising. What is harder to determine is to what degree apparent negative relationships between survival and recruitment indeed reflect biological relationships (life-history trade-offs, compensatory mechanisms, etc.) as opposed to MCMC sampling correlation during the estimation process. In the revised manuscript, we have placed more emphasis on this distinction and adjusted our conclusions linked to life history trade-offs based on new results on sampling correlations and derived generation times.

Many thanks for this interesting contribution.

Mark Hewison

## REVIEWER 1 (Anonymous)

In this study, the authors developed a multi-area integrated distance sampling model on willow ptarmigan in Norway. This modelling approach combines line transect data (citizen data) and radio-telemetry data to estimate spatio-temporal variation in vital rates (survival, recruitment), densities, population growth rates and the effect of small rodent occupancy on recruitment rates at several sites (41) across Norway.

I have enjoyed reading this manuscript. The paper is well-written, the citizen data collected at 41 areas across Norway is impressive, the modelling approach including multiple areas is novel, and the authors provide workflows that can be set up in a reproducible way by wildlife managers. I have really appreciated the figures (maps) showing the medians AND the uncertainty of demographic parameters at the different locations. Uncertainty is not always shown, and I think plotting maps with both median and uncertainty is very nice.

Here are some few comments and suggestions. I hope the authors will find them useful during the revision process.

### Main comments:

(1) One of my main comments is about the use of radio-telemetry data that comes from one single area (Lierne municipality), whereas all line transect data come from 41 areas across Norway. This means that the demographic rates estimated within the multi-area integrated distance sampling model result from radio-telemetry data from one area + distance-sampling data from 41 areas.

o One key point of the paper is to explore and demonstrate spatio-temporal variation in demographic rates (including survival). But the telemetry data comes from one site only. Naively, one can think that if there are multiple data sources for one site (at Lierne with telemetry and distance sampling), whereas for other sites there are only distance-sampling data (with telemetry data coming from Lierne), most of the information on survival should be driven by telemetry data collected at Lierne. In other words, can telemetry data “constraint” or “drive” estimates of demographic parameters in one way or another? I guess so, as it is the philosophy of integrated population model: extract demographic information shared among different data sources. If Lierne is very different from other sites (ecological context, type of birds monitored, see my comment below), with demographic rates and their temporal variation differing a lot compared to other sites, how could telemetry data that is collected at Lierne impact (or not) estimates of demographic rates for all areas? I am not sure how this could be addressed. Of course, the best way would be to use telemetry data in all areas. In absence of data, maybe simulations could be useful to assess the sensitivity of demographic rates in all areas to telemetry data collected at Lierne.

>> With this being an integrated model, telemetry data can indeed “constrain” other parameters in the model. However, the degree to which this is happening in the current setup (with telemetry data available for only one out of 41 areas) is very small. The reasons for this are that survival and recruitment in our model are indeed identifiable without telemetry data, and that the model does not directly extrapolate telemetry data from a

single location to the entire spatial extent of the study. Rather, it assumes that there is an underlying distribution of area-specific survival values, one of which is partially informed by telemetry data. While information does indeed flow from the telemetry data via survival probability in the Lierne area to the hyperparameters that define the distribution of survival values across all areas, the amount of information contributed that way is very small compared to amount of information provided by the line transect data. Even if survival in Lierne happened to be completely different from all other areas, the little bit of information from telemetry data would be “washed out” by the bulk of the line transect data. We have now visualized this by comparing posterior distributions obtained from the full integrated model to a second model that does not use telemetry data but is otherwise equivalent (folder “Comparison\_noTelemetry” in our OSF repository). This shows that the use of telemetry data is crucial for estimating the seasonal decomposition of survival within the Lierne area (where telemetry data are collected) but has negligible effects on other parameter estimates. We now also refer to this comparison and its results in the Methods and Results sections of the manuscript.

o In the plots showing temporal variation in vital rates and detection probabilities, the opaque blue area marks the period of time for which line transect surveys have been conducted in the area. This means that for the rest of the period, in all areas except Lierne, estimates are only driven by shared variation among areas (random factors) and CMR resulting from telemetry in one site, is that correct? This is directly linked to the previous comment. I think it would strengthen the manuscript a lot if the authors could demonstrate how sensitive the results are to different telemetry data in Lierne.

