Dear Editor-in-chief,

Please find attached an substantially updated version of our manuscript “Optimal sampling design for spatial capture-recapture”. We have revised the manuscript in accordance with the very thoughtful and constructive reviews. Broadly, this involved:

1. More thoroughly establishing the novelty of the work by emphasizing it as a general statistical tool for design generation that formalizes existing rules of thumb, and with the added advantage of being, demonstrably, flexible to any landscape, whereas previous research on design have been restricted to unrealistically square landscapes with regular trapping grids.
2. Addition of a third design criterion ($Q_{\overline{p_b}}$) criterion that balances captures ($Q_{\overline{p}}$) and spatial recaptures ($Q_{\overline{p_m}}$).
3. We have clarified and developed terminology throughout to address the lack of clarity identified in parts of the manuscript.

In addition to these general revisions, we have updated the manuscript throughout in response to the reviewer’s specific comments. Overall we were encouraged by the reviewers comments and would like to thank you, the handling editor, and both reviewers for the time taken to consider our work, and for giving us the opportunity to further strengthen our manuscript which we hope is now suitable for publication as a Statistical Report in *Ecology*. Below we have provided detailed point-by-point responses to the reviewer’s comments.

Many thanks,
Gates Dupont, Andy Royle, Ali Nawaz, Chris Sutherland

Author responses to editors and reviewers comments
Here we provide point-by-point responses to each of the comments in the review of our submission. Comments raised by reviewers are shown in red and our responses are indented gray text including manuscript line numbers when necessary indicating exactly where changes have been made.

Reviewer 1

Summary

I liked the manuscript and I think it can be a useful contribution to the theory of spatially explicit capture-recapture models (SECR oSCR). As a general impression, **I have struggled to see the novelty of the work**. Results **confirmed largely what was already known in Efford and Boulanger, 2019 and Sun et al. 2014** (cited in the text; see authors' comment at line 269 and 305). I consider it a small but interesting step forward within the SECR literature, but I advise authors to better stress the novelty of their work.

> Thank you. We were encouraged to see that you and reviewer 2 agreed that the work had value. We also thank you for your thoughtful comments which have helped highlight these useful contributions. We also note that as a result of your comments we have made two major updates to the package: the inclusion of other detection functions, and the development of a detailed vignette for using scrdesignGA() including defining design parameters and generating the required objects.
>
> In hindsight, we concede that we didn't do the best job of defining exactly what the purpose, and hence novelty, of this work was. To clarify: rather than develop another set of recommendations (heuristics) that are case and context specific, and largely untested in practically relevant landscapes, this work focuses more on providing a statistical framework (and open-source tool) that we believe is not constrained by heuristics, and places an emphasis on defining the objectives of the study (via a criteria; we provide two but others could be defined), knowledge of the species and system, and, importantly, resources.
>
> We believe strongly that this is an advance. By way of an example: Efford and Boulanger (2019) provide a metric for evaluation which will select the best design among a set of designs, but there is no guarantee that any of those designs are 'good' or whether better options exist. Likewise, Sun et al. offer heuristics that are tested in regular grids of possible trap locations, which we argue is difficult to achieve (especially for large

carnivores), and includes advocating for clusters, but those clusters are resource-intensive and the entire design could potentially be sensitive to the location of the first trap or cluster. We note also that even with the existence of admittedly straightforward rules of thumb, the design of spatial capture-recapture studies deviate wildly from these recommendations which is problematic, especially for difficult to monitor species that are of conservation concern. For example, Tobler et al (2013) and Suryawanshi et al (2019) demonstrate this failure to adhere to common design principles for jaguars and snow leopards respectively.

Our approach of statistically-informed automation of the design generation process is absolutely novel and addresses both of the arguably arbitrary decisions in the existing recommendations, and in doing so moves the topic of design forward substantially while ensuring minimum requirements are met as a direct output (trap spacing, spatial coverage etc...).

