Biometrics 00, 1–??
?? 2018
DOI: 10.1111/j.1541-0420.2005.00454.x

A latent capture history model for digital aerial surveys without recapture identification

D. L. Borchers[1,*], P. Nightingale[2], B.C. Stevenson[3], and R.M. Fewster[3]

[1]University of St Andrews, Centre for Research into Ecological and Environmental Modelling, St Andrews, Fife, UK

[2]Department of Computer Science, University of York, Deramore Lane, Heslington, York, UK

[3]Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand

email: dlb@st-andrews.ac.uk

Summary: Line transect aerial surveys can be a relatively cheap way of surveying wildlife populations, and we anticipate that they will increasingly be carried out by unmanned aerial vehicles. Because aircraft move quickly, these surveys typically violate the conventional line transect assumption of certain detection on the line if animals are not continuously available for detection. In this case, mark-recapture distance sampling with two observers may be effective. Here we develop a mark-recapture method for surveys with two high-resolution digital cameras, which involves marginalizing over the latent capture histories that arise when animals cannot be individually identified but may move and change their availability between the passage of the two cameras. The survey transect is partitioned into small segments within which there is ambiguity over animal identity, and the likelihood is obtained by enumerating all possibilities within each segment. We describe the new method as 'Latent Capture-history Enumeration', or LCE. The method requires that the species of interest be identified on each camera, but does not require identification of "recaptures" on different cameras.

We compare the LCE method with the recently-developed cluster capture-recapture (CCR) method, which tackles the same problem using a Palm likelihood approximation and does not involve a true likelihood. We apply both methods to the same data taken from a survey of harbour porpoise (*Phocoena phocoena*) in the North Sea. The LCE method is found to be approximately unbiased, to have close to nominal confidence interval coverage, and to be slightly more precise than the CCR method. It is less able to cope with a very large number of latent capture histories than is the CCR method, but is more extendable in some respects.

lets see...

RMF - added a bit more detail and LCE name. Removed CCR ref for a smoother read.

Key words: Availability bias; Double-observer survey; Line transect; Mark-recapture; Movement model; Poisson process.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

RMF - Suggesting some changes in emphasis, taking into account your latest comments about MRDS. It seems to me the usefulness of this work is:

1. A method for MRDS from a single UAV for ANY aerial survey: animals that move but might not necessarily dive.

2. Derivation in full generality including both movement and diving model. The diving part of the model doesn't need to be used if not applicable.

3. Suitable for diving animals if external data is available for τ.

4. Validation of CCR.

I've had a go at re-angling the Intro with this structure. Can we change the title so it doesn't narrow it down to marine mammals, e.g. "digital aerial surveys of wildlife"?

Aerial surveys of wildlife populations allow large areas of land or sea to be surveyed at relatively low expense. We anticipate that aerial surveys with human observers will increasingly be replaced by unmanned aerial vehicle (UAV) surveys using digital video or still cameras to detect animals. Such surveys present some new statistical challenges. In this paper we address these challenges and develop a method of estimating animal density from cameras deployed from UAVs.

Traditional aerial surveys using human observers involve a reasonably wide field of view, perhaps 1000m or more to either side of the aircraft. Detections of animals decrease with distance from the aircraft, an effect that is modeled using a detection function. By contrast, aerial footage from UAV-mounted cameras has a much narrower field of view — perhaps 100m to either side — and detectability can be assumed to be constant within this zone. The narrow field of view can be compensated for by deploying UAVs on-effort for much longer periods of time than human observers.

Conventional line transect analysis methods assume that all animals at distance zero from the transect line (the aircraft's path in our case) are detected with certainty.

David removed this sentence: "In the case of aerial surveys, this implies that all animals directly underneath the aircraft must be detected, so that the numbers missed at higher distances can be estimated." and changed a few words in the paragraph because aerial surveys with humans can detect things 500m or more ahead of the plane.

If this assumption cannot be met, extensions based on mark-recapture methods are employed: see Burt et al. (2014) for an overview. The basis of mark-recapture extensions to line transect methods is to have two observers who search the area independently of each other. The two observers serve as two "capture occasions", and animals detected by both observers constitute "recaptures". The mark-recapture design enables us to estimate the shortfall of detections on the line, as long as detection of each animal by each observer can be considered a random event governed by estimable probabilities.

When employing mark-recapture based methods, it becomes important to distinguish between two sources of detection failure, commonly described as *availability bias* and *perception bias*. Animals are said to be unavailable when they are guaranteed to be undetected by both observers: for example, a cetacean is unavailable to both observers while it is diving.

DLB removed " for the same reason"

DLB removed "Detections of unavailable animals cannot be treated as the outcomes of random trials for the two observers, because there is no chance of detection by either observer." because they can be treated as random trials if availability is a random variable, the the key point is the next sentence below.

Because the unavailable portion of the population is effectively unsampled, there is no information from which to estimate how large this portion is. It is therefore important that any systematic causes of unavailability, such as the diving cycle for cetaceans, are explicitly accounted for by survey design and analysis methods.

By contrast, perception bias describes the case where an animal may be detected by one or the other observer

DLB removed ", and the circumstances that cause it to be undetected by one observer do not affect the other" because this seems to contradict the sentence below saying detections may be dependent.

. Mark-recapture methods can readily accommodate this, because successful detection can be treated as a random outcome for each observer. Models that incorporate dependence between the outcomes for the two observers are well-developed – see Laake (1999); Borchers et al. (2006); Buckland et al. (2010). In aerial UAV surveys, with two cameras mounted on the same aircraft, a particular animal might be obscured from a forward-mounted camera by a bush or shadow, but detectable by a rear-mounted camera, or may move and become available to one camera when it was unavailable to the other.

In the case of cameras mounted from UAVs, adjusting for detection failure using a forward-mounted camera records each section of line several seconds before a rear-mounted camera records the same section. Separation of the cameras in both time and viewpoint offers a solution to the unavailability problem. For example, although there might be a strong correlation between the diving status of an animal at the passage of the two cameras, there is still a chance that the animal might dive or surface in the time between their passage. The separation therefore generates data with which we can model the availability cycle, although if the cycle is much longer than the time between cameras passing there may be inadequate data to model the cycle.

DLB: I felt we needed to say something about availability cycles much longer than time between cameras, as method won't work if every detected animal is either available or unavailable to both observers, and won't work well if almost all are.

There are however two complications to this design. Firstly, animals may move between the passage of the first and second camera, so there is uncertainty whether animals detected from the first and second camera are the same animal or two different animals. We describe this

as *uncertainty in capture history*. Each detected animal has a true capture history specifying which of the two cameras detected it: camera one only (capture history 1,0), camera two only (0,1), or both (1,1). When animals are detected from the air, there are usually inadequate visible features for distinguishing between individuals, so recaptures are determined purely on the basis of spatial location and time. The longer the time elapsed between the passage of the two cameras, the more difficult it is to distinguish between recaptures of a single individual, and captures of two different individuals. Rather than the capture histories being observed data, as in conventional capture-recapture studies, they are now latent variables.

Secondly, although separation in time enables us to deal with availability processes such as diving, there remains dependence between the animal's availability state at the passage of the two observers, so we are forced to adopt a model that accommodates this dependence. The dependence is reduced as the time delay between the passage of the two cameras increases, although if animals move, the dependence never goes to zero (as we demonstrate below). Moreover, longer delays exacerbate the problem of capture-history uncertainty. If both cameras are deployed from a single UAV, the maximum time delay between them is limited by practical considerations. If cameras are deployed on different UAVs there may be considerable practical difficulties ensuring that the two searched strips overlap sufficiently as wind and other operational factors can result in deviations from intended flight paths.

It strikes me that on land, and with GPS-enabled UAVs this may not be the case! I think it is not very challenging to program a drone to fly a pretty exact path, automatically compensating for wind etc.

Here, we develop an analysis framework suitable for aerial surveys conducted from a single UAV carrying two cameras, which we anticipate will be a common survey mode in the future. We show that for unbiased inference it is necessary to model the movement of animals into and out of the narrow detection strip during the time between the passage of the two cameras. This time lag creates an "in/out" availability process that induces dependence between the

two observers. For diving animals, we further consider the "up/down" availability process by explicitly modeling the diving cycle. As noted by Stevenson et al. (2018), two-observer survey data do not contain sufficient information to identify all parameters of the diving model, so the mean dive-cycle duration must be estimated from external data. Additionally, we deal with the problem of uncertain capture histories by identifying sections of the transect line with ambiguous animal identities and enumerating all possibilities within these sections. Our approach is the first fully likelihood-based method that does not require capture histories to be assigned to individuals, and incorporates both in/out and up/down availability process.

RMF - now go into previous work. Focus it on the CCR paper and point out that another output of this work is to validate CCR: a rare opportunity to do so with a proper likelihood. For uncertain duplicates, I can put in Hamilton et al 2018, Pike and Doniol-Valcroze 2015, and Southwell et al 2002. We'll need to slip Hiby and Lovell in but downplay it. How do the Borchers and Langrock methods fit into the context? What do they do with respect to time lag between observers, or is there just one observer?

DLB: Borchers and Langrock do have two observers, but although they model availability to account for $g(0) < 1$, they identify recaptures. The only reasons to mention Borchers+al:13, Langrock+al:13 and Borchers+Langrock:15 here are (a) they model availability - which is relevant here, and (b) it will increase my citation rate!
Good if you could put in the refs you mention above, but I had also thought that papers like Vale, R.T.R., Fewster, R.M., Carroll, E.L., and Patenaude, N.J. Maximum likelihood inference for model Mt,α for capture-recapture data with misidentification. Biometrics, 70, 962971, 2014, and related stuff on latent capture history models done by Brett McClintock and by Bill Link would be relevant.

DLB: I have commented out the rest of the text in the Intro, as it is superceded by what is written above.

2. Models for movement and availability

RMF - subsection structure edited and various small edits throughout this section

Two observers move along a transect line, one behind the other, searching a strip of half-width w. The observers move at constant speed v and are separated by a time-lag l and distance vl. We use the general term 'observers' and develop the model for lags l of any size, although we anticipate that the two observers are most likely to be two cameras mounted on the same UAV.

We use two coordinates for location: the forward coordinate along the transect line, and the transverse coordinate perpendicular to the line. We say that an observer 'passes over' an animal at the instant that their forward coordinates coincide, regardless of the animal's transverse coordinate at that instant. We assume that observer speed exceeds animal speed, so the time at which each observer passes over each animal is well-defined.

