1 Reply to Associate Editors' Comments

This paper presents a nice approach to estimating population totals from a sample of plots containing count data. The authors use change-of-support methodology to model counts in images as an IPP using spatial basis functions. In general, I think this paper is well written and do not have much to add beyond the comments of the two referees.

Thank you for the comment above, and for providing the following suggestions. We have incorporated all of them, and the changes have improved the manuscript substantially. We respond to each comment below.

Comments:

- 1. Some of the figures may be easier to read in color. One of the referees had specific suggestions regarding the figures.

  Changed as suggested.

- 2. 7, 27: Why is $K_F \geq 4K_C$? Perhaps you could say more about this choice?

  We added the following clarification "...generally $K_F \geq 4K_C$. Note that Cressie and Johannesson (2008) use 3 scales with approximately 3 times as many knots at the next finer scale. Here, because we only have two scales, we use 4 times as many knots at the finer scale." We recognize that this is *ad hoc*, and as explained later, the issue of knot selection requires more research.

- 3. 8, 20-23: Are these recommendations for the bivariate case or are these knot choices for the univariate case only?

  All of the cited literature was in a spatial context, so they are all bivariate. We clarified that by adding "spatial" when referring to knots pertaining to this literature.

- 4. 9, 5: I believe in R (using optim) you would need to use LBGF instead of Nelder-Mead if wanted to constrain the range?

  We added the following clarification "To ensure boundary conditions, say a as a lower bound and b as an upper bound for one of the elements in $\boldsymbol{\rho}$, we used a tranformation $\rho = a + (b-a)\exp(\rho^*)/(1+\exp(\rho^*))$, and then optimized for unconstrained ρ^* (note that a was a sliding lower boundary for ρ_C, but it would stabilize as ρ_F found its optimum)."

Minor Comments:

- 1. 3, 57: Should be "Royle et al. (2007)"

  Corrected.

- 2. 4, 6: "..., but we do not..." should be "..., we do not..."

  Corrected.

- 3. 5, 54: There should not be a space before the period causing it to be pushed to the subsequent line.

  Corrected.

- 4. 6, 45–47: Another potentially appropriate reference is Wikle and Berliner (2005, Technometrics).

  Added as suggested.

- 5. 9: Comma after Eq. (10).

  Corrected.

2 Reply to Reviewer #1 Report

In this manuscript the authors propose a model-based approach for estimating abundance from counts obtained during surveys of areal sample units (plots) whose locations are selected without benefit of randomization. The authors approach of using a spatial point process as a conceptual model makes perfect sense. They show that the observed counts can be derived from the assumptions of a Poisson process model, and they use Gaussian basis functions to specify spatial dependence among plots as a function of distance between plots. However, I do agree with the authors assessment that the whole issue of knot selection needs further research (p. 18). That said, this does not diminish the value of the authors current contribution, which I believe to be quite high.

The manuscript is generally well written and well organized. Below I offer several suggestions to improve an already fine paper.

Thank you for the summary above, and for providing the following suggestions. We have incorporated almost all of them, and the changes have improved the manuscript substantially. We respond to each comment below.

- 1. p. 1, line 56: D is not defined. Should this be R?

  Corrected.

- 2. 1st paragraph of Section 1.2: The authors conceptual framework is very similar to that described by Barber and Gelfand (2007) "Hierarchical spatial modeling for estimation of population size" Environmental and Ecological Statistics 14, 193–205. At a minimum, the authors should cite this article.

  Thank you. It is a good reference that we missed. It has been added.

- 3. p. 4, lines 29–31: The condition of unbiasedness is a lot to ask, and Im not sure the authors have demonstrated that their estimator is unbiased. Why not simply require that the estimator be consistent?

  That is a good point. Consistency is difficult here, and not exactly what we are after. Because of the whole finite population setting, as n increases to N, and we simply sum observed counts, any estimator would be consistent in that sense. We are after unbiasedness for fixed sample sizes over many realizations of a vaguely defined process. That will be difficult to prove in general, so we waffled and changed the wording to "The estimator should be approximately unbiased (demonstrated through simulations), and. . .". We hope that is acceptable.