With that being said, we didn't make this as clear or central as we had hoped. In several places throughout the revised manuscript, we have been more explicit about what we see as the novelty and value of the work. We hope the response above, and the corresponding changes (see below) sufficiently highlight the novelty and clearly distinguish this work from existing literature

Changes to the manuscript that highlight novelty:

> Lines 62-64: *"Generally speaking, sampling design for SCR can be conceived as a problem of selecting a subset of all possible trap locations that maximizes some SCR-relevant objective function. Here we develop an analytical framework that directly addresses this challenge."*
>
> Lines 104-105: *"Herein lies one of our novel contributions: we suggest three design criteria:..."*
>
> Lines 245-247: *"In this study, we develop a conceptual and analytical framework for generating near-optimal designs for SCR studies. We suggested two intuitive and statistically-grounded design criteria that can be optimized to produce candidate designs."*

Tobler, M.W. and Powell, G.V., 2013. Estimating jaguar densities with camera traps: problems with current designs and recommendations for future studies. Biological conservation, 159, pp.109-118.

Suryawanshi, K.R., Khanyari, M., Sharma, K., Lkhagvajav, P. and Mishra, C., 2019. Sampling bias in snow leopard population estimation studies. Population Ecology, 61(3), pp.268-276.

Specific comments

Line 31 As far as I know the first SCR models were developed by M. Efford. I think it would be correct to cite his pioneer work.

We have added these citations, thank you for pointing this out to us!

Line 41 the definition of "spatial captures" is not clear. Individuals with small home-ranges are never observed in multiple traps but can be captured several times. Are these recaptures unimportant in terms of sampling design?

We agree that this definition is narrow given the context in which it is used. We have now changed this sentence to reflect the broader point that the spatial pattern of individual observations is the important third consideration relevant to study design in SCR:

"... the spatial pattern of individual encounter histories." (line 40)

Line 60. This seems a contradiction. The current paragraph informs the reader on how to optimize spacing and cluster.

Thanks for pointing out that this sentence comes across as an inconsistency to the reader. We added a brief phrase at the end of this sentence to clarify our point that there exist only a few recommendations, and that they provide only basic guidelines without being informative about how to create an 'optimal' SCR design according to some objective. The added phrase is underlined here:

"In summary, the idea of optimal sampling design for SCR remains poorly understood beyond these few, basic recommendations." (lines 57-58)

Line 63. As far as I understood, landscape matters only if it constraints or influences animal home-ranges. The focus should be on animal movements not necessary on landscape (even if they can be related).

> What we refer to in this section is spatial structure in population density that is driven by habitat structure, such that areas of more suitable habitat are able to support more individuals per unit area compared to areas with less suitable habitat. As can be seen in our methods, we simulated this effect by parameterizing the coefficient of the landscape effect as 1.2, representing a weak but identifiable and biologically-relevant association between landscape (habitat) and population density.
>
> It is the case that SCR is generally robust to not accounting for spatial variation in density **if the landscape, or more specifically, the density gradient is representitively sampled**. So, from a design perspective, we wanted to investigate whether the idea of optimality holds when populations are spatially structured (which it does). You raise a very good point about whether the notion of optimality holds when space use is not-circular as is typically assumed. That is beyond the scope of this paper, in which we present the tool and demonstrate the ideas that criteria can be generated and design optimised with respect to those criteria.

Line 65. This topic has been already addressed (see previous paragraph)

> We agree that the topic has received some attention. However, the previous paragraph describes in detail the heuristic nature of these investigations. We have clarified this by closing the previous paragraph with:
>
> > "*In summary, the idea of optimal sampling design for SCR remains poorly understood beyond these few, basic recommendations. In particular, it is unclear whether existing design heuristics hold for spatially-varying density patterns, or in highly-structured landscapes where recommended regular trapping arrays can not be accommodated, and guidance of generating clustered designs is lacking"* (lines 57-61).
>
> Moreover, we now begin the next paragraph in general terms:
>
> > *"<u>Generally speaking,</u> sampling design for SCR can be conceived ..."* (line 62).

Indeed, the work conducted on design to-date has been instrumental in our development of the criteria we use. We now acknowledge this in line (<LINE>):

> *"...two model-based criteria <u>related to current thinking about the relationship between data quality and estimator bias and precision</u>..."* (line 69-70).

Line 92. The half normal distribution has a shoulder at short distances. In my experience this is not always realistic. Would it be possible to include the formula for the negative exponential function? It would make the application more general.

This is a good point. Although we focus on the half-normal, any other detection function can be used in the optimization procedure. In fact, this comment has motivated us to add an option for selecting among common detection function models which will be available in the next package update - thank you.

