*The Editor said this:*

*You will note that the earlier Stevenson et al paper is mentioned more than once, and so it would be useful in a revision to emphasise the differences between the results obtained from each method - one referee questions whether the approach in the current paper truly represents an improvement - especially with the application here. The application itself is also brought into question, with regard to how well it suits the methodology proposed - although as noted elsewhere, you do acknowledge this as a limitation.*

Dear Dr Brewer

We thank the Editor, AE and all three referees for their constructive comments.

Regarding the issues raised by the Editor, we discuss the pros and cons of the LCE and CCR methods in the Discussion section, and we have responded below to the specific issues regarding one referee’s comments about the utility of the LCE approach.

We agree with the AE about the importance and breadth of the simulation study and are fortunate that we were able to compare our results directly with those of another method (the CCR method) in Section 7

We have addressed all comments and give our detailed responses to the referees below, in red, following each comment by a referee. 

**Referee: 1**  
  
Referees comments to the Author. These may be passed without any edits.  
This is a very well-written and interesting article. I too believe that the future of mark-recapture, line transect surveys, distance sampling, etc. on wildlife population will be more or less replaced by digital aerial surveys, drones and other modern technologies. So it makes sense to develop new robust and computationally efficient statistical methods that can handle these types of platforms.  
  
My main concern is that the real data example doesn't completely align with the proposed methodology in its full entirety. Instead, a special case of the model is used. As an applied statistician, I was hoping to see the entire model fitted on some real data. However, I think it's good that the author's have at least acknowledged this limitation at the start of page 15.

Yes, we do not fit to real data because of a lack of data, although in principle we agree with the referee. Our application using the semi-synthetic data of Stevenson et al. (2019) is the closest we can come to an analysis of real data.

Also, I understand that the proposed method is "new" because it uses a full likelihood which leads to some nice modelling advantages; however it seems that the methods of Stevenson et al. (2019, <i>Biometrics</i>) give similar performances (as mentioned on page 20: "The performance of the LCE and CCR estimators is very similar, with the LCE estimator making a slight gain in precision as sample size increases"), and the CCR method can handle larger data sets (as mentioned on page 21: "The CCR method, by contrast, scales well and is able to deal with much larger numbers of detections.").

This is correct. The main advantage of the LCE method is that it is a maximum likelihood estimator and so enjoys the asymptotic properties of such estimators, including expressions for asymptotic variance. As such, it also provides a useful benchmark against which to compare the CCR method.  
  
Minor comments:  
  
- Page 1, first sentence: Could you add a reference for this opening claim?

Done  
  
- Page 5, 3rd line: The Zhang et al. citation is missing "in press".

Now published; new citation inserted.  
  
- Page 5, line 13: I've never heard of a Palm likelihood, is there a reference or some further theory for this?

Added a reference to a paper by Tanaka that contains details of the likelihood. Palm distributions are covered in most books on spatial statistics.  
  
- Page 5, line 16: What do you mean by "true likelihood"?

Changed wording to say “the CCR ``likelihood'' is an approximation to the actual likelihood”, which we hope clarifies the point.  
  
- Page 6, equation (1): I think the support, that is, t > 0 should also be included here?

Included it.  
  
- Page 6, second paragraph of Section 2.2: I feel like a diagram could be added here? It may help readers visualize the movement and availability on the line transect, and to see where the introduced notation fits in.

We have added a figure, as suggested (new Figure 1).  
  
- Page 8, line 5: Could the authors expand more on the assumption that the up/down state is independent of the in/out state? Not sure if this was explored in the simulations. Would this really hold in practice?

The assumption should hold in practice because animals will be unaware of whether they are in or out of the searched strip. We have changed the wording from

“Assuming that the up/down state is independent of the in/out state, the matrix of transition probabilities between these states at time separation t is the Kronecker …”

to

“We assume that the up/down state is independent of the in/out state, which is reasonable because animals do not know whether they are in or out of the strip. The matrix of transition probabilities between these states at time separation t is then the Kronecker..”  
  
- Page 10, line 6: I couldn't quite understand what is meant by "there are additional observations on the time delay t between detection by the first and second observers"?

To clarify, we have changed the wording from this:

“For animals with capture history w3 = (1; 1), there are additional observations on the time delay t between detection by the first and second observers, providing information about the movement parameter”

to this:

“For animals with capture history w3 = (1; 1), we observe the time delay t between detection by the first and second observers, in addition to the locations of the animals at these times. The time that it takes animals to move between the two locations at which they were observed provides information about the movement parameter”  
  
  
Page 10, line 15: Theta is introduced here but not fully defined. You could add: "see end of Section 4.2" after \theta in introduced?