- 4. p. 5, lines 20–23: Replace dx with ds to keep notation consistent.

  Corrected.

- 5. p. 5, line 48: Here is first appearance of Poi(μ(A)). Indicate that this notation is shorthand for Poisson with mean μ(A).

  Corrected.

- 6. p. 6, line 19: "to make inference on $\lambda(\mathbf{s}|\theta)$ from data" sounds awkward. Why not simply say, "to infer $\lambda(\mathbf{s}|\theta)$ from data"? Also, in the same sentence replace "aerial support" with "areal support".

  Both corrected.

- 7. p. 7, line 12: Need to indicate that $\rho > 0$ is required by the Gaussian basis function.

  Corrected.

- 8. p. 7, lines 21–22: Here authors promise to discuss later why they treated the parameter γ as fixed, not random. I may have missed it, but I could not locate their discussion.

  Sorry. That sentence is an orphan from an earlier version. That topic seems too complicated to get into in this paper. The sentence has been removed.

- 9. p. 7, line 31: Replace "linear" with "log-linear"

  Corrected.

- 10. p. 7, line 44: The simplification from β_0^* with offset to β_0 seems unnecessary. Instead, just change (6).

  Changed as suggested.

- 11. p. 8, lines 29–30: Technically speaking, $Y(B_i)$ and $Y(\mathbf{s}_i)$ are not identical. Only $Y(B_i)$ has a Poisson distribution.

  We meant that to be equivalent notation where $Y(\mathbf{s}_i)$ emphasized the centroid, but it seems unnecessary, so we removed $Y(\mathbf{s}_i)$ as it is not used again.

- 12. p. 9, line 51: Replace "Reimann" with "Riemann"

  Corrected.

- 13. p. 11, line 30: Replace "Overdisperion" with "Overdispersion"

  Corrected.

- 14. Section 2.8: I'm not really keen on the authors' approach of trying to account for spatial clustering by simply inflating the variance estimator. I think this is the weakest part of the paper and is not really necessary. Sure the Poisson process with spatial dependence is not entirely adequate, but wouldn't it make more sense to extend the Poisson model to account for a process that generates extra zeros (e.g., effects of poor-quality habitat)? The variance-inflation approach seems ad hoc to me.

  We agree that an extension of the Poisson model is more elegant, but would require modeling a non-stationary variance, or a zero-inflated model, and would change the whole estimation scheme that we use. It would surely add computational demands. The current iterative algorithm, coupled with the variance inflation, is very fast and can accommodate very large data sets, which is part of the appeal, and part of the trade-off.

We think that the variance inflation is not much more *ad hoc* than method-of-moments estimators in general, and of course it is in common use in the quasi-Poisson models. In fact, it is the application of the variance inflation locally (in the higher abundance areas) that is one of the main innovations of the paper over a simple quasi-type model. It is necessary to obtain (nearly) correct confidence intervals for the simulations. The lack of proper coverage (really, very poor) of the simple quasi-type inflation is what initially drove us to this solution.

- 15. p. 14, line 16: The confidence interval for abundance T_t appears to be based on an assumption of normality (but shouldnt 1.654 be 1.645?); however, there is no theoretical basis for this assumption since $\hat{T}(\mathcal{U})$ is a nonlinear function of the models parameters. If anything, I would have thought that the authors would have used a lognormal distribution and formed the CI by computing lognormal parameters that imply a lognormal mean of $\hat{T}(\mathcal{U})$ and a lognormal variance $\hat{v}_{k,t}$. If the value of $\hat{T}(\mathcal{B})$ is small, using a normal distribution could potentially yield a negative value for the lower confidence limit of abundance, which is obviously undesirable.