>> It is correct that parameter estimates for time periods in which no line transect data was collected are extrapolated based on shared variation among areas and years and area-specific averages estimated from the available years of line transect data. The telemetry data itself also only spans a subset of years (2015-2020) and information contained in it makes only a very small contribution to extrapolation as a whole, for the reasons we have stated above. By comparing our model results to those from a model without telemetry data (see above) we believe we now support well that the results are not sensitive to the telemetry data from Lierne.

(2) In Lierne municipality, birds are equipped with VHF collars. I see that references to previous works are provided (Israelsen et al. 2020, Arnekleiv et al. 2022), but still, it could be helpful to provide a little bit of details about the study area (habitat type, hunting pressure, environmental conditions). I think this information is important to better understand to what extent the Lierne area can be representative of other areas (see also my comment above). Similarly, it could be helpful to have information on the birds equipped (sex, age). I guess individuals of different sexes/ages may have different vital rates, that can in turn influence demographic rates estimated for other areas?

While there is large variation in both habitat and hunting pressure across the willow ptarmigan distribution in Norway, the study area in Lierne is by no means an “extreme” case. Rather, it is assumed to be quite representative of much public land in Central Norway. The ptarmigan habitat in the very southern and northern parts of the distribution is somewhat different, but the harvest management appears to be more similar across this gradient. We now briefly mention this in the description of the telemetry data. We have also clarified in the methods that the telemetry data included birds of both sexes, as well as young-of-the-

year (8-9 months) and older birds. Previous studies did find little differences in survival for males vs. females overall (Israelsen et al. 2020). The same study also found slightly lower survival in birds <1 year old compared to older birds; however, as birds only got fitted with transmitters at 8-9 months old, age information in the telemetry data is only partially useful for estimating the difference of first year vs. adult survival. Age-dependent survival also can not be estimated from the line transect data, and we therefore worked with a model that had age-independent survival (and pooled telemetry data across ages to contribute information to that parameter). Even if local differences in age-/sex-composition made telemetry data from Lierne less representative, the effects of that would be proportional to the influence of the telemetry data. Since we now show that in the current setup, with telemetry data from only one area, that influence is very small, we are not overly concerned about this particular type of bias.

(3) If I understood correctly, there is no sex effect in the model, assuming same vital rates for both sexes. Is that realistic biologically for that species? If so, it should be easy to add few sentences to clarify this. If not, it seems that line transect sampling allows collecting information on the sex of the individuals (L. 176), so maybe this information could be used in the model?

>> As mentioned in our reply to your comment (2), previous studies have shown little difference in annual survival of males and females in general. We now mention this in the methods section. When it comes to reproduction, it is very common for bird species with life histories like ptarmigan to assume that reproduction is not limited by males, i.e. every breeding female will be mated. Sex-specific reproductive rates are therefore not relevant here, at least not with how the population model is currently specified. That said, you are indeed correct in that information on sexes is collected as part of the line transect sampling. However, determination of sex is impossible for the juveniles, and even for the adult, there is a substantial fraction of “unknown sex” individuals in our data. That means that if we were to model sex-specific densities (and vital rates), we would require some sort of mixture model. While that is not impossible, it may be most relevant for more local-scale studies; mixture models are computationally expensive, and having one at the scale we are currently operating at (41 areas) is unlikely to be feasible (unless accompanied by software development and sampler optimization that is beyond the scope of this paper).

(4) Another major comment is about the uncertainty around estimated demographic rates. Some credible intervals (CRI) are very large. More precisely, I am wondering whether the model converges correctly and whether all demographic rates are actually estimable by the model.

>> The model does converge (we now mention convergence assessment in the manuscript), but even at convergence there are parameters that retain large uncertainty for different reasons. We will clarify this through the examples you have provided.