We have added the words “(see Section 4.1)” (having removed the old Section 4.1, as suggested below).   
  
- Section 4.1: Is a seperate section (Section 4.1) really needed for the homogeneous density? Could this be just be mentioned as a special case after equation (11) is introduced?

We have removed Section 4.1 as suggested. It is dealt with as a special case in the first part of what was Section 4.2  
  
- Page 17, line 20: It wasn't clear to me why 100 knots is added here? How is this simulated and why was it set to 100?

We have changed these words

“with an observer speed of 100 knots, which is around the typical speed of marine aerial surveys such as the harbour porpoise survey above”

to these

“with an observer speed of 100 knots, which is around the typical speed of marine aerial surveys such as the harbour porpoise survey above”

We have not given the details of how this was simulated, because it is fairly straightforward and Biometrics page length restrictions mean we don’t have the luxury of being able to describe all the details of the simulation. For information, the simulation was done by working in the time domain and converting from times between observers passing animals to distances between them, using an observer speed of 100 knots: (distance plane moved between locations) = (time between locations) \* 100 knots.

- Page 17, last line: L=1100 km seems like a very large distance to travel ... doubling and tripling this amount, even more so. Does this mean the plane or drone will continuously fly for 1100 km taking photos or information? Is this realistic?

The simulations were constructed to mimic a plane survey like the harbour porpoise survey of Section 6. A distance of 1100km at 100 knots corresponds to a flying time of just under 6 hours, which is very feasible for the aircraft, and is similar to the actual time flown.

A battery-powered drone would fly much slower (perhaps 10 km per hour) and for much shorter (probably less than an hour). A petrol-powered drone would have similar capabilities to the plane used on the harbour porpoise survey.  
  
- Table 1: There is no CI coverage probability % for CCR? Were there these worse, better or similar cf with LCE?

We did not estimate coverage probabilities with CCR, as our main focus is on the LCE method. The CCR paper of Stevension et al. (2018) does not contain coverage probabilities for the CCR method either.  
  
**Referee: 2**  
  
Referees comments to the Author. These may be passed without any edits.  
This paper presents a new method to estimate density (and other relevant disturbance parameters) when considering feeds of two cameras that have a time lag between them. The methods are illustrated with harbor porpoise derived data and the performance evaluated under simulated data. I suspect these and similar methods will have demand soon as surveying with drones becomes more and more common. The method is contrasted with an earlier method to deal with the same problem by Stephenson et al in Biometrics.  
  
I present below some comments and suggestions that might help to improve the manuscript. I really don’t have any “Fundamental issues”, so there’s only a few (2) “General comments”, which are issues that are rather general in nature (a few of them might mostly reflect me thinking about this for the first time and possibly getting the wrong end of the stick!) and (3) “Specific comments”, which are much more self-contained and mostly only editorial in nature.  
  
General comments  
  
When one reads the abstract one is left wondering…”Because detection of animals from the air is imperfect” – the way I see it, this method will help with availability bias, but not perception bias – or, in other words, what is perception bias in an image? What is the problem you are proposing to solve?

This is not quite correct. As we note in the paper, in some circumstances we can estimate the parameter p, which is the perception bias parameter. The main problem we solve is this: How can we use mark-recapture line transect methods when we do not know capture histories? We have changed these words in the abstract “We obtain the likelihood by automatically …” to these: “We obtain a likelihood for mark-recapture line transects without capture histories by automatically …”

I am a bit confused, as you state “Animal movement” as being the issue in page 1, but surely, if the two cameras are mounted on the same airplane, the speed of the animals would have to be ridiculously high for it to cause issues? What am I missing? You then later state that “In practice, the time delay will need to be sufficiently long relative to the duration of the diving cycle to ensure that the data are adequate to fit the availability model.” – I find it hard to believe that for most whales the time delay on a rear-facing and a forward-facing camera is enough to be on the order of the say tens of seconds successive breaches might take to occur. Ah.. but then you state “Mounting both cameras on the same UAV has the advantage of creating a different viewing  
aspect for the two cameras: an animal that is obscured from one camera by a bush or shadow might be detectable from the other camera.” –I like that idea. Gives one the opportunity to turn some otherwise availability bias into effectively perception bias. So maybe you chose your example wrongly, whales would not have been ideal, but say deer on a low canopy forest perhaps?

