Review of Kissling et al., "Integrating multiple datasets to align biological and statistical populations for abundance estimation", Revision 1.

Within this manuscript, Kissling and colleagues study how spatial temporary immigration can be considered in the case of a species with complex monitoring (a highly mobile, non-territorial seabird in a dynamic habitat). Two types of populations are studied: the biological population, which takes temporary immigration into account, and the statistical population, which does not. The use of telemetry data coupled with count data makes it possible to estimate the biological population. Overall, the authors propose a model to add 'to the toolbox of demographic models that account for spatial temporary emigration'.

The manuscript is well-constructed, with an introduction in which the objectives are clearly set out, and the hypotheses are clearly stated. Similarly, results and discussion respond well to the introduction and discussion provides details of the assumptions specific to the model and the improvements that could be made.

However, I have a few concerns which I detail below:

Firstly, there is a lack of explanation of the methods used to calculate the statistical population and the biological population, even though this is the core of the study. It is clearly explained that detection is defined by several processes: probability of presence, detection, and identification. However, important information is missing, such as the content of the data used to calculate these detections: y_{pd}, y_{pp}, y_{pk} are not defined in the equations and should be specified. Moreover, once the detection parameters have been obtained, there is no information about the next stage, which consists of using these parameters to estimate the statistical population and the biological population. The equations used to obtain the estimates of these two populations are essential given the aim of this study. The code in Appendix 4 is nonetheless well-commented but is not sufficient to understand the analyses in detail.

My second comment concerns the use of the term 'integrated' model. Perhaps it stems from my limited understanding of the model, but with the elements I have here: telemetry data inform p_p, and the distance sampling data inform the statistical population. Parameter p_p is then used to obtain the biological population from the statistical population. However, an integrated model is defined with 'at least one parameter that is identical across data sets being combined' (Kery & Royle 2020). However, there don't seem to be any parameters jointly estimated by the two data sets. There is a need to clarify and justify the term 'integrated model' in the introduction.

Finally, another comment here is that the results of the study are quite focused on improving the precision of abundance estimates for the biological population versus the statistical population. This comment is also perhaps linked to my limited understanding of abundance estimations, but I wonder about the variability associated with each of the parameters and therefore the gain in precision obtained. The variability in abundance estimates comes from many sources that are not always under control. The number of birds equipped and the number of telemetry flights to locate individuals seem to introduce differences in precision, as shown in Appendix 1. It might be important to standardize the use of telemetry data? Also, there is great sensitivity to the number of boat surveys. Would it have been possible to compare the statistical population obtained between boat survey 1 and boat survey 2? Furthermore, to take all sources of uncertainty into account, shouldn't the state space model be an integral part of the 'integrated model'?

Below, I've included some line-by-line and figure/table comments.

Comment on the text:

Line 31. Specify ‘reduces the CV of the abundance estimate’. Abundance estimate *of the biological population compared with the statistical population*?

Line 35. Specify what ‘r’ is and specify *temporal?* ‘trend’.

Line 39. ‘Unbiased’ may be a strong word. Can we be sure that there is no bias in these abundance estimates?

Line 76. It's necessary to clarify what is meant by 'population mismatch' here because this term is regularly used afterward. Is it a mismatch ‘between statistical and biological’ populations?

Line 112. The word 'paradoxical' doesn't seem right here: if the area is large, it would seem logical that it would be difficult to sample sufficiently, so it's not paradoxical? Maybe just delete it or clarify.

Line 129. Can parameter p_a also be considered as a momentary unavailability (e.g., an individual underwater)?

Line 161. Be cautious when using 'unbiased'.

Line 165. Why not enter the last objective as ‘(4)’?

Lines 176-192. On Figure 1, Gull Island, Caetani River, and the glaciers are difficult to place on the map. If a reader needs to understand the ecology of the species and the complexity of field monitoring by following the geography of the bay, it probably deserves a map with more information, perhaps in an appendix if it is too difficult to include all the information in Figure 1. Or simplify the paragraph.

Line 196. How long is the interval (minimum/maximum depending on the year) between the two boat monitoring sessions?

Line 200. Do the transects shown in Figure 1 correspond to the transects carried out each year? In the introduction, it's explained that sampling cannot always be the same, but are the transects mentioned being followed as closely as possible? Please, be more specific and also refer to Table 1 for information on sampling effort. Have the environmental conditions been noted to consider the variability of detection depending on the conditions encountered (waves, etc.)? Moreover, in Table 1, what explains such a big difference in truncation between years? The whole area is sometimes 100% monitored, can we be sure that there are no double counts between two transects? Please clarify in this paragraph.

Line 211. What does ‘in and near Icy Bay’ mean? It's a term used several times, it's important to clarify it. Is Icy Bay just the Main Bay and Taan Fjord strata and 'near' the other strata?