Here are few examples:

- o By looking at the posterior distributions of survival and recruitment, for each area, each year: it seems that some parameters are hardly estimable (e.g. survival in 2021 for all areas).
- >> Large uncertainty in estimates of annual survival and recruitment is expected when estimation goes “beyond” the data, i.e. for year-area combinations where no line transect

data was collected. These estimates are the results of converged MCMC chains, just high degrees of uncertainty remain as the predictions are based on other parameters (means and random effects) and not directly constrained by observational data. The example of 2021 survival illustrates this point very well. The parameter represents survival from 2021 to 2022 and since no data beyond 2021 was included in the analyses, not a single area contributes information on the year or residual random effect for that prediction. That estimate is thus a true “forecast” and has very high uncertainty as would be expected. Since survival for 2021 is completely beyond the study period, we have decided to remove it from the manuscript figures to avoid confusion surrounding it.

o Similarly, CRI for the effect of rodent occupancy on recruitment rates are very large in some areas. Could the data points be added on the all the graphs (in addition to model prediction)? I have the feeling that there are very few data points for large rodent occupancy values.

>> This is indeed the case: across areas, there are only few data points for high rodent abundance values. Instead, as we mention in the discussion, the rodent observation data is sparse and/or suffers from 0-inflation in some areas, which is definitely not conducive to estimating covariate effects with high precision. We have followed your suggestion and added the actual data points to the supplementary figure “Rep\_betaR.R.pdf” on OSF.

o In Budal Fjellstyre or in Os Fjellstyre, the plot showing the average population density over time seems to indicate a population size equals to zero with a large CRI as well as for annual survival probability.

>> Population density / size estimates are not estimated as zero. The graphs may make it look that way because the posterior medians for population density are low relative to the uncertainty in estimates (= spread of the entire posterior distribution). The reason for the large uncertainty is that data collection in Budal, Os, and a few other areas started relatively late in the study period and any population size estimates prior to data collection are “hindcasts”. Similar to forecasts, hindcasts accumulate uncertainty and that is what we are seeing here.

o Similarly, uncertainty around estimated proportions of variance explained by spatial, temporal and residual variation is very large for survival probabilities (figure 7).

>> The estimated proportion of survival variance explained by spatial factors is fairly precise, but equivalent estimates for temporal and residual variation are admittedly much more uncertain. We have ruled out strong sampling correlations between the variance parameters as the main cause of this (the posteriors are not strongly correlated). We also ran a simplified model that omitted residual variation in survival and recruitment, but this did not lead to increased precision of these estimates either. Instead, residual variation in detection probability became inflated, indicating it “absorbed” unaccounted for variation in survival and recruitment when the residual terms were removed. As the area- and year-specific survival probabilities are estimated with decent precision and chain mixing (see supplementary figure “PostDens\_tS\_tR.pdf” on OSF), it seems likely that the hierarchical model underlying the survival probabilities would need improving to increase precision (see below).



Maybe the model is overparametrized? Maybe it is worth trying some more parsimonious models (e.g. remove time-dependent detection probability in some areas where it seems constant)? Maybe also some areas could be removed, in particular when the amount of collected data is low (e.g. Fjellstyre)? Maybe some key environmental drivers (e.g. weather covariates, hunting pressure) could be added in the model to “help” estimating demographic rates?

>> Due to its dimensions (and associated computational demands) our model is poorly suited for an elaborate hunt for a “most parsimonious model”. We did, however, explore whether a simplified model that omitted residual variation in survival and recruitment would yield more precise parameter estimates and/or improved mixing. That was not the case. We absolutely agree that including key environmental drivers and/or other structuring factors (e.g. spatial correlation) in the hierarchical models for survival and recruitment is a promising way forward to – hopefully – increase precision of estimates. We do delve into this in the discussion but do consider practical exploration as outside the scope of this manuscript (we are planning follow-up work for looking into this). The main purposes of this manuscript are i) to showcase a semi-automated workflow for large-scale integrated population assessments and ii) highlight the potential of sharing data across locations. The latter is also the reason why we do not consider dropping areas with little data as relevant here; by retaining them and being upfront about the uncertainties, we believe we manage to show what may and may not be possible with data integration across areas and years.