The utility of our method certainly depends on the length of the lag between the two observers relative to the availability cycle length, and for some species (e.g. some whales), mounting two cameras on one aircraft moving at 100 knots will not provide sufficient lag. We do address this issue in our simulation study, varying the lag by a factor of 8 (from 10 seconds to 80 seconds). We have now added the words “dive cycle length is 110 seconds” to the caption of Table 1 so that readers can see more easily how long the lag is relative to the dive cycle length.

One could of course obtain a longer lag by slowing the aircraft down. If it moved at 10 knots, as might a quadcopter drone, the lag would increase by a factor of about 10, for example. In this case animal movement becomes more of an issue. Our manuscript provides a general methodology, adaptable for a wide variety of scenarios; we illustrate it with one particular scenario and then investigate a total of 36 variants of this scenario by simulation.

You state “Ensuring that the narrow search strips of two UAVs overlap adequately can be difficult in some environments and a cheaper alternative is to mount two cameras on a single aircraft:”. While difficult, it does not seem impossible, and hence this might open the question to what might be the optimal strategy, depending on how likely this is. Two have two cameras on 1 UAV, with the small-time delay that this necessarily leads to, versus two UAV’s traveling with a delay, or even the same UAV getting back on itself every now and then, say. Have you thought about it? You later state “because animal movement in and out of the field of view of the cameras is itself an availability process.” Which seems to imply that the methods could deal also with only partly overlapping strips?

We have thought about these issues. In fact, we started by considering data from a two-plane visual aerial survey in which the second plane attempted to follow the path of the first, but with a lag of a minute or two. The practical difficulty of one plane following exactly the path of another became apparent to us from these data: the paths very frequently diverged by 200m or more. This is not such a problem for visual surveys where observers search out to more than 1,000 m either side of the transect line, but for a camera survey with strip half-width around 125m, there was too little overlap to make mark-recapture line transect methods feasible.

We do mention a method involving one aircraft circling back over its own path periodically, but this has the same problems as the two-plane method mentioned above. If the two aircraft could be made to follow almost the same paths, then both the two-plane and the circle-back methods may be feasible.

Our method could be extended to deal with partial overlap of searched strips, but we do not develop that extension in this paper. We present a framework that could be extended by others in this, and other ways.

Similarly, we have not investigated optimal strategies, but our framework provides the tools that others can use to investigate this.  
  
In page 5 you state that “we use a hidden Markov model formulation of the likelihood” – but it seems rather obscure what this means at this point since you’ve not defined what the states might be? It will become obvious later, but still…

We have changed the wording as follows:

“… we use a hidden Markov model formulation of the likelihood, with states reflecting animals' availability to be detected.”

I was confused when reading that “Previous literature has devoted substantial attention to each of the problems of availability and uncertain capture histories, but rarely together”, since there’s no point to think about capture histories, either certain or uncertain, if you are not interested in estimating availability. Can you perhaps reword?

We believe that the referee is incorrect. Mark-recapture distance sampling methods typically do not estimate availability explicitly, but do use capture histories – so there is indeed some point in thinking about capture histories even when you are not interested in estimating availability.  
  
In page 6 you say “They did not allow animal movement in the direction of aircraft travel” – given airplane vs animal speed, would that be a big problem for them?

No, it would be no problem for them, and indeed it is no problem for our application (we did investigate the effect of not allowing animal movement in the direction of aircraft travel) – because the aircraft moves so much faster than the animals. It would become a problem for a survey platform that moved much slower relative to animals, as might be the case for some drones for example. We have added the words “Neglecting animal movement in the direction of aircraft travel is not a problem for observers that travel much faster than the animals, but it is a problem for much slower-moving observers.”

In page 8, The matrix derived from equation 5, U(t)=exp(Qt) does not seem to have rows adding up to 1 (which should be the case for a state transition matrix, as say equation 4)? Is there some normalizing constant missing, or am I missing something?

The exponential in this equation is a matrix exponential, not a matrix consisting of each element of Q exponentiated. We have added the words “, where here exp() indicates a matrix exponential function”  
  
Might section 5.2 be an appendix? In some sense, this is just a description of the computational algorithm you used, so it becomes quite distracting while there’s no need to read it to understand the methods/message of the paper?

We have moved this section to ab appendix.

Figure 1 implies that correlation is higher when animals are moving. I would have assumed intuitively that if animals are moving, hence changing position, that would increase independence, not correlation? While not moving then there’s no possibility of going from in to out or out to in, and hence there are only two relevant states (up or down), and the correlation would not tend to increase – am I being deceived by intuition?

