Supporting Information for
A latent capture history model for digital aerial surveys

by

D. L. Borchers, P. Nightingale, B. C. Stevenson, and R. M. Fewster.

1. Web Appendix A: Derivation of $f_T(t)$

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

$$-vl + vt \;=\; \sigma W_t \quad \Rightarrow \quad \frac{vt}{\sigma} - W_t \;=\; \frac{vl}{\sigma}\,. \tag{1}$$

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. Now if a particle follows Brownian motion with drift parameter c, such that its location at time t is $X_t = ct + B_t$, then the random variable $T = \inf\{t : X_t = a\}$ is the first passage time to location a, and has probability density function $f_T(t) = a \exp\left\{\frac{-(a-ct)^2}{2t}\right\}/(\sqrt{2\pi t^3})$. 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 as Eqn (**??**).

2. Web Appendix B: 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 $\mathcal{P} = \langle \mathcal{X}, \mathcal{D}, \mathcal{C} \rangle$. The CSP $\mathcal{P}$ has a set of decision variables $\mathcal{X}$, each of which has a set of possible values that it may take, called its *domain*, where $\mathcal{D}(x)$ is the domain of $x \in \mathcal{X}$. 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) $\subseteq \mathcal{X}$. 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 $x_{i,j}$ with domain $\{0, 1\}$. Variable $x_{i,j}$ is equal to 1 in a solution if and only if the two observations are paired. $\mathcal{X}$ is the set of all such decision variables $x_{i,j}$. $\mathcal{D}$ is the function $\{(x_{i,j}, \{0, 1\}) \mid x_{i,j} \in \mathcal{X}\}$.

Suppose we have two distinct sets, $s_1 = \{i, j\}$ and $s_2 = \{k, l\}$, where i may be paired with j, and k may be paired with l, but the two sets are not disjoint: in other words $s_1 \cap s_2 \neq \emptyset$. In all such cases we add the constraint $c_{s_1,s_2} = (x_{i,j} = 0 \vee x_{k,l} = 0)$ to prevent such pairing. The set $\mathcal{C}$ contains all such c_{s_1,s_2} and no other constraints. All components of the CSP $\mathcal{P} = \langle \mathcal{X}, \mathcal{D}, \mathcal{C} \rangle$ have now been defined.

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

3. Web Appendix C: The relationship between σ_{palm}, σ and mean animal speed

The σ of Stevenson et al. (2019), 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, which is normally distributed with mean zero and 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 $\sigma\sqrt{2}/\{\Gamma(0.5)\sqrt{l}\}$.

4. Web Appendix C: Code to reproduce results of the paper

Source code for an `R` package called `LCE_paper` to fit LCE models, and code to fit to the porpoise data and to conduct the simulations described in the paper, is available here: `https://github.com/david-borchers/LCE_paper`. The package and code is also available at the Biometrics website on Wiley Online Library.

References

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

Stevenson, B. C., Borchers, D. L., and Fewster, R. M. (2019). Cluster capture-recapture to account for identification uncertainty on aerial surveys of animal populations. *Biometrics* **75,** 326–336.