Animals may move between the passage of the first and second observers. We model animal movement as a Brownian motion, such that the animal's displacement over time t follows a bivariate normal distribution with mean $(0, 0)$ and variance $\boldsymbol{\Sigma}(t) = \sigma^2 t \boldsymbol{I}$, where $\boldsymbol{I}$ is the 2×2 identity matrix.

2.1 *Forward movement*

RMF - how to get the appendix labels correct? I would give up and do it manually.

For a given animal, the time elapsed between the passage of the first and second observers overhead is a random variable T. We show in Appendix 8 that T has probability density function (PDF)

$$f_T(t) = \frac{vl \exp\left\{-\frac{v^2(l-t)^2}{2\sigma^2 t}\right\}}{\sqrt{2\pi\sigma^2 t^3}}. \quad (1)$$

RMF - removed comment about forward distance of the animal during time t: not used.

2.2 *Transverse movement and in/out availability*

RMF - changed from $Y(0)$, $Y(t)$ to Y_0, Y_t for conciseness / convention. Also corrected $Y_t - Y_0 \sim N(0, \sigma^2 t)$.

The survey design involves two observers detecting animals within a strip of width w either side of the line. As animals may move into or out of the detection zone between the passage of the two observers, we consider the survey area to constitute a wider strip of width $b > w$ either side of the line, where the buffer b is chosen such that there is negligibly small probability that animals beyond b at the passage of the first observer will be within the searched strip w at the passage of the second. The buffered strip of width $2b$ therefore covers all animals that may be exposed to detection.

Let Y_0 be the signed distance of an animal to the right of the transect line when the first observer passes overhead, and let Y_t be its distance to the right time t later. We assume that $Y_0 \sim U(-b, b)$, independently for all animals. Our movement model implies that $Y_t - Y_0 \sim N(0, \sigma^2 t)$.

For each animal within the buffered strip b at the passage of the first observer, let the binary random variable Z_t be 1 if the animal is within the detection zone of width w at time t, and 0 otherwise. Thus Z_t describes the animal's in/out availability for detection at time t. The probability $\mathbb{P}(Z_t = 0 \mid Z_0 = 1)$ that an animal moves from inside to outside the detection zone during time t, and the probability $\mathbb{P}(Z_t = 1 \mid Z_0 = 0)$ that it moves from outside to inside the detection zone, are

RMF: corrected lower integral limit in eqn 4 from -w to 0; also added curly brackets and re-ordered eqn 5 to remove the double negatives.

$$p_{IO}(t) = \mathbb{P}(Z_t = 0 \mid Z_0 = 1) = \frac{1}{w} \int_0^w \left\{ \Phi(y - w; \sigma^2 t) + \Phi(-y - w; \sigma^2 t) \right\} dy \qquad (2)$$

$$p_{OI}(t) = \mathbb{P}(Z_t = 1 \mid Z_0 = 0) = \frac{1}{b - w} \int_w^b \left\{ \Phi(y + w; \sigma^2 t) - \Phi(y - w; \sigma^2 t) \right\} dy \qquad (3)$$

where $\Phi(\cdot; \sigma^2 t)$ is the cumulative distribution function of a normal random variable with mean zero and variance $\sigma^2 t$.

We model in/out availability as a two-state Markov process with transition probabilities over an interval of time t given by Eqns (2) and (3):

$$\boldsymbol{M}(t) = \begin{pmatrix} 1 - p_{IO}(t) & p_{IO}(t) \\ p_{OI}(t) & 1 - p_{OI}(t) \end{pmatrix}. \quad (4)$$

The stationary distribution of the in/out Markov chain, which gives the long-term proportion of time spent in each state, is $(w/b,\ 1 - w/b)$.

2.3 *Diving behavior and up/down availability*

We model animal diving behavior using a two-state continuous-time Markov chain such that the time spent in state 1 (the near-surface state) is an exponential random variable with expected value κ, and the time spent in state 2 (the diving state) is an exponential random variable with expected value $\tau - \kappa$, where τ is the expected dive cycle duration. The Markov transition rate matrix $\boldsymbol{Q}$ is

$$\boldsymbol{Q} = \begin{pmatrix} -\frac{1}{\kappa} & \frac{1}{\kappa} \\ \frac{1}{\tau-\kappa} & -\frac{1}{\tau-\kappa} \end{pmatrix}. \quad (5)$$

The transition probability matrix for transitions between up/down states at time separation t is $\boldsymbol{U}(t) = \exp(\boldsymbol{Q}t)$. The stationary distribution of the Markov chain is $(\gamma,\ 1 - \gamma)$, where $\gamma = \kappa/\tau$.

RMF - put the stationary distribution $\boldsymbol{\pi}$ inline, and we don't need a notation for it.

2.4 *Combined availability model*

The possibilities of being in or out of the detection zone, and up or down with respect to diving, generate four states that animals can occupy: (up and in), (up and out), (down and in), and (down and out). We number the states 1 to 4 in that order. 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 product $\boldsymbol{\Gamma}(t) = \boldsymbol{U}(t) \otimes \boldsymbol{M}(t)$. Using a matrix formulation for $\boldsymbol{\Gamma}(t)$ provides an extendable and computationally efficient way of dealing with the hidden states 2, 3, and 4.

The stationary distribution for the four-state Markov process is

$$\boldsymbol{\delta} = \left(\gamma \frac{w}{b},\ \gamma \left(1 - \frac{w}{b}\right),\ (1-\gamma)\frac{w}{b},\ (1-\gamma)\left(1 - \frac{w}{b}\right) \right). \tag{6}$$

3. Detection model

We assume that the probability that an animal is in each state at the time the first observer passes over it is given by the stationary distribution $\boldsymbol{\delta}$, and hence that its state distribution after a waiting time t, when the second observer passes it, is $\boldsymbol{\delta}\boldsymbol{\Gamma}(t)$.

Define the binary variable X_{ij} to be 1 if animal i is detected by observer j and zero otherwise. We model X_{ij} as a state-dependent Bernoulli random variable with parameter $p_j(c) = \Pr(X_{ij} = 1 \mid C_{ij} = c)$ where C_{ij} is the state of animal i when observer j passes over it, and $c \in \{1, 2, 3, 4\}$. It follows that X_{ij} $(j = 1, 2)$ are observations from a Markov modulated Bernoulli process.

RMF - no need to define t_{i1} and t_{i2} here. Acronym MMBP never used again so removed.

It is convenient to arrange the state-dependent probability mass functions of X_{ij} in a diagonal matrix (see Zucchini et al., 2016, Eqn 2.13). For observer j, this matrix is

$$\boldsymbol{P}(x_{ij}) = \begin{array}{c} \begin{array}{cccc} \text{up,in} & \text{up,out} & \text{down,in} & \text{down,out} \end{array} \\ \begin{pmatrix} \text{Bern}(x_{ij}; p_j(1)) & 0 & 0 & 0 \\ 0 & 1 - x_{ij} & 0 & 0 \\ 0 & 0 & \text{Bern}(x_{ij}; p_j(3)) & 0 \\ 0 & 0 & 0 & 1 - x_{ij} \end{pmatrix} \end{array} \tag{7}$$

where $\text{Bern}\big(x_{ij}; p_j(c)\big) \equiv p_j(c)^{x_{ij}} \{1 - p_j(c)\}^{1 - x_{ij}}$. The above matrix allows for animals to be

detected in the 'down' state, but not in the 'out' state. In what follows, we assume that $p_j(3) = 0$, so that only animals in the near-surface 'up' state can be detected.

Let t_i be the time elapsed between the passage of the first and second observers over animal i. Conditional on t_i, the probability of observing capture history $\omega_i = (x_{i1}, x_{i2})$ for animal i can be expressed as the following matrix product, which efficiently sums over hidden states:

$$\mathbb{P}(x_{i1}, x_{i2} \mid t_i) \;=\; \boldsymbol{\delta}\boldsymbol{P}(x_{i1})\boldsymbol{\Gamma}(t_i)\boldsymbol{P}(x_{i2})\mathbf{1}\,, \tag{8}$$

where $\mathbf{1}$ is a column vector of ones. We label the three observable capture histories as $\omega_1 = (0, 1)$, $\omega_2 = (1, 0)$, and $\omega_3 = (1, 1)$, and define $p_k(t) = \mathbb{P}(\,\omega_k \mid t)$ as given in (8). The overall probability of capture history ω_k is then

$$\tilde{p}_k = E_t\left\{p_k(t)\right\} \;=\; \int p_k(t) f_T(t)\, dt\,. \tag{9}$$

RMF - removed the joint probability statement above, as it isn't needed beyond the expectation; and the $p_{01}(t_i)$ etc notation, which isn't needed and can be confused with transition probabilities. Likewise replaced their only other occurrence in Section 7. Also simplified notation around ω and $\tilde{p}_k$.

4. Survey model

RMF - removed mention of $2bD(s)$, as the meaning of intensity is number per unit area. Simplified likelihood notation and derivation.

We assume that the number and locations of animals in the forward direction, within distance b of the transect line at the time that the first observer passes overhead, are governed by a Poisson process with intensity $D(s)$ at along-transect location s. We derive the overall likelihood under the supposition that the capture history ω of each animal is known. We will later revoke this requirement by marginalizing over all possible assignments of detections to capture histories.

Let $\boldsymbol{s} = (s_1, \ldots, s_n)$ be the observed forward locations of the n detected animals at the time

of first detection. We can write $\boldsymbol{s} = \left(\boldsymbol{s}^{(1)}, \boldsymbol{s}^{(2)}, \boldsymbol{s}^{(3)}\right)$, where $\boldsymbol{s}^{(k)}$ corresponds to locations of animals with capture history k for $k = 1, 2, 3$. Each set of locations $\boldsymbol{s}^{(k)}$ arises from a thinned Poisson process with thinning probability $\tilde{p}_k$. Because multinomial splitting of a Poisson process produces independent Poisson subprocesses, the likelihood of $\boldsymbol{s}$ is the product of the three likelihoods from the thinned subprocesses. For animals with capture history $\omega_3 = (1, 1)$, there are additional observations on the time delay t between detection by the first and second observers, providing information about the movement parameter σ. The conditional PDF of waiting time T, conditional on the capture history being ω_3, is

$$f_{T|\,\omega}(t \mid \omega_3) = \frac{f_T(t)\,\mathbb{P}(\omega_3 \mid t)}{\mathbb{P}(\omega_3)} = \frac{f_T(t)\,p_3(t)}{\tilde{p}_3}\,, \tag{10}$$

where the right-hand side of (10) is obtained from Equations (1), (8), and (9). This PDF is included as an auxiliary component to the Poisson process likelihood for $\boldsymbol{s}^{(3)}$.