Line 215. How long do birds keep their radio transmitters? Are the frequencies unique to each bird? This is important information for understanding how these telemetry data are used.

Line 223. Is the study area completely covered: ‘aimed’ for complete coverage or ‘assumed’ complete coverage? Are birds detected outside the strata still considered part of the biological population (detections outside the strata are visible in Figure 1)? What happens when a bird is not

detected? Does the sampling continue, or is it assumed that the bird is not present in the area? Can a bird change state during the same telemetry flight? Please clarify these aspects.

Line 270. Are these individuals not part of the biological population?

Line 298. Is it assumed that the detection conditions are the same between two surveys? Are there any environmental effects to consider?

Line 311. Clarify if p_k is considered the same between two surveys during the same year. I get the impression, with Appendix 4, that p_k is indexed by survey.

Line 312. Are there strata for p_p? This is not mentioned above.

Lines 321/330/345. y_p is not defined.

Line 328. p_p is indexed by individual (j). In appendix 4, if I understand the code correctly, p_p is equivalent to 'beta', which seems to be a constant parameter that does not depend on individual locations.

Line 339. Clarify 'i' in this equation. Line 332 'i' is mentioned as being survey 1 or 2 of the year, but line 299 mentions that p_d is the same every year. So I think 'i' doesn't correspond to the survey. Why is σ dependent on the group 'q' given that the group effect is not considered (line 303)?

Line 388. Specify λ here. Line 352, λ is the average group size, and I don't think it's referring to the group here.

Line 403. '516 telemetry locations used to estimate p_p'. It's fewer if we consider the 3-day windows length?

Line 414. It's not surprising that the estimates are close between years, given that the same telemetry data are used, or at least a common proportion. For example, in 2012, Appendix 1 shows that the exact same telemetry data were used. It would be more accurate to construct this Figure 3 by looking at how p_p varies using independent telemetry data over the same year.

Line 422. Is the tidal stage always binary? Is there never an in-between? Why is the scale linear on Figure 4 if there are just two categories?

Line 437. "suggesting that Kittlitz's murrelets in Icy Bay were functioning as a single biological population". Remove this here and keep it for discussion with more details.

Line 477. Clarify 'We improve […] abundance estimates […]' of the *biological population*? compared to what?

Line 481. Only 'if p_p varies'? If it doesn't vary, is one boat survey enough?

Line 487. Clarify 'we increased […] abundance estimates […]' of the *biological population*? compared to what?

Line 488. Clarify '*temporal*? trend'.

Line 499. Note that this is not always proportional (at least in the short term): there was an increase in the biological population between 2009 and 2010, whereas there was a decrease in the statistical population between 2009 and 2010 (Figure 7).

Line 563. ‘additional data’ such as?

Line 571. Do we have any idea how many individuals need to be equipped?

Line 577. “relax the assumption related to representativeness of the tagged animals” please explain why.

Comments on the tables/figures:

Figure 1. Explain what the grey area means. ‘Nest' appears in the figure legend but does not correspond to anything on the map. Specify the area concerned. How is the ocean state defined: why does it stop there and not before/after?

Figure 2. Why don't c and d have the same boxplot format as a and b? Specify what the blue area on plots a and b represents.

Figure 3. Write 'probability of presence' on the axes. Why choose the standard errors rather than the 95% credibility intervals as in Figure 2?

Figure 4. It might be interesting to add the linear relationships between p_p and the environmental covariates estimated by your generalized linear mixed model to this figure.

Figure 5. Write ‘probability of presence’ on the x-axis and ‘abundance of the statistical population’ on the y-axis. The ’**mean** $\boldsymbol{p_p}$’ should represent the average p_p obtained from all the telemetry data as mentioned in line 384. Perhaps don't indicate '**mean**' p_p here to avoid confusion. Why choose the standard errors rather than the 95% credibility intervals as in Figure 2?

Figure 6. Replace the legend ‘with pp / without pp’ with ‘biological population / statistical population’. It would also be consistent to apply the same color code between Figure 6 and Figure 7, with, for example, the statistical population in red and the biological population in blue.

Comments on Appendix:

Appendix 1. Note that survey 1 is in red (not black), survey 2 is in blue (not red). If the asterisks indicate that the same telemetry data were used to estimate p_p between survey 1 and 2, why, for the year 2012, was the same data used for the 1-day window length (meaning that the boats' passage took place on the same day?) but for the 5-day or 7-day window length, the data were not the same? Please clarify.

Appendix 3. ‘We did not ‘location’ any murrelets in very close pack ice’. Replace it with ‘locate’.

Kéry, M., & Royle, J. (2020). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in r and bugs: Volume 2: Dynamic and advanced models: Vols. Ch 10* (1st ed.). Elsevier. https://doi.org/ 10.1016/C2015-0-04070-9