#### **Minor suggestions:**

- L. 63-64: there can be substantial amounts of variation in population dynamics and life history within a species (among populations), as shown in Nilsen et al. 2009 JAE on roe deer. Maybe also highlight that these differences can be driven by contrasting ecological contexts (hunting pressures, weather conditions, interspecific interactions, etc.)?

>> Thank you for the suggestion. We have adopted this in the introduction now as follows:

“Such variation can arise from differences in ecological contexts (including local habitat and weather conditions, hunting pressures, and interspecific interactions, e.g. Nilsen et al. 2009; Bond et al. 2021) and needs to be accounted for when developing sustainable strategies for area use, harvest management, and species and biodiversity conservation (Williams 2002).”

- L. 171: consist of -> consists of

>> Fixed.

- L. 210: from from: delete one “from”

>> Fixed.

- L. 303: proportion variance -> proportion of variance

>> Fixed.

- I am wondering whether estimates obtained at Lierne municipality (i.e. survival, recruitment, density, etc.) with the multi-area integrated distance sampling model match to estimates obtained in an earlier study (Nilsen and Nater 2024). Maybe clarify this point?

>> The estimates from the earlier Lierne case study and the multi-area analysis are not completely equivalent. The former focused on the primary study area for telemetry data, which is just the western part of Lierne, while the latter focused on the entire Lierne municipality. Nonetheless, there is a good match between recruitment and population size estimates from the two studies despite the fact that the earlier Lierne case-study did not allow for time-variation in survival. We do see some differences in the time-series of estimated detection probabilities though, and this is likely due to a combination of the different assumptions for survival (constant vs. time-varying) and the larger area / partially different set of observers in the multi-area study.

· L. 343: Estimates for population density are provided in all areas, which is very great. I am wondering whether age-specific densities (adults/juveniles) could be provided as well. This could be interesting in order to assess whether age structure varies among areas, and over years. In addition, I guess estimates of demographic rates are not only available at the area level, but also at the transect level, right? I think providing this information could be very interesting as well.

>> We would argue that for assessing variation in age structure across years, it would be most useful to look at either age ratios or proportions of the population in specific age classes (instead of separate juvenile and adult densities). Age ratios are already visualized as our recruitment rate is defined as juveniles/adult (we now specify this in the caption of supplementary figure "TimeSeries\_rRep.pdf"). In addition, we have added a series of plots showing, for each area, how the proportion of juveniles in the population has changed over time (supplementary figure "TimeSeries\_ageStr.pdf"). Should someone nonetheless be looking for the age-specific density estimates, they will be able to find them in the deposited posterior samples. Regarding transect-level data: age-specific population densities are estimated at the transect level (and available as such from the posterior samples) but remaining parameters were estimated for specific area-year combinations under the assumption that they are the same for all transects within an area in a given year. As transects cover a limited amount of space relative to the movement capacity of ptarmigan, estimating transect-specific vital rates does not make much sense unless these are linked to potentially relevant transect-level covariates (e.g. environmental characteristics such as habitat type or habitat quality). The same applies for detection probabilities, although there transect-specific estimates even using random variation may make more sense given that different transects are sampled by different observers. However, we did not look into this in the present study.

· L. 417: "variation in detection over time was modest". From the "timeseries\_pdetect.pdf" file, I would say that variation in detection is area-specific: detection seems constant over time in some areas, and highly variable in others. Maybe the range of values could be given (the lowest value reported and the highest one), just to give an idea to the readers? I guess the detection probability depends on how much the fieldworkers are trained (as well as the dogs!) (L. 170). Is the ID of the fieldworkers and the dogs known, and if so, can it explain spatio-temporal variation in detection probability? Along the same line, can spatio-temporal variation in detection probability be explained by varying environmental conditions?