Let L be the total transect length of the survey, and let n_k be the number of observations of capture history $k = 1, 2, 3$, with $n_1 + n_2 + n_3 = n$. Let $\tilde{p}_\cdot = \tilde{p}_1 + \tilde{p}_2 + \tilde{p}_3$ be the overall probability of detection. We write $\boldsymbol{s}$, $\boldsymbol{\omega}$, and $\boldsymbol{t}$ for the locations, capture histories, and (where available) time delays for animals $i = 1, \ldots, n$. The parameter vector is $\boldsymbol{\theta}$. The likelihood is:

$$\mathcal{L}(\boldsymbol{\theta}\,;\,\boldsymbol{s}, \boldsymbol{\omega}, \boldsymbol{t}) = \frac{\exp\left\{-\int_0^L D(u)\tilde{p}_\cdot\,du\right\}}{n_1!\,n_2!\,n_3!} \left\{\prod_{i=1}^{n} D(s_i)\right\} \tilde{p}_1^{\,n_1}\,\tilde{p}_2^{\,n_2} \prod_{i\,:\,\omega=\omega_3} \left\{f_T(t_i)\,p_3(t_i)\right\}. \tag{11}$$

4.1 *Homogeneous density*

In the homogeneous case, where density is constant throughout the survey, we have $D(s) = D$. The likelihood is:

$$\mathcal{L}(\boldsymbol{\theta}\,;\,\boldsymbol{s}, \boldsymbol{\omega}, \boldsymbol{t}) = \frac{\exp\left(-LD\,\tilde{p}_\cdot\right)}{n_1!\,n_2!\,n_3!}\,D^n\,\tilde{p}_1^{\,n_1}\,\tilde{p}_2^{\,n_2} \prod_{i\,:\,\omega=\omega_3} \left\{f_T(t_i)\,p_3(t_i)\right\}. \tag{12}$$

4.2 *Model parameters*

The model has four kinds of parameters:

Density parameters: In the case of the homogenous Poisson process there is one pa-

rameter, θ, such that $D = e^{\theta}$. When density varies with covariates, θ is replaced by a linear predictor involving a parameter vector.

Dive cycle parameters: The two-state dive cycle model described above is parametrized in terms of the mean dive cycle length, τ, and the mean proportion of time in the near-surface state, γ, which are linked to parameters α_τ and α_γ via log and logit links: $\tau = e^{\alpha_\tau}$ and $\gamma = e^{\alpha_\gamma}/(1 + e^{\alpha_\gamma})$.

Movement parameters: The animal movement model has one parameter, σ, which we model using a log link: $\sigma = e^{\phi}$.

Detection parameters: Assuming that animals are only detectable when in state $c = 1$ (up,in), we have two Bernoulli parameters to model: $p_1(1)$ and $p_2(1)$. These can be modeled using logit link functions. If the observers are identical digital detectors, it may be reasonable to assume these two probabilities are identical, i.e. $p_1(1) = p_2(1) = p = e^{\beta}/(1 + e^{\beta})$.

As is the case for density, covariates can be incorporated into the other three models by replacing the corresponding scalar parameter on the link scale with a suitable linear predictor involving the covariates.

RMF - don't need notation $\boldsymbol{\theta}^*$.

For the rest of this paper, we focus on the constant density model with identical detectors and no covariates. There are five parameters: $(\theta, \alpha_\gamma, \alpha_\tau, \phi, \beta)$. Stevenson et al. (2018) show that not all of these parameters are identifiable from the two-observer survey design. For the detection model, they assume that $p = e^{\beta}/(1 + e^{\beta}) = 1$. This is reasonable for digital aerial surveys conducted in calm sea states, if we define the near-surface state to be "at or breaking the surface": a state that is easily observed. The field of view of a digital camera is such that objects towards the periphery of the image are as easily detected as objects in the centre of the image, so a detection function that drops off with distance from the line is not needed.

Stevenson et al. (2018) also show that even when p is known, only two of $(\theta, \alpha_\gamma, \alpha_\tau)$ are identifiable, so one of these parameters must be estimated using external data. We follow Stevenson et al. (2018) and Hiby and Lovell (1998) and assume that the mean dive cycle duration $\tau = e^{\alpha_\tau}$ is estimated separately, so we treat it as known in the present survey. In what follows, we therefore assume that detection of animals in the up/in state is certain ($p = 1$); we use external estimates to set τ; and we estimate the remaining three parameters. These constitute the density, D; the mean proportion of time in the near-surface state, γ; and the movement parameter, σ. The parameter vector is therefore $\boldsymbol{\theta} = (\theta, \alpha_\gamma, \phi)$.

RMF - corrected estimating κ to estimating γ above. Also removed last sentence which is relevant to the next section.

5. Marginalising over the latent capture histories

The likelihoods (11) and (12) are formulated under the supposition that the capture histories ω_i are known for animals $i = 1, \ldots, n$. However, the core problem when observers are separated in time is that the capture histories cannot be known with certainty: they are latent variables. Here we address this problem by enumerating all plausible combinations of latent capture histories. We marginalize the likelihood by summing over the individual likelihoods for every plausible capture history combination. We refer to each combination of capture histories as a "pairing", since once the pairs of detections with capture history $\omega_3 = (1, 1)$ have been decided, the capture histories $(0, 1)$ or $(1, 0)$ of all other detections are determined, because we know for each detections which of the two observers made the detection.

Calling the mth set of pairings $\boldsymbol{\omega}^{(m)}$, and the associated vectors of first-detection locations and time delays $\boldsymbol{s}^{(m)}$ and $\boldsymbol{t}^{(m)}$ respectively, we obtain the likelihood for the parameters $\boldsymbol{\theta}$ as

$$\mathcal{L}(\boldsymbol{\theta}) \;=\; \sum_{m=1}^{M} \mathcal{L}\left(\boldsymbol{\theta}; \boldsymbol{s}^{(m)}, \boldsymbol{\omega}^{(m)}, \boldsymbol{t}^{(m)}\right) , \tag{13}$$

where M is the number of plausible pairings.

RMF - this implicitly assumes that all plausible pairings are equally likely, which I don't think is the case. If we applied this formula to literally all pairings, not just the ones we've deemed plausible, the way that the implausible ones would be disregarded would be by having an additional term for $P(\omega^{(m)})$ which would be zero for implausible pairings. As it is, we are saying that a pairing with a large separation is just as valid as 'data' as a pairing with a low separation, and assuming that the model will 'believe' the data from the latter more than the former and estimate σ appropriately. (Which it probably will, once you segment, so I don't expect the estimates will be badly affected.) It also means that cases where M is large make more contributions to the likelihood than cases where M is small.

Do you want to make any comment about this? I'm inclined not to as it opens a can of worms. It might be a target for referees, but it would be less obvious to referees if you write $\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{s}, \boldsymbol{t}, \boldsymbol{\omega}^{(m)})$ instead of $\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{s}, \boldsymbol{t} \mid \boldsymbol{\omega}^{(m)})$, and omit the comment about conditional likelihood. EDIT: I've now done this, because it is written like this later in (14).

I also added dependence on m to $\boldsymbol{s}^{(m)}$ and $\boldsymbol{t}^{(m)}$.

While this likelihood is easy to write down, it is challenging to evaluate because we need to enumerate all M plausible combinations $\boldsymbol{\omega}^{(m)}$. For any but very small sample sizes, the number M of possible pairings is very large. We tackle this problem by first partitioning the location vector $\boldsymbol{s}$ into subsets between which paired detections are impossible, to reduce the number of plausible pairings, and then using a constraint programming technique for efficient enumeration of all possible pairings within subsets.

5.1 *Subdivision of* $\boldsymbol{s}$

We partition $\boldsymbol{s}$ by "cutting" the transect line immediately after detections by observer j for which the distance to the next detection by the other observer is greater than a maximum possible distance that an animal could have moved between the two observers passing over it (d_{max}). This distance d_{max} must be decided using knowledge of the movement speed and behavior of the target species. A suitable value for d_{max} can be chosen by doing inference at a range of plausible values to find where estimates become insensitive to d_{max}. The cost of setting d_{max} too large is in computational speed; the cost of setting d_{max} too small is positive bias in estimation of D, since setting d_{max} too small will result in some animals with true capture history $(1,1)$ being assigned capture history $(0,1)$ or $(1,0)$.

RMF - changed C and c to R and r below, as C and c are HMM states.

Having divided the transect line into R segments, we enumerate the possible pairings $\boldsymbol{\omega}^{(m_r)}$ for segments $r = 1, \ldots, R$. Let M_r be the number of possible pairings in segment r. We calculate the likelihood as

$$\mathcal{L}(\boldsymbol{\theta}) \;=\; \prod_{r=1}^{R} \sum_{m_r=1}^{M_r} \mathcal{L}\left(\boldsymbol{\theta}; \boldsymbol{s}^{(m_r)}, \boldsymbol{\omega}^{(m_r)}, \boldsymbol{t}^{(m_r)}\right) . \tag{14}$$

When d_{max} is substantially smaller than most of the distances between detections by different observers, segmentation can lead to a massive reduction in computation time, making it quite feasible to compute what would otherwise be an intractable likelihood.

5.2 *Constraint programming for enumerating all* $\boldsymbol{\omega}^{(m)}$

For efficient enumeration of the possible pairings within one segment, we define a simple *constraint satisfaction problem* (CSP) (Russell and Norvig, 2010, Chapter 6). A CSP is a triple $\mathcal{X} = \langle \mathcal{V}, \mathcal{D}, \mathcal{C} \rangle$. The CSP $\mathcal{X}$ has a set of decision variables $\mathcal{V}$, each of which has a set of possible values that it may take, called its *domain*, where $\mathcal{D}(x)$ is the domain of $x \in \mathcal{V}$. In addition there is a set of constraints $\mathcal{C}$ that restrict the combinations of values that may be taken by the variables. A constraint $c \in \mathcal{C}$ is a relation defined on a set of variables

scope$(c) \subseteq \mathcal{V}$. A *solution* is an assignment of values to variables such that each variable is assigned a value from its domain, and all constraints are satisfied.