Yes, that intuition is misleading. Independence and lack of correlation is achieved when the time between the two observers passing over an animal (the lag) is long enough that the animal being available (or unavailable) at the time the first observer passes over it contains no information about its availability when the second observer passes over it. When animals are stationary, availability is determined only by the up/down process and with a dive cycle length of 110 seconds and mean time at near-surface of about 0.73\*110=80 seconds, independence is acheived at a lag of something over 200 seconds. But with estimated mean displacement of 0.58 m/s and searched strip half-width of 125 m, there is still a substantial probability that an animal initially inside the searched strip will still be inside it 200 seconds (and more) later, and therefore initial availability is not independent of availability 200 seconds (and more) later.  
  
A correlation of 0.75 between two estimators means that almost (r^2=0.56) 50% of the variation remains unexplained. That seems a quite large value to me, assuming that both estimators are unbiased… I would assume a larger correlation across estimators – what are the sources for between estimator variability? The fact that you get differences as large as those observed in practice not more than 20% of the times does not fill me with confidence either. What does this mean… one can easily expect to be 20 % off the truth when one chooses one vs. the other when one does not know which is closer to the truth?

No, that is not the correct interpretation. The correct interpretation is that were one to use the LCE and CCE estimates from the harbour porpoise survey to test a null hypothesis that the expected value of the LCE and CCE estimators is the same, one would not reject this null hypothesis at any significance level less than 20% (and so not reject it at the commonly-used 5% significance level).

The fact that R^2=0.56 is of no concern. The LCE and CCR estimators use somewhat different kinds of information and so it would be a bit surprising (although admittedly comforting) if they were very highly correlated. Suppose you had two estimators that were equally good in terms of (lack of) bias, variance, and mean squared error. The fact that they had (let us suppose) very low correlation would not make either one of them a poorer estimator. Correlation between different estimators is not an issue when we are concerned with evaluating the adequacy of any particular estimator, as we are here.

We don’t really know what the source of between-estimator variability is – beyond the fact that the two estimators use somewhat different information and are mathematically different.  
  
I find it a bit confusing in figure 3 (which I find hard to read), but it seems like the mean bias increases for larger samples sizes, in particular for the crosses (CCR)?

The bias does not increase. We have added these words to the caption: “From left to right, the mean percentage difference within each of the three groups of estimates is 2.8, 2.9 and 2.4 in the case of the LCE estimator, and 1.3, 3.0 and 2.8 in the case of the CCR estimator”  
  
In the discussion, you state: “For UAV surveys of non-diving animals, including land surveys, the parameters \tau and  \gamma are not needed,”, but one could come up with an availability process for terrestrial animals too… even without animal movement, a forward-facing camera might not see an animal obstructed by a bush that is visible to a backward-facing camera… will this be able to be accounted for?

It seems to us that it would not in general be sensible to model the availability process as a Markov process when availability is determined by whether or not the view of an animal is obscured (by bush, or whatever). Better to treat the probability of being obscured or not as part of the observation process and in this case the probability that an animal is obscured by bush would be reflected in the estimate of the parameter p.

We have added this sentence: “When animals on a terrestrial survey are missed because view of them is obscured (by bush, for example), the probability of being obscured would be reflected in the estimate of p.”  
  
Specific comments:  
  
In page 2, I suggest “aircraft's path projected on the ground” would be more accurate than just “aircraft's path”

Done  
  
In page 3, while the example “and birds or amphibians may be available only when vocalizing” is correct, it does not apply in the case of cameras, so since this is the 3rd example, I’d remove it?

We want to keep it because we see (and in fact are pursuing) applications of LCE-type estimators in acoustic surveys. We have split off the birds and amphibians into a new sentence: “For example, whales are unavailable while diving, and seals are unavailable at haul-out sites while they are at sea. Or, in the case of acoustic surveys of calling animals, animals are only available when vocalising.”  
  
In page 4, should “in whether animals detected in similar locations by the two cameras is the same animal” be “in whether animals detected in similar locations by the two cameras are the same animal”? Difficult one, the correct would depend on whether there are 2 animals or 1, which we don’t know! If you replace “animals” by “detections” my suggestion works though.

Changed to “there is uncertainty in whether detections in similar locations by the two cameras are the same animal or two different animals”.  
  
Page 6, something is wrong with “Zhang et al., ress”

Fixed, with updated citation.  
  