Line 159 A bracket is missing.

This has been fixed, thank you for pointing this out to us.

Despite the work can benefit environmental managers to help them on choosing the number and location of detectors (see line 354), it is not clear how to use the scripts on a real case. Authors considered, or have been inspired by, a real case of an irregular and heterogeneous landscape. It will be interesting to add a paragraph to illustrate how to use the scripts. For example, "Let's consider an area of …, Appendix 8 can provide the optimal number of detectors given criteria 1…." something on these lines.

In addition to the fully-commented scripts, we have developed a detailed vignette for using scrdesignGA() including a step-by-step workflow for defining design parameters and generating the required objects. We link to this vignette at the very beginning of the scripts, and mention it in the manuscript:

> *" We provide a detailed description of the general GA, the k-of-n adaptation, and our implementation in the R package oSCR in Appendix 1 and Appendix 4."* (lines 118-120)

Reviewer 2

Summary

The paper addresses an important topic that has been given relatively little attention: study design for spatial capture-recapture surveys. As the authors highlight, advice on study design is largely heuristic and has not been rigorously tested in real-world applications. In particular, as the authors highlight, chosen study designs may have unforeseen consequences when applied to populations with heterogeneous density.

In the paper, the authors use a genetic algorithm to determine trap placement, a key decision in study design, and use a simulation study to determine the efficacy of the design this algorithm produces. The authors claim this design aid to be robust to heterogeneous density, irregular study regions, and constrained sampling effort (i.e. the number of available traps).

> We are very pleased to hear that we were successful in presenting these exact arguments, and that we effectively communicated the novelty of our work.

I have some major comments on the paper that I think should be addressed. Minor comments are some small things I have noticed during the review.

> We very much appreciate your feedback and having addressed all of your comments we feel the manuscript is clearer, and has been substantially strengthened. We hope you agree.

Major comments

1. Optimal: The paper is entitled "Towards optimal..." and the word "optimal" is used in several places (e.g. Line 67, "Our approach generates an optimal sampling design"). I believe the optimality refers to the near optimal performance of the genetic algorithm (Wolters [2015]) and not the optimality of the designs for SCR inference. There are no theoretical results showing the algorithm produces optimal designs in an idealised scenario and it is unlikely anyway that any design algorithm would have optimality across real world scenarios. I think it would be better if the paper was framed as providing a new design *aid* to practitioners that can lead to designs that in the scenarios

considered seem to perform well, but avoid making any claims of optimality. I will also add the same caveats to the claims of robustness: the simulation results provide some weak evidence that the designs are robust to different geometries and heterogeneous densities, but it can mislead the reader when it is stated (Line 73) “we show that the designs are robust”.

We agree. As you point out, the performance of the genetic algorithm is ‘near-optimal’ and our optimization approach produces a design that is near-optimal **with respect to the criteria used.** We have attempted to address these important distinctions when the concept of optimality and the genetic algorithm are first introduced. In addition, the extent to which we have addressed these points was somewhat dictated by the page limits, but we believe we have struck the balance of clarifying the text as per your comments, and remaining within space.

Line 65: changed *“optimal”* to *“near-optimal”*

Lines 68-69: changed *“optimal designs”* to *“designs”*

Line 103: changed *“for generating optimal designs”* to *“for optimizing SCR designs”*

Lines 107-108: changed *“metric for identifying optimal SCR designs”* to *“metric for efficient identification of optimal SCR designs”*

Line 207: changed *“designs generated using the optimal design algorithm”* to *“designs generated using the genetic algorithm”*

Lines 245-246: changed *“generating optimal designs”* to *“generating near-optimal designs”*

In the “Optimizing method” section we provide a far more explicit, honest, and hopefully satisfactory, definition of optimal, noting that genetic algorithm produces near-optimal solutions:

Changed: *“To identify the optimal subset of locations that minimize Qp or Qpm, we used a genetic algorithm implemented by the function scrdesignGA() in the oSCR package (Sutherland et al. 2019).”*

To: *"We applied a genetic algorithm (GA) to the task of finding a design that minimizes any criterion, noting that optimality here is with respect to the defined criteria, and in the context of the GA is 'near-optimal' (see Appendix 1 & Goldberg 1989)." (lines 110-112)*

We chose to remove "Towards" from the title. Originally, we had included this term to clarify that we do not claim that our framework obtains exactly an optimal sampling design for SCR that could be generically applicable (as you point out).