We define a CSP for a segment as follows. Two detections by different observers may be paired if and only if the distance between them is less than or equal to d_{max}. For each set $\{i, j\}$ of two observations that may be paired, we define one decision variable $c_{i,j}$ with domain $\{0, 1\}$. Variable $c_{i,j}$ is equal to 1 in a solution if and only if the two observations are paired.

Suppose we have two sets, $s_1 = \{i, j\}$ and $s_2 = \{u, v\}$, where i may be paired with j, and u may be paired with v, but the two sets are not disjoint: in other words $s_1 \cap s_2 \neq \emptyset$. The two sets cannot both be paired simultaneously because they share an observation. In all such cases we add the constraint $c_{i,j} = 0 \vee c_{u,v} = 0$ to prevent both sets being paired simultaneously.

We use a backtracking search procedure with forward checking (Russell and Norvig, 2010, Chapter 6) to enumerate all solutions to the CSP. The set of solutions to the CSP corresponds one-to-one to the set of valid pairings within the segment. When a solution is found, the part of the likelihood pertaining to that pairing is calculated, avoiding the need to store the set of pairings and allowing efficient calculation of $\sum_{m_r=1}^{M_r} \mathcal{L}\left(\boldsymbol{\theta}; \boldsymbol{s}^{(m_r)}, \boldsymbol{\omega}^{(m_r)}, \boldsymbol{t}^{(m_r)}\right)$.

5.3 *Interval estimation*

We estimate the variances of parameters using the inverse of the Hessian obtained in the fitting process. Confidence intervals for the parameters D, σ and γ are gained from the inverse log transformation of confidence intervals for θ and ϕ, and the inverse logit transformation of α_γ, assuming normality of the maximum likelihood estimators of these parameters.

6. Application

We use the term 'Latent Capture-history Enumeration' method, or LCE, to describe our framework.We developed this method in anticipation of digital aerial survey data becoming

widely used, but, pending the availability of analysis methods such as the LCE method developed here, such data are not yet available. We therefore estimate density from the semi-synthetic data used by Stevenson et al. (2018). These data are taken from an aerial survey of harbor porpoise (*Phocoena phocoena*) in the North Sea using human observers, compiled from periods when the aircraft circled back over its transect after a lag of $l = 248$ seconds. The two observers correspond to the two passes of the aircraft. Only data in a narrow strip of half-width $w = 0.125$ km are included, to mimic the narrow field of view and perfect near-surface detection characteristic of digital observers. As noted by Stevenson et al. (2018), a lag of $l = 248$ seconds is longer than any plausible value for τ and as a consequence the surfacing states of an animal at the times the two observers pass are independent and the estimator is robust to unknown τ. For shorter lags τ needs to be speficied (see Section 7).

RMF - Ben didn't need to specify τ for this analysis because the two observers were considered independent for such a long lag. How did you deal with this? There hasn't been a previous mention of τ not being needed if you can assume independence. In some ways this should come after the comments done in Fig 1, but I can see why you want to put the application first; so some sort of comment is needed.
DLB: I set τ in the estimation code, to be 110 seconds, as in Ben's simulations, but also checked that estimates were insensitive to τ by estimating with some other values. Added the two sentence at end of paragaph above.

Following Stevenson et al. (2018), we use a buffer of $b = 2$ km, beyond which we assume no animal could enter the detection zone between the passage of the two observers. Using a Palm likelihood approximation termed cluster capture-recapture (CCR), Stevenson et al. (2018) obtained the following estimates, with 95% confidence intervals in brackets: $\hat{D} = 1.05$ (0.84, 1.60) pods per km^2; $\hat{\sigma}_{palm} = 0.15$ $(0.11, 0.19)$ km; and the expected proportion of time in the surface state, $\hat{\gamma} = 0.86$ (0.56, 1.00). Using the enumerated likelihood via our LCE formulation, we obtain $\hat{D} = 1.24$ (0.97,1.6) pods per km^2, $\hat{\sigma}_{palm} = 0.09$ (0.07,0.11) km, and

$\hat{\gamma} = 0.73$ (0.55,0.91). The estimate $\hat{\sigma}_{palm} = 0.09$ corresponds to a mean rate of movement over $l = 248$ seconds of 0.58 m/s, with 95% confidence interval (0.47,0.71) m/s.

Stevenson et al. (2018) estimated the coefficients of variation (CV) of $\hat{D}$, $\hat{\sigma}_{palm}$ and $\hat{\gamma}$ to be 19%, 16%, and 13%, respectively. The corresponding estimated CVs from the LCE method are 13%, 10%, and 13%, respectively.

The estimates from the two methods are broadly consistent; the LCE method estimates there to be substantially less animal movement, slightly less time at the surface, and a higher animal density. As we cannot evaluate the relative merits of the methods on the basis of a single survey with unknown density, we investigate their performance by simulation.

7. Simulation study

RMF - removed the bit about MRDS and availability. It opens a can of worms, because for MRDS you have simultaneous observers so movement isn't relevant. With LCE, we deliberately introduce a time lag to enable us to estimate availability, but this means we have to create models for movement and diving, AND we still need external data to estimate τ. So why not just use a single-observer transect (no need for MRDS with digital cameras), and use external data to estimate the proportion of time available, instead? Comparisons with MRDS here might invite the reviewers to ask awkward questions...

DLB: I am OK with this change, although in general it is not true to say that MRDS has simultaneous observers, and that movement isn't relevant (it is true for digital cameras). The reason we do not use single-observer transect methods is that we do not want to have to use external data to estimate the proportion of time available - because these data are typically difficult to get, and the proportion may change from place to place and time to time, so that you may be using the wrong proportion. This difficulty was one of the main reasons for developing our method. Us needing external data for τ is unfortunate (we did not know we'd need this when starting to develop this method). The density estimate may be less sensitive to misspecification of τ than the single-observer method is to misspecification of the proportion, but I guess we don't want to invite referees to ask us to look into that.

Recall that X_{i1} and X_{i2} are detections of animal i by two observers separated by a time lag. With lags close to zero, X_{i1} and X_{i2} are highly correlated because animals available to one observer are almost certain to be available to the other. As lag increases, we expect this correlation to decrease. A pertinent question is whether dependence can be removed by choosing a suitably long lag. To investigate this we look at the correlation between X_{i1} and

X_{i2} as a function of lag. We consider a range of γ values from 0.1 to 0.9, and lags from 0 to 500 seconds.

In our model, X_{i1} and X_{i2} are Bernoulli random variables with expectation $\gamma w/b$. The correlation between these variables when there is a separation of t seconds between the two observers passing over an animal, is

$$\rho(t) \;=\; \frac{\sum_{x_{i1}=0}^{1}\sum_{x_{i2}=0}^{1}\left(x_{i1}-\gamma\frac{w}{b}\right)\left(x_{i2}-\gamma\frac{w}{b}\right)\mathbb{P}(x_{i1},x_{i2}\mid t)}{\gamma\frac{w}{b}\left(1-\gamma\frac{w}{b}\right)}\,, \tag{15}$$

where $\mathbb{P}(x_{i1}, x_{i2} \mid t)$ is given in (8).

RMF - replaced notation $p_{X_1X_2}$ which I removed from Section 3.

The dark line in Figure 1 shows the correlation as a function of the lag (l) for $\tau = 110$ seconds and γ and σ equal to the estimates obtained in the previous section. It also shows the correlation for $\gamma \in \{0.1, 0.2, \ldots, 0.9\}$ and the correlation under the assumption that animals do not move but do become unavailable by diving.

[Figure 1 about here.]

It is clear that increasing the lag to τ or more reduces correlation to approximately zero if animals do not move (grey lines), although in the presence of animal movement there is still correlation between observers due to in/out availability. This corroborates the observation of Stevenson et al. (2018) that correlation between observers due to up/down availability can be removed by setting a lag greater than τ. With such long lags, the up/down availability model requires only the single parameter γ, corresponding to the proportion of time spent at the surface. This has the considerable advantage that there are no unidentifiable parameters, so no external data are needed to estimate density.

RMF - removed the bits about needing a model for availability and can't design ourselves out of correlated detections. If we're talking about lags at all, we already have a model for availability. I thought the real point about a long lag is that we no longer need external data to estimate the extra parameter τ. This links back to the comment in Section 6 about what you did about τ there? I've made a suggested edit, but this needs to be tidied up more carefully by a comment in Section 6.
DLB: There is received wisdom in MRDS that increasing lag decreases dependence, but no received wisdom that dependence never goes to zero. Hence the sentence seems worthwhile to me. But removing it is OK.

In practice, we are primarily interested in methods for surveys with two cameras on one aircraft, and with this configuration and fast-moving aircraft, lags of more than some tens of seconds are unlikely to be achievable. In light of this, and the results of Figure 1, we present simulations for (a) a scenario designed to imitate the porpoise survey above, and (b) scenarios with lag l of 10, 20, 50 and 80 seconds, and γ equal to 10%, 20%, 50% and 80%. We do this for σ equal to 1.5 m/s (the speed estimated by Hiby and Lovell, 1998), 0.95 m/s (the speed estimated by Westgate et al., 1995), and 0.5 m/s (a speed lower than that estimated above or by Stevenson et al., 2018). In all cases, we use the same τ in estimation as was used in simulation. For the short-lag scenarios in (b), we perform simulations with true density $D = 1.24$, as estimated in the previous section, and with an observer speed of 100 knots, which is around the typical speed of marine aerial surveys.

DLB: added this sentence

Need to say how many simulations.

7.1 *Simulation based on harbor porpoise data: lag of 248 seconds*

For this scenario we use the estimates of Stevenson et al. (2018) as the generating values, corresponding to $D = 1.05$, $\gamma = 0.86$ and $\sigma_{palm} = 0.15$. We investigate by simulation the bias and precision of the LCE estimator. In the light of our results in Section 6, where we

obtained an LCE estimate that was 18% greater than the CCR estimate of Stevenson et al. (2018), we also investigate whether this discrepancy is within the bounds expected due to estimator variability.

RMF - moved paragraphs around to reflect order of importance.

The empirical bias of the LCE and CCR density estimators from the simulations are 9.9% (CV=29.8%) and 12.7% (CV=37.9%), respectively. The biases reduce to 4.3% (CV=18.0%) and 5.1% (CV=21.4%), respectively when sample size is doubled while holding density constant, and to 2.9% (CV=14.6%) and 3.3% (CV=16.7%) when sample size is trebled.