Thank you very much. In looking at eq. (10), we just assumed that sum would allow the central limit theorem to work for us. However, there is very high correlation. After examination, the estimator is skewed, even for large $\hat{T}(A)$. We adopted your suggestion, and added a paragraph on confidence intervals and new equation (18). We re-computed all of the tables for the simulations as well, but the overall patterns and recommendations are unchanged. We also fixed 1.645.

- 16. p. 17, line 27: Replace "Figure 1" with "Figure 2"

Corrected.

- 17. Figure 2: It's difficult to distinguish the spatial pattern with grey scale. I recommend using a color figure.

Changed as suggested.

- 18. Figure 7: Scale uses dark grey for low abundance and light grey for high abundance, which is exactly opposite that of Figure 2! I think these two figures should use the same ordering of color scale.

Changed as suggested.

3   Reply to Reviewer #2 Report

This manuscript presents a model-based methodology for estimating population sizes from an irregularly-sampled set of plots. The authors address the issue of how to introduce a finite population correction in a model-based context and deal with overdispersion. The work is well-motivated by a real-world application in counting seals in Alaska.

I found this to be a well-written, clearly-motivated, practical, careful piece of work. I have a few substantive questions that the authors should be able to readily address.

Thank you for the comment above, and for providing the following suggestions. We have incorporated almost all of them, and the changes have improved the manuscript substantially. We respond to each comment below.

Major Comments:

- 1. I'm unclear on the possibility of a selection bias here. The authors say that the camera is turned off where seals are necessarily absent, but I think they then go on to estimate abundance in those locations from the model rather than treating them as known zeros. It seems to me there is a selection bias possibility that the model may estimate non-zero abundance based on smoothing from nearby areas that have seals even though the counts are known to be zero (perhaps just very close to zero?) in those areas where the camera was turned off. Provided there are data near to the "camera-off" areas that have low abundance, any bias is probably limited, because low abundance will be inferred. But if there were an area of high abundance next to an area where the camera was off, there would be an overestimate for the unsampled area from the smoothing. It would be helpful if the authors can comment on this issue. Perhaps Im misunderstanding something?

  Sorry that we did not make that clearer. All areas where a seal could not haul out were eliminated by the boundary file. We tried to explain that near the end of section 1.1 by including discussion of 'donut holes.' However, to make it clearer, we added a sentence to address your question. That is, we make every attempt to eliminate the selection bias problem.

- 2. The motivating goals listed in Section 1.3 and the interest in the actual number of seals rather than the mean intensity, as well as the interest in frequentist performance, are all nicely laid out, and I thank the authors for their clear presentation of the methodological motivation.

Thank you!

- 3. The authors use maximum likelihood for estimating basis function coefficients. In many contexts this would be done with some penalization (penalized splines, Bayesian estimation etc.) and doing unpenalized maximization carries a risk of having very high uncertainty. In this context, the authors use a limited number of knots and their estimation does account for uncertainty, so perhaps this is not a major concern, but a sentence or two recognizing this as a potential issue would be warranted. I guess the main concern comes as the number of coefficients gets large relative to the number of observations, in which case asymptotic justification for their variance estimator on page 10 breaks down.

Both penalized splines and Bayesian approaches are now mentioned at the end of section 2.5, and a caution is provided regarding variances and the number of coefficients after equation (14).

- 4. On page 10, line 23, the authors decompose the prediction variance into two terms, which I interpret as (1) the variability of the true number of seals around the mean (which is of course based on the intensity function) and (2) the uncertainty from estimating the basis coefficients and therefore in estimating the intensity function. On page 11, line 20, the authors describe the two terms as (1) the variance of predicting the intensity surface given the regression parameters, and (2) the variance in estimating the regression parameters. Is my interpretation of the first term off-base? I dont understand what the authors mean by the variance of predicting the intensity surface. If the regression model is correct (which I think is an assumption that is implicitly being made here), then the intensity surface is just a deterministic function of the coefficients and there is no variability when conditioning on the regression parameters. The variability comes because of the realization of the actual counts given the intensity. On a related note, in the application, I think it would be very valuable to show the variance decomposition for the actual prediction variance. My intuition is that the second term may often be more important than the first, following my point #3 above.