>> There is definitely variation in how strongly detection probability fluctuated over time in the different areas. We have rephrased the statement in the results section to better reflect



this: *“Variation in detection over time was modest on average but the degree of temporal changes varied by area, with some areas having nearly constant detection while others showed variation by factors larger than 1.5 (SF "TimeSeries\_pDetect.pdf").”*

Regarding potential factor that can influence detection probability (beyond the ones already mentioned in the discussion of the paper): we do have the IDs of the field workers at least for the recent years since we started using the App for data recording. Furthermore, we have just started collecting additional data on the dogs for a pilot project. We believe that the level of experience of both human and canine observers will have substantial impacts on detection probability, and since the same volunteers typically monitor the same transects for several years, it is not unlikely that this may also manifests itself as temporal and/or spatial variation. For interest, this issue was explored in some detail in a recent MSc thesis supervised by E.B. Nilsen, and here strong effects of both habitat type (mainly contrasting open vs closed habitats) and terrain ruggedness (lower detectability in more rugged terrain) was found. In addition, there were effects of both time of day (higher in the morning compared to mid day), temperature (higher with colder temperature) and number of dogs (even though only one dog can be used at a time, some people bring more dogs and can use different dogs across the day – so that the dog conducting the search is always fresh and fit). Please stay tuned for future work, as we hope to investigate this further.

· On the plots showing the detection probability over time for each site, maybe the first year with a detection of 1 can be removed?

>> We had a closer look as to why detection probability was estimates very high in the first year in some areas. Upon closer inspection, we found that the way we parameterised initial population densities of juveniles and adults could lead to overestimation of detection probability and underestimation of population size. Using tests on simulated and real data, we discovered that this estimation bias could be avoided by initializing population size differently, i.e. by having the same prior on adult density as before, but calculating initial juvenile density as adult density times year 1 recruitment rate (this is equivalent to how juvenile densities are calculated at  $t > 1$ ). We have re-run all analyses with the updated version of the model. The plots still contain detection probability in the first year, as that is no longer close to 1 for any areas with the new model.

· Difference in life histories among populations is mentioned at several places throughout the manuscript. Generation time is a relevant metric to range a population along the slow-fast continuum of life history variation. As the integrated model allows estimating demographic rates for multiple populations, I think generation times can easily be estimated for all populations, and then compared?

>> We have now added calculation of generation time to the manuscript. We calculated generation time both as the inverse of fecundity elasticity (after Brooks & Lebreton 2001) and as per-generation population growth rate (after Caswell 2000). Both approaches gave very similar results, and these are now presented both in text and in supplementary figures.

· L. 625: observed -> observer

>> Fixed.

· L. 654: may possible -> may be possible

>> Fixed.

Again, thanks for this manuscript. This is a very interesting study, and the analysis of large-scale citizen data offers exciting opportunities for wildlife management and conservation.

## REVIEWER 2 (Todd Arnold)

This is an impressive data set, rigorous and well-explained analysis, and well-written paper. Almost every concern or question for further consideration that I jotted down while reading early sections of the paper was ultimately addressed in more detail in the Discussion. I have only a few big picture issues for you to consider, if they seem relevant.

I presume many of the volunteers might be hunters who enjoy the opportunity to work their hunting dogs during the pre-season, but perhaps also identify good places to go hunting after the hunting season opens? If this is the case, and depending on whether hunting is additive or compensatory, and depending on whether the decision to go hunting is dependent on how many broods you observe during the survey, there might be some potential unintended feedback in this monitoring system. In Michigan, most young American Woodcock are banded with the help of hunters who appreciate the opportunity to train their dogs during the non-hunting season, but if these same hunters go back and hunt areas that had lots of broods, the resulting harvest and survival rates might be biased. Not sure if this is a possibility in Norway's ptarmigan surveys, but if so, perhaps you should include some Discussion of possible impacts (does the embargo you mention on line 187 related to this, and does it last until the hunting season is over?).

>> Indeed, many of our volunteers are passionate hunters and dog-handlers that enjoy the opportunities of doing the transect surveys. While we cannot rule out the possibility that hunters do use the publicly available line transect data to decide where to go hunting, we do not think this will have a major impact. Access to ptarmigan hunting on public land in Norway is starting to become a limited commodity, and most licences for the few first weeks of the season (when hunting pressure is at its highest) are purchased anyway. Later in the season, the results from the surveys might be more likely to have some effect on where hunters prefer to hunt – but then the hunting pressure is in any case much lower. Also – as you correctly point out, this is also related to the embargo. Estimated population densities are made publicly available prior to the hunting season, but only on a rather coarse spatial scale. The raw data is under embargo for 1-2 years after the surveys, and hunters will thus not have access to high resolution data from the current season.