BCS: Changed from 7.1% (CV=24.6%) and 10.9% (CV=33.3%) for original sample size, 3.5% (CV=20%) and 4.1% (CV=23%) for doubled sample size, and 2.9% (CV=17%) and 3.5% (CV=18%) for trebled sample size. I suspect some of the new values are larger because my simulations had $N \sim \text{Poisson}(E(N))$, but perhaps the old ones had N fixed? I think other differences are because the original values only used 150 iterations, but these used 1000. Should I redo with fixed N? Can easily be done.

BCS: Changed from 0.86

The correlation between LCE and CCR density estimates from the simulations is 0.75, while the probability of getting a relative difference as large as, or larger than, that observed is approximately 20%, from which we conclude that the observed difference is not large enough to raise concerns about the validity of either estimator with the porpoise data.

BCS: Changed from 18%

The LCE estimator formulates the delay in encounter times between the two observers as a random variable, due to animal movement towards or away from the second observer, while the CCR method does not, and instead assumes these times to be equal to the lag time between the observers. We anticipate that this will cause the expected values of the two estimators to diverge for long lags, or for the case where animal speeds are non-negligible relative to observer speeds, and this may cause the CCR estimator of density to become biased. Here, however, with the speed of the observer being some 50 times greater than the

mean speed of animals, and the standard deviation of the difference of encounter times from lag time between observers being only 2.4% of the lag time, the effect on the CCR estimator is very small.

7.2 *Short lag scenarios*

Here we investigate the bias and confidence interval coverage of the LCE density estimator under short lag scenarios, and compare these with those obtained from the CCR estimator of Stevenson et al. (2018). There are 36 simulation scenarios corresponding to all combinations of lag $l \in \{10, 20, 50, 80\}$ seconds, $\gamma \in \{0.1, 0.2, 0.5, 0.8\}$, and $\sigma \in \{1.5, 0.95, 0.5\}$ m/s. Simulation results are summarized in Table 2. Boxplots of the density estimates for each of the 36 scenarios are shown in Figure 3.

RMF - Table 1 has numerous strange rounding errors for γ, e.g. 0.21 instead of 0.2, etc. Also the values of speed include 0.65 which differs from the values specified above. The γ and lag columns only need 1 decimal place for γ and don't need decimal places for lag.

[Figure 2 about here.]

[Figure 3 about here.]

[Table 1 about here.]

[Table 2 about here.]

The LCE density estimator is unbiased or very nearly unbiased

Need to update after lots of simulations. BCS: Still seems to be the case.

in all 36 scenarios. Figure 5 shows the empirical bias as a function of the mean number of detections by each observer, together with the empirical bias of the CCR estimator fitted to the same simulated data. The bias of the two estimators is very similar. The correlation between the two estimators varies from 0.888 to 0.995across the 36 scenarios, and the mean

BCS: Changed from 0.92–0.998

difference of the estimator means from the true density, as a percentage of the true density, is 1.08% in the case of the LCE estimator, and −0.1%in the case of the CCR estimator.

BCS: Changed from 0.75 and -0.20%

[Figure 4 about here.]

[Figure 5 about here.]

The coefficients of variation of the LCE and CCR density estimators for all 36 scenarios are shown in Figure 7. The CVs decrease with sample size, as expected. The CV of the LCE estimator is smaller than that of the CCR estimator by a small percentage of the LCE estimator CV, with the percentage improvement due to the LCE method increasing as sample size increases

BCS: No longer true in general. With the smallest sample sizes CCR is more precise.

. We interpret this to be a consequnce of the fact that the CCR estimator is not a maximum likelihood estimator, being based instead on an approximation to the Palm likelihood of pairwise comparisons between detections, so it does not have the asymptotic efficiency of a maximum likelihood estimator

BCS: But I think this is still a sensible conclusion.

. Nevertheless, the difference in precision of the two estimators is very small, with the difference being no more than about 8% relative to a CV of about 9% for the largest sample size considered, which has about 270 detections by each observer

BCS: I haven't changed these values because I was not sure how they were calculated.

DLB: I think the sentence meant abs(CV_{CCR}-CV_{LCE})/CV_{LCE}=0.08 for the largest sample size scenario, and CV_{LCE}=0.09 for this scenario. That statement is consistent with the new Figure 6, although strangely not with the old Figure 6. Maybe I had a preview of the new Figure 6 when I wrote it? To simplify, we could end the sentence after "very small". Figure 6 tells the story better.

.

[Figure 6 about here.]

[Figure 7 about here.]

Coverage probabilities at the 95% level for the LCE estimator are shown in Table 2. In all cases coverage probability is close to the nominal value.

Coverage tends to be a bit high - suspend judgement until lots of simulations done.

RMF - my best guess for coverage is that it's due to summing the likelihoods over all the ω pairings, without taking account of the different probabilities of each pairing. So I don't think it is strictly an MLE, but I can live with not drawing attention to this. Intuitively I think this will overweight the influence of less-plausible pairings with higher σ, and therefore estimate confidence intervals that are wider than they should be.

BCS: With the new simulation results, coverage looks pretty good! So I guess we don't have to bother thinking about all this :)

We conclude from this simulation study that, for these shorter-lag scenarios, the LCE estimator is approximately unbiased, with close to nominal confidence interval coverage. The performance of the LCE and CCR estimators is very similar, with the LCE estimator making a slight gain in precision.

8. Discussion

It might be sensible to include here some words on other latent capture history models? Rachel, you know that literature best I think, having developed one of the methods yourself?

Burt et al. (2014) note that most MRDS models do not allow for animals to be at different distances from different observers, and say "A more satisfactory approach would be to develop models that incorporate movement, but this is not straightforward and remains to be done.". We have done that here, although our detection function is customized for digital aerial

surveys, in which observers detect all animals that are at or close to the surface within distance w of the transect line at the time of coming abeam of the observer.

If capture histories *are* known, as is assumed with MRDS models, the LCE model provides a framework for extending MRDS methods to incorporate animal movement and allow for the fact that the two observers on MRDS surveys very commonly observe the same animals at different perpendicular distances. A complication when such an extension with human observers is that, unlike cameras on aircraft, human observers typically have a wide range of forward distances in view at once, so that animals may be within detection range for very much more than an instant. In this case, one would need to model the observers' detection hazard functions rather than their detection probability functions, as has been done by Langrock et al. (2013); Borchers et al. (2013); Borchers and Langrock (2015); Borchers and Cox (2016), for example.

As noted by Stevenson et al. (2018), if capture histories are known the CCR model can be viewed as a kind of SCR area search model, and the same is true for the LCE model. Looking at Figure 1, one might be tempted to think of the LCE model as a kind of M_b SCR model, in which capture probability is elevated by first capture and then decays slowly over time. The LCE model is not quite that though, because capture probability for the second observer changes with time since an animal was available for detection by the first observer, whether or not the first observer detected it.

The LCE method has some advantages over the CCR method, but it does not scale well as density increases. While we were able to deal with moderately large sample sizes above, this is because density is low enough that the transect line can be divided into many segments with relatively few possible combinations of capture histories within each segment. The number of possible capture histories increases very rapidly when each observer detects more than a few animals within a segment, and the LCE method computation will be infeasible in this

case. The number of possible capture histories is

$$N_{CH} = \sum_{m=0}^{n_2^*} \binom{n_2^*}{m} {}^{n_1^*}P_m, \tag{16}$$

where n_1^* is the larger of n_1 and n_2, and n_2^* is the smaller of them. For $n_1 = n_2 = 2, \ldots 10$, N_{CH} is 7; 34; 209; 1,546; 13,327; 130,922; 1,441,729; 17,572,114; 234,662,231. We speculate that the LCE estimation method will become too slow to be practically useful on typical desktop computers when the number of detections by each observer within a segment is greater than about 8 or 9. The CCR method, by contrast, scales well and is able to deal with much larger numbers of detections within segments.

Being a maximum likelihood method, the LCE method has the advantage of being able to use the extensive inference results and machinery associated with maximum likelihood estimators, including likelihood-based model selection criteria such as AIC, asymptotic efficiency, and associated interval estimation methods. The CCR estimator requires interval estimation by bootstrap, is slightly less efficient than the LCE estimator and cannot take advantage of likelihood-based model selection methods. It is also not able to accommodate varying times between encounters of animals due to animal movement, although in the scenarios we considered this has negligible effect. Finally, the LCE mthod provides an inference framework that allows inclusion of covariates in all parameters mentioned in Section 4.2. (While covariates were not available for our application, we anticipate that they can be collected on future surveys.) It is not clear how easy it will be to include covariates that change continuously along the transect line in the CCR estimation framework, although it should easily accommodate covariates that are constant within, but different along, sections of transect.

As was shown by Stevenson et al. (2018), it is not possible to estimate all parameters of interest from two-observer data with unobserved capure histories. However, varying the lag, or having more than two cameras operating may make one more parameter identifiable. It

is possible that incorporating covariates into some of the parameters will also make more parameters identifiable. This is an area worth exploring in future.

We anticipate that the framework provided by the LCE and CCR methods may facilitate substantial reduction in the cost of processing double-observer data by allowing the estimation process to be automated. To do this, we need only an adequate automatic identifier of the species in question in the video stream from each "observer" separately – we do not need recaptures to be identified. If false negatives affect only the availability process (e.g. remove animals underwater and partially visible), the only cost of using a strict identification criterion in order to avoid false positives, is reduced sample size. If, however, false negatives affect the detection process (e.g. remove some individuals because although they were as available as possible, a wave broke over them as the observer passed) then the assumption of $p = 1$ may be violated and bias may ensue if lag between observers is short. Surveying at different lags may allow estimation of p (in additon to D, σ and κ) so that it will likely be possible to automate inference from digital surveys by using automated object identification criteria that are sufficiently strict so as to reduce the probability of false positives to virtually zero, providing that the survey involves effort and detections at more than one lag.

Received ?? 2018. *Revised* 20??

Appendix A. Derivation of $f_T(t)$

RMF - I've made several edits to tidy up the derivation here.