In page 9, perhaps better to be explicit and replace “so that only animals in the `up' state can be detected” by “so that only animals in the `up' and `in' state can be detected”

Done  
  
Re notation, choosing \theta for the spatial density parameter is misleading since the entire unknown parameter vector what bold \theta – I think choosing a different letter helps to avoid confusion since if I got it right from equation 12 the bold \theta will have several non \theta parameters within it too.

We have replace theta with eta.  
  
In page 12 out of 31 you have “For the rest of this paper, we focus on the constant density model with identical detectors and no covariates, which has five parameters” – I’d put the \bold theta there, before the 5 parameter set. Presumably, that’s a good way of reminding the reader what it is, the set of parameters.

Done

I’d had a “typically” or some similar qualifier to the statement “The field of view of a digital camera is such that objects towards the periphery of the image are as easily detected”, as if the cameras were really oblique to increase the timing between detections one could observe considerable distortion at the edges of the image?

Added “typically” as suggested.

The sentence “A CSP is a triple P = (X,D,C)” seems meaningless to me? And what is a simple P?

\*\*\*\*\*\*\*\*\* return to this once hear back from Peter \*\*\*\*\*\*\*\*\*\*\*\*  
  
Equation 16 – what is P\_m?

We have replaced ^n\_1^\*P\_m by n\_1^\*!/( n\_1^\*-m)! for clarity  
  
Figure 2 – I find it awkward that one can’t know which scenario is which? Or is that considered irrelevant?

We have added vertical lines separating estimates into the three groups shown in the next figure, and added these words to the caption “Scenarios are numbered in the same order as they appear in Table 1. Vertical dashed lines separate simulations according to the three sample size groups shown in Figures 4 and 5.”  
  
Reference typos:  
  
In 3 Borchers references, twice “markov” instead of “Markov”, and also “poisson” vs “Poisson”

Fixed

Typo in Hamilton et al 2018 – “andgroup size” and “354?362”

Fixed  
  
In link et al 2010 – “multinomial”, not “multi-nomial”

Fixed

Pike & … - “sightingsfrom” and “arctic” and “resreport”???

Fixed  
  
There’s a doi for Zhang et al, aleready more than in press at online early – use it ? <https://doi.org/10.1111/biom.13030>  
  
Done

Referee: 3  
  
Referees comments to the Author. These may be passed without any edits.  
The manuscript addresses the important and challenging problem of animal population size estimation using digital aerial surveys. The bottom-up approach for the construction of the state-space models is very interesting, and the proposed estimation methods seems effective and was able to produce consistent estimates of model parameters with those by a different approach reported in the authors' recent work. The presentation of the work is also of high quality. I have only a few comments for the authors to consider in a revision:  
  
1. The model for in-out availability of the  animals was based on a 2-D Brownian motion model for individual animal movement, but the up-down availability is simply assumed to follow a Markov process with up/down times following exponential distributions, in contrast to the inverse Gaussian distributions for the in/out times. This apparent inconsistency of the models for horizontal and vertical movements of the animals seems a bit odd. I understand that it might lead to technical difficulties to use a 3-d Brownian motion (with truncation/reflection at the surface) for animal movement, but is it possible to give some sort of justification to the modelling choice in this paper, or show that the choice does not have material influence on the final estimate of the abundance parameter?

We think that using different models for horizontal and vertical movement is appropriate. This is because the vertical movement process is a different process to the horizontal movement process. Vertical movement is governed by the animals’ need to breath and to dive, and is necessarily somewhat cyclical, alternating between being up and diving, while horizontal movement does not have this alternating feature. Unlike the diving behaviour where an animal at the surface must return to the surface within some relatively short time after leaving it, there is no need at all for an animal to return periodically to its initial location in the east or north directions.  
  
2. An interesting point of the reported approach is that the capture history of the detected animals is not assumed to be known with certainty. However, this makes the form of the data a little ambiguous. In the classical capture-mark-release-recapture experiment, the data effectively consists of 3 numbers: the number of animals captures in the first capture occasion only, that in the second occasion only, and that in both occasions. It seems that the data to be modeled in the current paper consists of the numbers of animals detected by the first and second observers respectively, and the coordinates of these animals. For the sake of clarity, it is perhaps worthwhile to specify the form of the data available somewhere in the paper.

We have added this sentence to Section 6:

“The data for the LCE and CCR methods comprise the two sets of locations of detected animals: $\bm{s}\_1$ for those detected by observer 1, and $\bm{s}\_2$ for those detected by observer 2. Which observer 1 detections are ``recaptured’’ by observer 2 is unknown, and the LCE method considers all possibilities.”