To me, this still qualifies as an IPM if you don't include the wee little bit of telemetry data, as your stage-structured surveys alone are sufficient to estimate fecundity, survival, and abundance. However, if the telemetry data are age-structured, or even if they are adults only, it would be possible to have age-structured survival estimates in your model (e.g. equation on line 257 could be  $S_j \cdot D_j + S_a \cdot D_a$ ), although I suspect HY vs AHY survival would be only partially identifiable. In North America, meta-analyses of radiomarked waterfowl indicate that they survive less well than birds receiving only legbands, so it might be worth using the telemetry data more as a prior, with understanding that it provides a lower bound on true survival of unmarked ptarmigan.

>> In principle, that is correct. Our radio-telemetry data is a mixture of adult birds and birds < 1 year old. The latter are typically 8-9 months at marking though, and thus beyond the period when we would expect there to be the largest differences between juvenile and adult survival (i.e. the first fall after birth). Based on that, our telemetry data may not give a good picture of differences in age-specific survival even if we were able to distinguish adults and

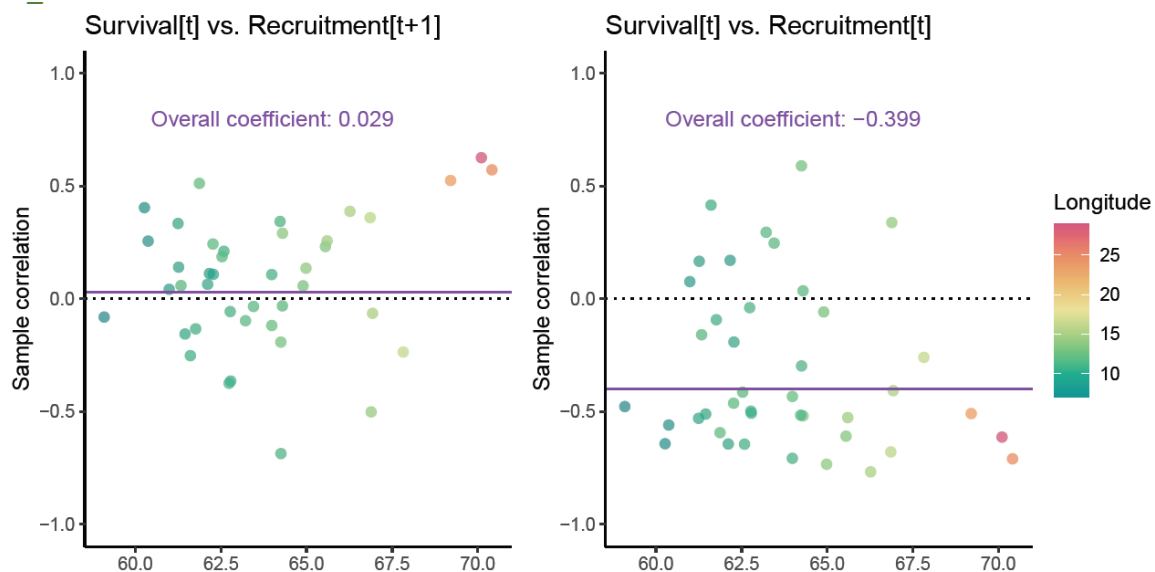
<1 year olds without fail when fitting radio transmitters. Regarding effects of radio transmitters on survival: several previous studies did not find an effect of radio-transmitters on ptarmigan survival, and we therefore do not expect this to be a major contributing factor here (see references in Israelsen et al. 2020). We have added this information to the updated description of the telemetry data.

Israelsen, M. F., Eriksen, L. F., Moa, P. F., Hagen, B. R., & Nilsen, E. B. (2020). Survival and cause-specific mortality of harvested willow ptarmigan (*Lagopus lagopus*) in central Norway. *Ecology and Evolution*, 10(20), 11144-11154.