Define the time and forward coordinate at which observer 1 passes over an animal to be 0. The animal's forward coordinate at time t is σW_t, where W_t is a one-dimensional Brownian motion. The forward coordinate of observer 2 at time t is $-vl + vt$. The time at which

observer 2 passes over the animal is therefore the minimum t such that

$$\begin{aligned} -vl + vt &= \sigma W_t \\ \Rightarrow \quad \frac{vt}{\sigma} - W_t &= \frac{vl}{\sigma}\,. \end{aligned} \tag{A.1}$$

The passage time for observer 2 is therefore $T = \inf\{t : vt/\sigma + B_t = vl/\sigma\}$, where $B_t = -W_t$ is also a Brownian motion. We now use the following standard result for the first passage time of a Brownian motion with drift. Suppose a particle follows Brownian motion with drift parameter c, such that its location at time t is $X_t = ct + B_t$. The random variable $T = \inf\{t : X_t = a\}$ is the first passage time to location a, which has probability density function

$$f_T(t) = \frac{a \exp\left\{\frac{-(a-ct)^2}{2t}\right\}}{\sqrt{2\pi t^3}}\,. \tag{A.2}$$

Substituting $c = v/\sigma$ and $a = vl/\sigma$, we obtain the probability density of the time T at which observer 2 passes over the animal:

$$f_T(t) = \frac{vl \exp\left\{\frac{-v^2(l-t)^2}{2\sigma^2 t}\right\}}{\sqrt{2\pi\sigma^2 t^3}}\,. \tag{A.3}$$

Appendix B. The relationship between σ_{palm}, σ and mean animal speed

The σ of Stevenson et al. (2018), which we call σ_{palm} here, is based on the displacement of animals from the midpoint of their two locations after time l has elapsed, being normally distributed with mean zero variance equal to σ^2_{palm}. If we let the signed distance between the first and second location be Y, then $Y/2 \sim N(0, \sigma^2_{palm})$ and hence $\sqrt{\{Y/(2\sigma_{palm})\}^2} = |Y|/(2\sigma_{palm}) \sim \chi(1)$. Using the fact that the expected value of a $\chi(1)$ random variable is $\sqrt{2}/\Gamma(0.5)$, we have that $E\left\{|Y|/(2\sigma_{palm})\right\} = \sqrt{2}/\Gamma(0.5)$, and hence $2\sigma_{palm} = E(|Y|)\Gamma(0.5)/\sqrt{2}$.

The distance Y between the initial location and the location after l seconds, of an animal following Brownian motion with rate parameter σ, has distribution $Y \sim N(0, \sigma^2 l)$, so that $E\left\{|Y|/(\sigma\sqrt{l})\right\} = \sqrt{2}/\Gamma(0.5)$ and $\sigma\sqrt{l} = E(|Y|)\Gamma(0.5)/\sqrt{2}$, and hence $\sigma = 2\sigma_{palm}/\sqrt{l}$.

As the average speed of an animal over a period of l seconds is $E(|Y|)/l$, the average speed over l seconds of an animal following Brownian motion with rate parameter σ can be written as $\sigma\sqrt{2}/\{\Gamma(0.5)\sqrt{l}\}$.

RMF - removed $E(v)$ above as we use v for the constant speed of the observers. Corrected bracket convention (here and throughout) to $[\{()\}]$.

Acknowledgements

This work was part-funded by the Royal Society of New Zealand through Marsden grant 14-UOA-155.

(TBC):

- David's IAA gant from St Andrews
- David's Leverhulme grant
- Ben: anything?
- Pete: anything?

Stephen Marsland contributed substantially to obtaining the correct expression for $f_T(t)$.

References

Borchers, D. L. and Cox, M. J. (2016). Distance sampling detection functions: 2d or not 2d? *Biometrics* **73,** 593–602.

Borchers, D. L., Laake, J. L., Southwell, C., and Paxton, C. G. M. (2006). Accommodating unmodelled heterogeneity in double-observer distance sampling surveys. *Biometrics* pages 372–378.

Borchers, D. L. and Langrock, R. (2015). Double-observer line transect surveys with markov-modulated poisson process models for overdispersed animal availability. *Biometrics* **71,** 1060–1069.

Borchers, D. L., Zucchini, W., Heide-Jørgenssen, M. P., Canadas, A., and Langrock, R. (2013). Using hidden markov models to deal with availability bias on line transect surveys. *Biometrics* **69,** 703–713.

Buckland, S., Laake, J., and Borchers, D. (2010). Double-observer line transect methods: levels of independence. *Biometrics* **66,** 169–177.

Burt, M. L., Borchers, D. L., Jenkins, K., and Marques, T. A. (2014). Using mark-recapture distance sampling methods on line transect surveys. *Methods in Ecology and Evolution* **5,** 1180–1191.

Hiby, L. and Lovell, P. (1998). Using aircraft in tandem formation to estimate abundance of harbour porpoise. *Biometrics* **54,** 1280–1289.

Laake, J. (1999). Distance sampling with independent observers: Reducing bias from heterogeneity by weakening the conditional independence assumption. In Amstrup, G., Garner, S., Laake, J., Manly, B., McDonald, L., and Robertson, D., editors, *Marine mammal survey and assessment methods.*, pages 137–148, Rotterdam. Balkema.

Langrock, R., Borchers, D. L., and Skaug, H. (2013). Markov-modulated nonhomogeneous poisson processes for unbiased estimation of marine mammal abundance. *Journal of the American Statistical Association* **108,** 840–851.

Russell, S. and Norvig, P. (2010). *Artificial Intelligence: A Modern Approach, 3rd Edition.* Pearson.

Stevenson, B. C., Borchers, D. L., and Fewster, R. M. (2018). Cluster capture-recapture to account for identification uncertainty on aerial suveys of animal populations. *Biometrics* **??,** ???–???

Westgate, A. J., Read, A., Beggren, P., Koopman, H. N., and Gaskin, D. E. (1995). Diving behaviour of harbour porpoises, *Phocoena phocoena*. *Canadian Journal of Fisheries and Aquatic Sciences* **52,** 1064–1073.

Zucchini, W., MacDonald, I. L., and Langrock, R. (2016). *Hidden Markov Models for Time Series: An Introduction Using R, Second Edition*. Chapman and Hall/CRC.

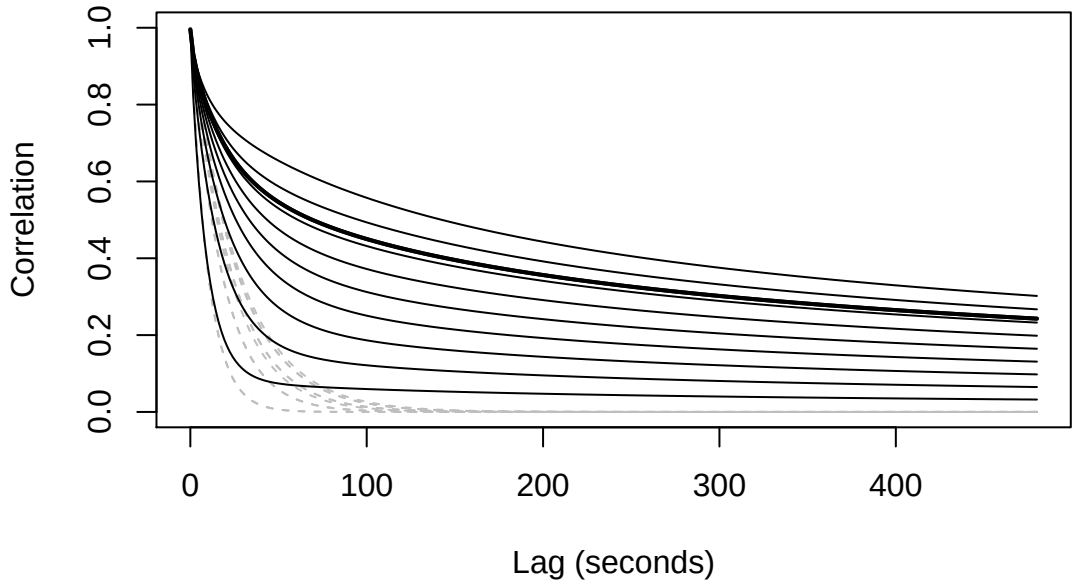

Figure 1. Correlation between detections by the two observers as a function of lag for mean proportions of time available $\gamma = 10\%$ (bottom black line), $20\%, 30\%, 40\%, 50\%, 60\%, 70\%, 80\%, 90\%$ (top black line). The thick black line is for $\gamma = 73\%$. The grey dashed lines show the correlation under the assumption of no animal movement.

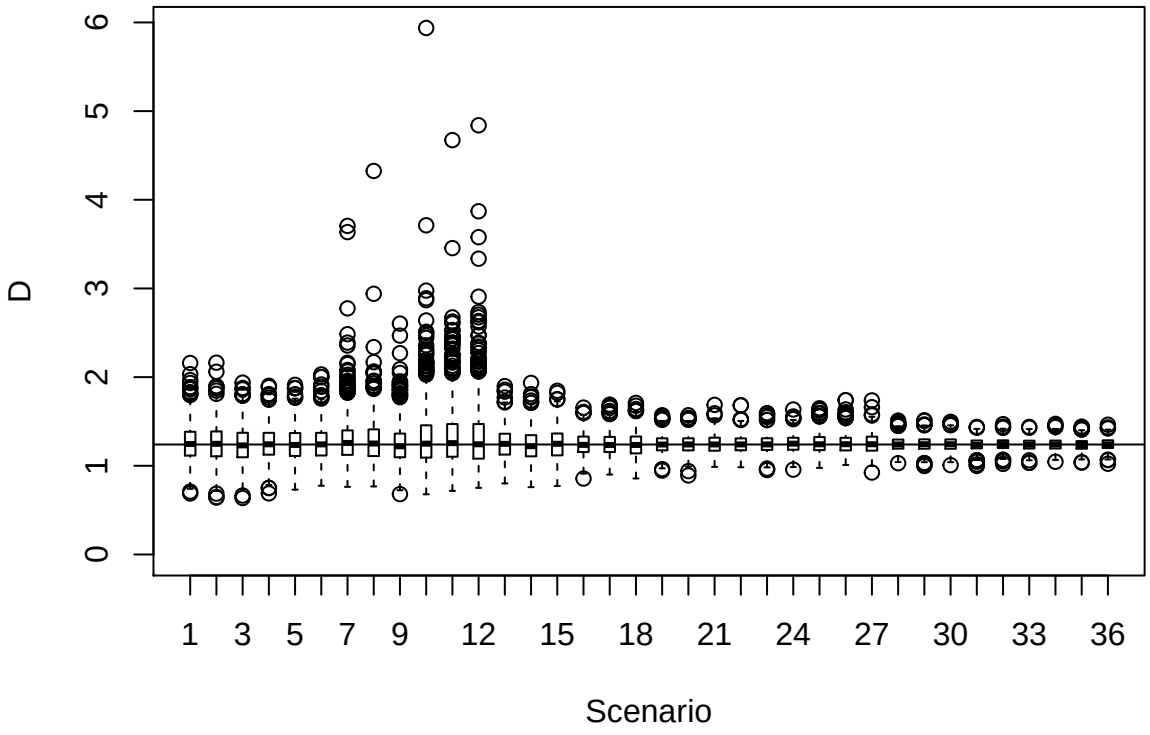

Figure 2. Box plots of estimated density for each of the 36 scenarios. The horizontal line is at true density $D = 1.24$.