Regarding the point about robustness, again we realize now that as presented, the claim can be read as misleading. We maintain that the designs are robust to the common constraints we included, with the key point being that this is true as long as there is a sufficient number of traps (effort). With fewer traps, performance suffers monotonically in terms of bias and precision, however, this is true not only for the criteria and designs we presented but also for the pre-existing 2-sigma design, and would be the case for any possible design that uses an insufficient number of traps. To amend this issue, we chose to restate the ending of the introduction:

"We explore design performances for scenarios where we vary the spatial coverage of traps, the landscape geometry, and deviations from uniform spatial distribution of individuals." (lines 70-72)

2. Alternative optimization criterion: Efford and Boulanger [2019] recently proposed an alternative optimization criterion that approximates the coefficient of variation for the density estimate in the case of a regular grid; in essence, it quantifies a compromise between $Q_{\overline{p}}$ and $Q_{\overline{p_m}}$. I note that in the author's code (and oSCR package) there is a p_comb that tries to achieve the same thing. It would be fitting for the paper to include the optimization criterion from Efford and Boulanger [2019] and compare it with the currently presented options in terms of bias and root-mean-square-error. This could also serve in the irregular case the role that the "2σ" design does in the regular case.

This is a good point. We actually included the p_combo (now $Q_{\overline{p_b}}$) criteria in an original draft of the manuscript, but decided to remove it prior to submission, mostly due to space. However, having the point raised is motivation enough for us to reintroduce this *'compromise criteria'*. The revised manuscript now also reports the $Q_{\overline{p_b}}$ design and corresponding simulations. We do note that, while conceptually similar in terms of achieving a compromise, in some separate simulation work, our colleagues found the performance of the Efford & Borchers (2019) approximation of CV(D) breaks down for small numbers of traps and small sample sizes, whereas $Q_{\overline{p_b}}$ appears to be more robust. As such, we included $Q_{\overline{p_b}}$ rather than the Efford & Borchers criteria.

In addition, we attempted to more clearly emphasize in the discussion that we propose three criteria, but that alternative criteria exist, and that our aid can easily accommodate those alternatives (as well as new ones).

3. Representative samples: The paper should include a fuller discussion about how the objective of obtaining a representative sample is also important in the study design. For example, in the case of 49 traps in the irregular design, it is not sensible a priori to place the traps according to the $Q_{\overline{p_m}}$ design as it clearly favours the south-east of the study region (likely driven by the irregular shape of the study region meaning p is maximised when the centroid of the trap array is surrounded by the area of greatest spatial extent). A cluster design would seem more appropriate in this case and it would greatly improve the paper if a sensible cluster design were compared with the currently presented design; the cluster design could be devised manually. I argue that the proposed $Q_{\overline{p_m}}$ design would not be sensible because if the intention is to expand inference to that entire study region, I would not be comfortable claiming with that design limited to the south-east that I had a representative sample (and so would likely restrict my inference to a smaller area). This, of course, depends on the species under consideration, but for many such species my concerns would arise. For example, I predict that in the case of an (unknown) north-south linear density gradient, the proposed $Q_{\overline{p_m}}$ design would do extremely badly compared to a sensible cluster design.

We appreciate the point about cluster designs, and the fact that they may perform well relative to the designs produced here using the three proposed criteria. We had considered adding a ‘cluster design’ alongside the 2-sigma design but actually found this extremely challenging due to a series of decisions that must be made on the way to generating a cluster design:

1. Number of traps per cluster, and whether this is fixed or not
2. Number of clusters (related to (1), of course)
3. Within-cluster trap spacing
4. Between-cluster spacing

We quickly realized that, especially for an irregular space where these decisions matter much more but where they have not been explored, it would be difficult to generate generalities about cluster designs that are independent of these decisions without doing many many more simulations.

[As an aside, this exact issue played out in a Twitter thread on design recently, showing how, when following recommendations, a clustered design could vary quite a lot: https://twitter.com/wild_ecology/status/1272016846770835456?s=20]

We note that the p_combo criterion ($Q_{\overline{p_b}}$) we have now added can best be described as "clustered space-filling": small clusters of traps emerge that fill the state space. Therefore, this criterion gets us some way towards understanding how the other designs perform relative to a design that exhibits clustering.