We think that we have the same interpretation, but we didn't explain it well. A simple analogy is predicting a new observation from an independent sample of normally distributed data, say from $N(\mu, \sigma^2)$. Given the mean, a prediction will have the variance of the normal distribution. Then, we add the variance of estimated the mean; e.g.,

$\sigma^2 + \sigma^2/n$. We were thinking of the $\mathcal{M}(\hat{T}(A))$ decomposition in that same way. We have clarified the language, and added the values to the example as requested. You were right; about $3/4$ of the variance is due to the second term.

- 5. Page 12, top: Why is weighted least squares rather than ordinary least squares the right thing to do here? I worry about robustness to outliers even in unweighted form. Is there a mathematical justification for the weighting - the authors merely state that they want to weight values with large expectation more. More generally, I would have thought that estimation for overdispersion in Poisson models is well-developed in the literature, yet the authors treat this as an open area for investigation. Can the authors confirm that this is still an open question and give the reader a few citations to point to the current understanding of this problem? Is the linear regression estimator completely new or have others proposed this? Given the simplicity of the linear regression estimator, if this has not been proposed before, is there a reason for that?

We have added material to the introductory paragraph on the overdispersion section. We have searched thoroughly, and several times, trying to find reasonable estimators in the literature. That would be desirable rather than proposing our own, but we have not found anything. Still, it's possible that we missed some. We tried many searches using words like 'overdispersion,' 'robust,' 'nonstationary,' and many more. Our situation is somewhat unique because we want to trim or downweight residuals with low expected values, so it is not a general prescription. Part of the reason that we list several estimators is that we do not know which is best. We suspect the trimmed mean is the best, in part for the reasons that you mention, but it does require a decision on the trimming percentage. We hope to continue research in this area, and hope others will as well.

Minor Comments:

- 1. Abstract, line 29: "unsampled area" => "unsampled areas"

Corrected.

- 2. Page 1, line 50: I didn't follow why "extensions to count data have been difficult". There is lots of work on spatial GLMs where the likelihood is Poisson. Is there some

difficulty in going from maps developed based on count data to abundance estimates? Are the issues involved basically those given in Section 1.3 of the manuscript?

In part due to Section 1.3, but we were mainly thinking about nonlinear link functions and change-of-support. We have clarified this sentence and added a reference to Cressie, 1993.

- 3. Page 3, line 24: "days" => "day's"

Corrected.

- 4. Page 8, line 50: It might be helpful to the reader to explicitly call this a block-wise coordinate descent algorithm.

Changed as suggested.

- 5. Page 9, line 7: As far as I can tell, R's optim() does not provide for constraints when using Nelder-Mead, only when using BFGS (see the help information on the 'lower' and 'upper' arguments). Can the authors clarify how they imposed the constraints? Furthermore, I believe the constraints in optim() are constants and wouldnt allow for a constraint such as $\rho_C > \rho_F$.

The AE had a question about this as well. It is now clarified.

- 6. Page 9, line 51: "Reimann" => "Riemann"

Corrected.

- 7. Page 11, line 30: "Overdisperion" => "Overdispersion" !!!

Correted.

- 8. Page 18, line 46: A side note that I believe similar convergence issues happen with spatial models with Tweedie (continuous observations with zero inflation) likelihoods as well, though I don't have a good citation offhand.

Thank you. We could only find this reference with some searching. I will contact Chris upon re-submitting this. It is an interesting connection. Thank you.

- 9. Fig. 4: It would be helpful to use different line types to distinguish the two solid lines in panel A.

Changed as suggested, and we think the whole figure is improved by using open circles as well.

- 10. In Fig. 2, dark is high abundance, while in Figs. 3 and 7, dark is low. It would be good to have the color ramp be consistent in direction. I'd also suggest color as a way of allowing the reader to more easily see the variations than a greyscale, particularly in Figs. 2 and 7, but I'll leave this comment as being one of personal preference.

These have been changed as suggested, and color is now used.