BCS: Looks ugly because all of the 'outliers' due to having 1000 iterations. Probably need to make this look nicer...

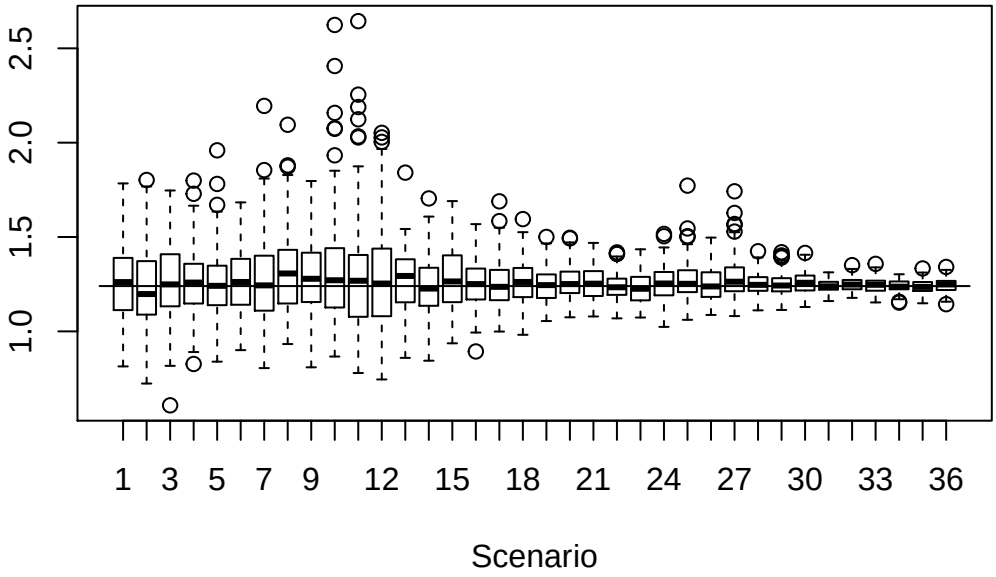

Figure 3. Box plots of estimated density for each of the 36 scenarios. The horizontal line is at true density $D = 1.24$.

BCS: OLD VERSION for comparison.

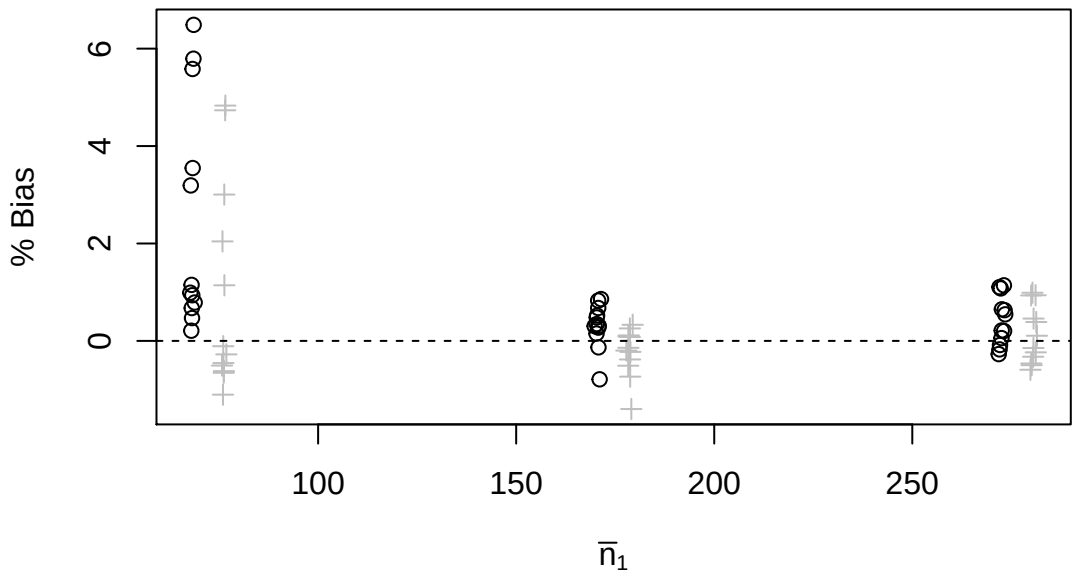

Figure 4. Percentage difference of estimated density from true density, as a function of mean number of detections by a single observer.The LCE estimator is represented by circles, the CCR estimator by crosses. Crosses are offset 8 points to the right, to avoid overlap with circles.

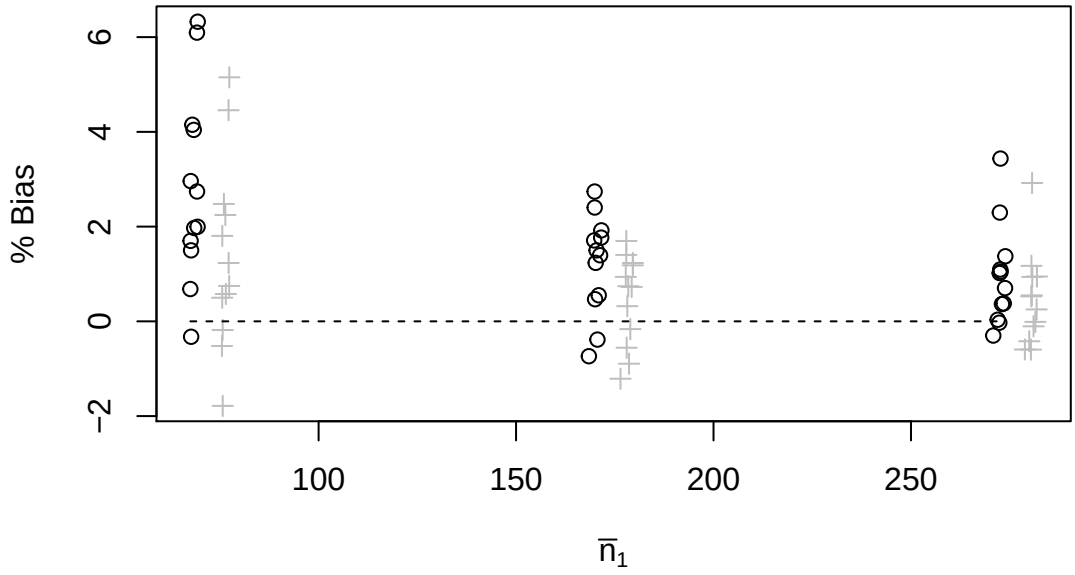

Figure 5. Percentage difference of estimated density from true density, as a function of mean number of detections by a single observer.The LCE estimator is represented by circles, the CCR estimator by crosses. Crosses are offset 8 points to the right, to avoid overlap with circles.

BCS: OLD VERSION for comparison.

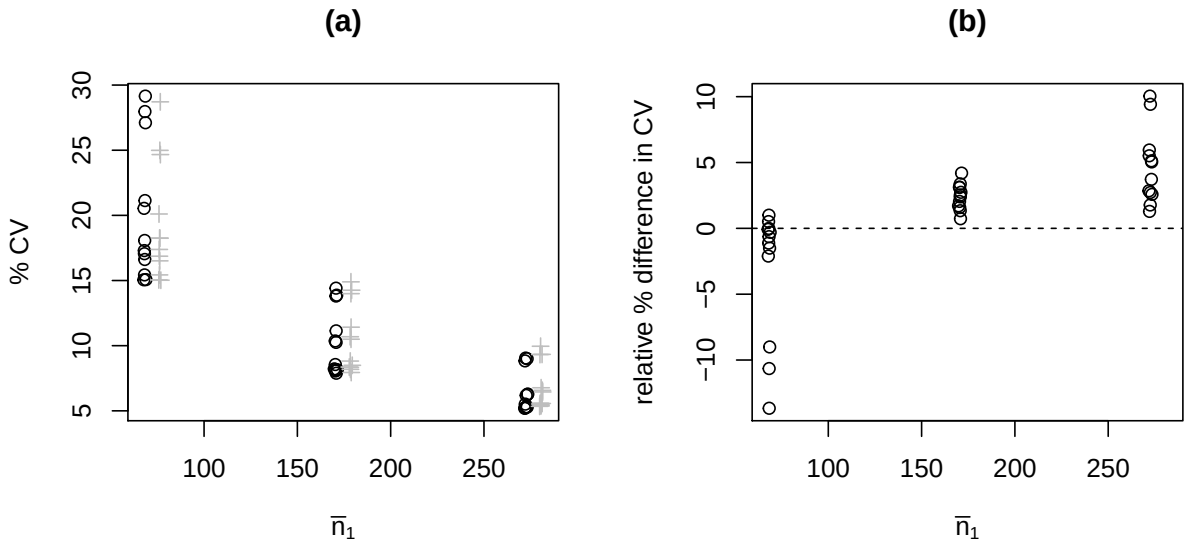

Figure 6. Percentage coefficient of variation (%CV), as a function of mean number of detections by a single observer. The LCE estimator is represented by circles, the CCR estimator by crosses. Crosses are offset 8 units to the right, to avoid overlap with circles. Panel (a) shows the %CV. Panel (b) shows the amount by which the CV from the CCR method exceeds that from the LCE method, expressed as a percentage of the LCE CV.