Line 288: Is it feasible to list the key feature(s) of this custom half-normal distribution?  
 >> Feasible, yes. However, we see little reason to delve into the details here when Michael Scroggie, the nimbleDistance package developer, has written good quality vignettes documenting the background, setup, and features of the custom distributions. Instead of repeating this information in the manuscript, we have now added a reference to the package documentation. We are sure that interested readers will find all the necessary information there.

For Figure 5b, I wondered if this was indicative of a cost of reproduction, or of correlated sampling errors in your model (i.e. if you miss high on Recruitment, you will underestimate Survival, and vice versa). You hint that this might be the case later in the discussion when describing the traceplots, but a more formal way to address this might be by looking at correlations between S and R across the sims.list within a single model run. If there is an inherent sampling correlation within sites, it seems like that might also show up as a process correlation among sites. Ultimately, it might require simulations with 0 process correlation between R and S to address this, and I'm NOT recommending you do that for this manuscript, but it may be worth exploring this a bit more and adding additional caveats if warranted.

>> We have followed your suggestion and have implemented a sampling correlation test between S and R. We tested for both correlation between  $S_t$  and  $R_t$  and between  $S_t$  and  $R_{t+1}$ . The results are shown below:



What we found is that there was indeed a non-negligible amount of sampling correlation between  $S_t$  and  $R_t$  (but not  $R_{t+1}$ ). While this assessment does not allow to accurately

quantify sampling vs. process correlation, it does give us the opportunity to better discuss the patterns we see in the results. We have revised the results and discussion sections with that in mind.

One of the benefits of a CAR approach to spatial variation might be the ability to predict ptarmigan populations and population trajectories in currently unmonitored areas, especially if you included appropriate landscape covariates. Not suggesting you do this, but you might add it to the already very thoughtful “suggestions for future work”.

>> A beautiful and a bit dangerous thought. We do absolutely agree that there is a fantastic opportunity here, but one that will be subject to all constraints with extrapolation. A relevant model would probably need to also take into account habitat and possibly weather covariates, but it’s definitely worth a shot. Given our dataset (plus the additional data from private lands) we may have a unique opportunity to test and validate an approach like this in the future. We now mention this opportunity in the discussion:

“Estimation of latent parameters in missing areas may be possible though (Perry de Valpine, personal communication; Schaub & Kéry 2021 chapter 19), and this may result in a unique opportunity for making predictions of ptarmigan population trends in unmonitored areas, provided that data for a sufficient number and range of areas are available. Here, we may benefit from the fact that the line transect survey data included in this study constitutes just the publicly available part of the data collected through "Hønssefuglportalen" but the programme also includes additional surveys on private land. [...]"

#### **Some minor wording issues:**

Line 19: “structured”

>> Fixed.

Line 72: “last decade” or “last few decades”

>> Fixed.

Line 78: omit “the” or add program or scheme after Swiss Biodiversity Monitoring

>> Fixed.

Line 84: “allows for greatly reduced costs and extended”, or insert subject (conservationists, analysts) between allows and to.

>> We have reconsidered the sentence and prefer to keep the original wording.

Line 113: change final “as” to “an”

>> Fixed.

Line 138-140: As a population, willow ptarmigan are plural, but as a species it is singular. So change remain to remains, and later “ptarmigan is considered a sentinel species that is sensitive...”

>> Fixed.

Line 148: “cycles”

>> Fixed.

Line 222: “switches”

>> Fixed.

Line 223: “from” instead of for

>> Fixed.

Fig. 1 legend: Not clear that “area” refers to the 41 reporting districts unless you refer to the text – recommend adding to the figure legend too, e.g. by adding “(41 reporting districts)” to the end of the second sentence.

>> Done.

Line 284: change on to one

>> Fixed.

Line 349: omit “the” in front of northern.

>> Fixed.

Line 374: When you first provide estimates of the random effects, recommend you remind readers of the scale they are measured on, e.g. SD = 0.169 on the logit scale (log scale for Recruitment).

>> Done.

Line 551: “in areas”

>> Fixed.

Line 566: “are managed”

>> Fixed.

Line 625: “observer”

>> Fixed.

Best,

Todd W. Arnold