The issue you present with limited effort (number of traps) and a strong gradient is an interesting one. We chose not to include a strictly linear gradient in our analysis and replaced it with the strong spatially-structured landscape, as we felt a linear gradient was not biologically realistic whereas a strong spatially-structured landscape was a better representation of the polar end of the scenarios we presented, opposite of uniform density structure. Considering the results from the strong spatially-structured landscape, we can see that your prediction is correct: design performance is not as strong, both in terms of bias and precision, and beyond our expectation of deterioration from decreased effort alone. However, on average, performance of $Q_{\overline{p_m}}$ designs in that scenario are not unacceptable, supporting the integrity of our framework and proposed criteria. Interestingly, in this scenario, the density-varying SCR model performs more poorly than the null model in recovering parameter estimates, and we have highlighted this nuance in our manuscript:

> *"This suggests that the low numbers of traps do not adequately represent the variation in the landscape, and therefore, the model is unable to estimate the underlying landscape effect (Figure 3, Table 1)."* (lines 225-228).

In summary, though, we definitely acknowledge that the general point is an issue - this comment was critical in our reshaping the manuscript, and we greatly appreciate the critique! While we dont add a cluster design, we have made the following changes that we believe acknowledge the issues raised:

1. The performance issue identified at low trap numbers is likely a matter of spatial coverage and that this could be resolved using cluster designs, but of course also by increasing resources.
2. Extending scrdesignGA() to include cluster-based trap selection (sensu Sutherland et al 2018) and developing alternative criteria (inc. cluster-based or spatial representation) are interesting avenues to explore.

4. Genetic algorithm details: The paper does not contain the details necessary to use the genetic algorithm. In particular, the paper should explain briefly how the algorithm works and what parameters must be tuned, e.g., the ngen parameter.

This was a very useful comment in helping us reshape our section on optimization, and we feel that we have thoroughly addressed this in our current manuscript submission. We did this in several ways:

First, we expanded on the genetic algorithm and k-of-n functionality in the "Optimization Method" section (lines 110-120).

Second, we included a separate appendix that describes the genetic algorithm in more detail based on our understanding of it, as well as a brief discussion of parameter tuning and suggestions (Appendix 1).

Third, we repeatedly point the reader to the original paper for the kofnGA() function by Wolters (2015), which we anticipate should be able to answer any additional questions. We do note, however, that Wolters admittedly spends little time on parameter sensitivity and tuning, suggesting that it's not critical to the final solution, and instead could be tuned more for the sake of efficiency, which we note in the appendix.

Finally, we include in the function itself and in the function help files the default parameters that we would suggest, serving as a starting point for users.

Overall, this should thoroughly address any and all questions about the function, both in terms of how it works as well as how it should be used.

Minor comments

(Line 31) It would be useful to readers to also cite the seminal papers on SCR: Efford [2004], Borchers and Efford [2008].

This was a regrettable oversight. We have added these important citations.

(Lines 43–45) The use of the word "captures" and "recaptures" is confusing here. Is *n* the number of unique captures and $g0$ controls the number of recaptures at a single trap?

Thank you for pointing this out, we have rewritten the sentence for clarity:

> *"The ability to reliably estimate density is directly related to these considerations: the number of captured individuals n is the sample size; the number of recaptures is directly related to the baseline detection probability,* g_0*; and the number and spatial distribution of recaptures are directly related to the spatial scale parameter,* σ*."* (lines 40-44)

(Lines 101–102) Equation between them would be clearer to read with brackets, i.e.

$$\prod_{j=1}^{J}\{1 - p(s_i, x_j)\}.$$

Same for sum in lines 107–108.

We have added these brackets as you suggested.

(Line 255) "lose" should be "loss"

Yes, it should. We have fixed this.

4_designs_irregular.Rcodefile: At the end of file, it saves variables that do not exist, e.g.designA.pcombo

The p-combo criteria was meant to serve as an intermediate to the two criteria we recommended in this manuscript, and one that we have now included again as per your suggestion ($Q_{\overline{p_b}}$). The file has been corrected accordingly.