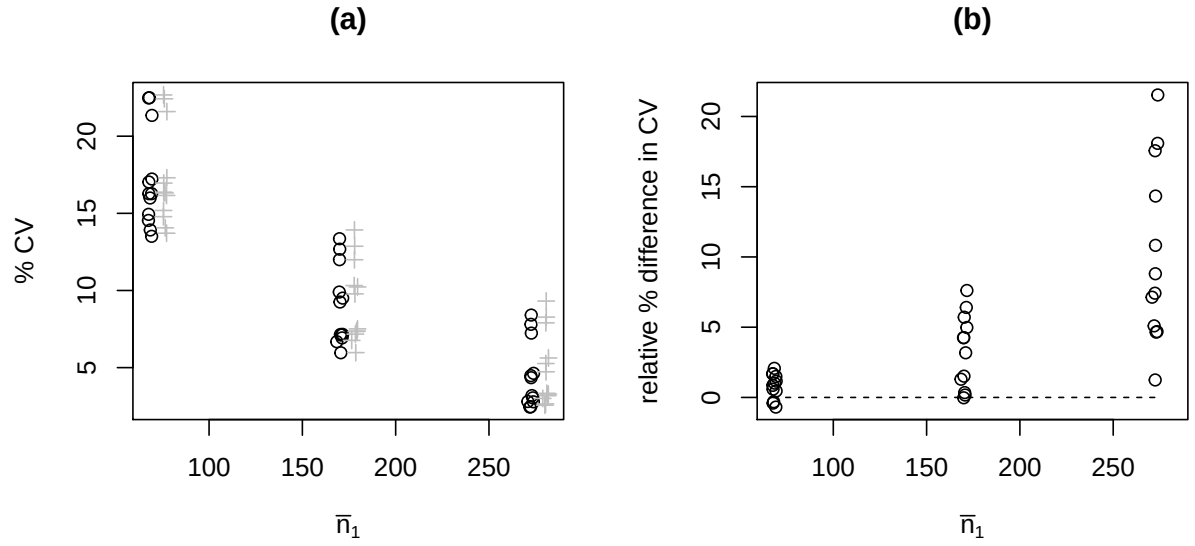

Figure 7. Percentage coefficient of variation (%CV), as a function of mean number of detections by a single observer. The LCE estimator is represented by circles, the CCR estimator by crosses. Crosses are offset 8 units to the right, to avoid overlap with circles. Panel (a) shows the %CV. Panel (b) shows the amount by which the CV from the CCR method exceeds that from the LCE method, expressed as a percentage of the LCE CV.

BCS: OLD VERSION for comparison.

Table 1

Full sims with Nsim = 1000. The columns mean(n) and mean(m) are the averages from my simulations. Is that correct, or are they supposed to be calculated analytically?

	gamma	lag	sigma	%BiasLCE	%cvLCE	%CoverLCE	%BiasPalm	%cvPalm	mean(n)	mean(m)
1	0.20	10.00	0.65	0.94	16.61	0.95	-0.62	16.51	68.23	44.39
2	0.20	10.00	0.95	1.15	17.06	0.96	-0.11	16.87	68.05	43.88
3	0.20	10.00	1.50	0.21	17.30	0.95	-1.10	17.39	67.98	43.85
4	0.20	20.00	0.65	0.99	15.06	0.96	-0.51	15.05	67.74	30.67
5	0.20	20.00	0.95	0.79	15.07	0.96	-0.28	15.02	68.84	31.00
6	0.20	20.00	1.50	0.67	15.43	0.96	-0.45	15.43	68.10	30.56
7	0.20	50.00	0.65	3.55	21.14	0.96	1.14	18.25	68.33	16.68
8	0.20	50.00	0.95	3.19	20.54	0.96	2.04	20.11	67.85	16.38
9	0.20	50.00	1.50	0.47	18.07	0.95	-0.65	18.25	68.17	16.66
10	0.20	80.00	0.65	5.58	27.97	0.96	3.00	24.99	68.31	14.14
11	0.20	80.00	0.95	6.49	27.11	0.96	4.83	24.67	68.58	13.76
12	0.20	80.00	1.50	5.79	29.15	0.95	4.73	28.71	68.53	13.70
13	0.50	10.00	0.65	0.83	13.81	0.95	0.26	13.99	170.70	143.78
14	0.50	10.00	0.95	-0.79	13.88	0.95	-1.40	14.26	171.07	143.97
15	0.50	10.00	1.50	0.68	14.41	0.94	0.08	14.90	170.71	142.57
16	0.50	20.00	0.65	0.27	10.25	0.96	-0.38	10.50	170.74	125.56
17	0.50	20.00	0.95	0.15	10.36	0.96	-0.51	10.69	170.29	124.59
18	0.50	20.00	1.50	-0.13	11.14	0.94	-0.73	11.42	170.78	124.12
19	0.50	50.00	0.65	0.35	8.17	0.96	-0.14	8.34	170.33	97.85
20	0.50	50.00	0.95	0.30	8.22	0.96	-0.20	8.36	169.81	96.64
21	0.50	50.00	1.50	0.48	8.55	0.96	0.09	8.82	170.34	95.50
22	0.50	80.00	0.65	0.31	7.89	0.96	-0.23	7.94	170.84	88.42
23	0.50	80.00	0.95	0.52	8.06	0.95	0.11	8.18	170.43	87.16
24	0.50	80.00	1.50	0.86	8.15	0.95	0.33	8.49	171.42	86.45
25	0.80	10.00	0.65	1.14	9.00	0.95	0.94	9.33	273.13	247.68
26	0.80	10.00	0.95	1.10	8.83	0.96	0.93	9.32	272.00	246.19
27	0.80	10.00	1.50	1.08	9.05	0.96	0.99	9.96	272.34	244.71
28	0.80	20.00	0.65	0.55	6.29	0.96	0.11	6.45	273.47	234.23
29	0.80	20.00	0.95	0.63	6.26	0.97	0.39	6.57	273.30	232.95
30	0.80	20.00	1.50	0.65	6.19	0.98	0.46	6.77	272.62	230.38
31	0.80	50.00	0.65	-0.08	5.51	0.96	-0.50	5.59	272.14	217.32
32	0.80	50.00	0.95	0.21	5.30	0.96	-0.15	5.45	272.58	216.46
33	0.80	50.00	1.50	-0.17	5.21	0.97	-0.46	5.51	272.04	213.68
34	0.80	80.00	0.65	0.06	5.28	0.97	-0.33	5.37	272.52	214.10
35	0.80	80.00	0.95	-0.27	5.20	0.97	-0.59	5.35	271.82	211.78
36	0.80	80.00	1.50	0.20	5.29	0.96	-0.23	5.56	273.18	209.54

Table 2

OLD VERSION: Preliminay sims with Nsim=150 (n,m are number obs by 1 observer, and number recaptures). The first column is scenario number.

	gamma	lag	sigma	%BiasLCE	%cvLCE	%CoverLCE	%BiasPalm	%cvPalm	mean(n)	mean(m)
1	0.20	10.00	0.65	1.50	16.27	0.96	-0.18	16.37	67.71	43.87
2	0.20	10.00	0.95	-0.33	17.03	0.95	-1.78	16.96	67.73	43.70
3	0.20	10.00	1.50	1.70	14.93	0.97	0.50	15.19	67.59	43.69
4	0.20	20.00	0.65	1.97	13.92	0.97	0.58	14.06	68.51	30.67
5	0.20	20.00	0.95	0.68	14.54	0.95	-0.52	14.78	67.53	30.41
6	0.20	20.00	1.50	2.74	13.51	0.97	1.23	13.71	69.22	30.71
7	0.20	50.00	0.65	2.00	17.22	0.97	0.75	17.30	69.36	17.07
8	0.20	50.00	0.95	6.10	16.27	0.99	4.46	16.16	69.20	16.13
9	0.20	50.00	1.50	4.04	15.99	0.99	2.25	16.32	68.39	16.09
10	0.20	80.00	0.65	6.33	21.35	0.97	5.15	21.59	69.39	13.98
11	0.20	80.00	0.95	2.96	22.48	0.96	1.80	22.67	67.62	13.75
12	0.20	80.00	1.50	4.15	22.49	0.96	2.48	22.42	68.03	13.66
13	0.50	10.00	0.65	2.74	12.00	0.98	1.40	11.99	169.90	142.49
14	0.50	10.00	0.95	0.47	12.68	0.98	-0.56	12.87	170.01	142.55
15	0.50	10.00	1.50	2.40	13.36	0.99	1.70	13.92	169.91	141.21
16	0.50	20.00	0.65	1.23	9.25	0.97	0.32	9.78	170.17	124.79
17	0.50	20.00	0.95	1.71	9.90	0.97	0.94	10.32	169.78	123.73
18	0.50	20.00	1.50	1.92	9.51	0.99	1.23	10.23	171.60	124.23
19	0.50	50.00	0.65	0.55	7.11	0.99	-0.16	7.34	170.93	97.32
20	0.50	50.00	0.95	1.77	7.16	0.97	1.18	7.51	171.56	96.85
21	0.50	50.00	1.50	1.50	7.15	1.00	0.74	7.17	170.40	94.97
22	0.50	80.00	0.65	-0.38	5.96	1.00	-0.89	5.97	170.61	88.75
23	0.50	80.00	0.95	-0.74	6.68	0.99	-1.21	6.77	168.44	86.32
24	0.50	80.00	1.50	1.40	6.93	0.97	0.73	7.37	171.27	85.65
25	0.80	10.00	0.65	2.30	7.81	0.97	1.17	7.90	272.50	246.65
26	0.80	10.00	0.95	1.05	7.24	0.99	0.94	8.27	272.71	246.63
27	0.80	10.00	1.50	3.44	8.41	0.98	2.92	9.32	272.66	244.53
28	0.80	20.00	0.65	1.09	4.35	1.00	0.55	4.73	272.59	232.97
29	0.80	20.00	0.95	1.02	4.48	1.00	0.53	5.27	272.46	232.08
30	0.80	20.00	1.50	1.37	4.63	1.00	0.95	5.63	273.89	231.49
31	0.80	50.00	0.65	0.03	2.44	1.00	-0.42	2.57	271.95	217.55
32	0.80	50.00	0.95	0.70	2.78	1.00	0.25	3.29	273.81	217.29
33	0.80	50.00	1.50	0.37	3.19	1.00	-0.10	3.34	273.03	214.03
34	0.80	80.00	0.65	-0.03	2.46	1.00	-0.60	2.65	272.37	214.35
35	0.80	80.00	0.95	-0.30	2.80	1.00	-0.59	3.00	270.83	210.97
36	0.80	80.00	1.50	0.38	3.04	1.00	-0.01	3.18	273.49	210.53