Biometric Practice

Cluster capture-recapture to account for identification uncertainty on aerial surveys of animal populations[†]

Ben C. Stevenson[1,2#,*], David L. Borchers[1], and Rachel M. Fewster[2]

[1]School of Mathematics and Statistics, University of St Andrews, St Andrews, Fife KY16 9LZ, United Kingdom

[2]Department of Statistics, University of Auckland, Auckland 1010, New Zealand

#Present address
*email: ben.stevenson@auckland.ac.nz

Additional Supporting Information may be found in the online version of this article.

Summary:

Capture-recapture methods for estimating wildlife population sizes almost always require their users to identify every detected animal. Many modern-day wildlife surveys detect animals without physical capture-visual detection by cameras is one such example. However, for every pair of detections, the surveyor faces a decision that is often fraught with uncertainty: are they linked to the same individual? An inability to resolve every such decision to a high degree of certainty prevents the use of standard capture-recapture methods, impeding the estimation of animal density. Here we develop an estimator for aerial surveys, on which two planes or unmanned vehicles (drones) fly a transect over the survey region, detecting individuals via high-definition cameras. Identities remain unknown, so one cannot discern if two detections match to the same animal; however, detections in close proximity are more likely to match. By modeling detection locations as a clustered point process, we extend recently developed methodology and propose a precise and computationally efficient estimator of animal density that does not require individual identification. We illustrate the method with an aerial survey of porpoise, on which cameras detect individuals at the surface of the sea, and we need to take account of the fact that they are not always at the surface. This article is protected by copyright. All rights reserved

1. Introduction

Capture-recapture (CR) methods (Seber, 1982) are widely used to estimate the size of animal populations. The key data for these methods are the capture histories: records of where and when each animal was detected. These data can be obtained without physical capture, for example through audio detections by microphones (e.g., Dawson and Efford, 2009), visual detections by camera traps (e.g., Borchers et al., 2014), or sightings from aircraft (e.g., Hammond et al., 2002, 2013). These indirect methods of capture have great potential for species and habitats that are otherwise difficult to sample, but they introduce the problem that individual identity is uncertain, so compiling detections into individual capture histories is prone to error.

In this paper, we focus on aerial surveys of marine mammals. These surveys are usually conducted using human observers, but we anticipate that they will in future be done using high definition video cameras instead, possibly mounted on drones rather than in piloted aircraft. Collecting CR data from aerial surveys requires at least two observers, each acting as a 'capture occasion', to account for imperfect detection of animals. Marine mammal surveys have the particular difficulty that animals are unavailable for detection when they are diving. This source of detection failure can only be adjusted for if the two observers, or cameras, are separated in time by some designated lag. Otherwise, if diving animals are always unavailable to both observers, the diving proportion of the population is effectively unsampled and there is no information from which to estimate the size of this proportion.

If the two observers are separated by a short time lag, dependence between them is strong, and there is little information on which to estimate the availability process. On the other hand, the shorter the time lag, the easier it is to correctly identify recaptures (detections of the same individual by both observers). With long lags, there can be considerable uncertainty about which detections by the two observers correspond to the same animal, because

individuals cannot generally be identified from the air. CR methods typically assume no uncertainty about recaptures. This creates a dilemma for the surveyor: longer lags create better independence between observers, but greater uncertainty in recapture identification.

Uncertainty about recaptures is common when detection is by remote devices or sightings rather than by physical capture. Existing methods to account for uncertain identification on CR surveys typically model the unknown true capture histories, or their frequencies, as latent variables. These approaches either compute a likelihood via a sum over all plausible sets of true capture histories (e.g. Hiby and Lovell, 1998), or sample from these sets as part of a Markov chain Monte Carlo (MCMC) algorithm (e.g., Chandler and Royle, 2012; Link et al., 2010; Tancredi et al., 2013; Wright et al., 2009). However, even for moderately sized data sets, the number of plausible detection-to-identity matchings can be prohibitively large. As a result, computational demands hamper both maximum likelihood and Bayesian methods: the large number of possible matchings can prohibit calculation of the likelihood function, and MCMC algorithms involve long computation times.

One recently proposed method (Fewster et al., 2016), however, does not reconstruct capture histories. This approach treats CR data with uncertain identification as a clustered point pattern. The survey measures some variable from each detection, chosen so that observed values across multiple detections of the same individual are more similar than observed values for different individuals. Observations from the same animal therefore form a 'cluster'; a cluster here corresponds to detections of the same individual, not a cluster of different individuals. The approach of Fewster et al. (2016) then applies a Poisson cluster process estimation method (Tanaka et al., 2008) to detection data. The parameters of the process can relate to those of interest in CR studies, such as animal density. This method considers the discrepancies between all pairs of detections, rather than summing over or sampling from plausible matchings, providing a computationally efficient estimator. Fewster et al. (2016)

used the term 'trace-contrast' for this type of estimation; here we adopt the alternative term 'cluster capture-recapture'.

In this paper, we generalize the methods of Tanaka et al. (2008) and Fewster et al. (2016) to develop a cluster capture-recapture estimator of animal density from aerial surveys using two observers separated by a time lag to account for the unavailability of animals, for example due to diving. We refer to the two observers as 'cameras', as we anticipate this will be a typical survey mode in the future. The data from these surveys consist of two sets of locations of detected individuals, one for each camera. Individuals detected by one camera cannot be matched to those detected by the other by any physical markings or identification features. Instead, information about identification is obtained from the fact that locations of detections of the same animal by the two different cameras tend to be more similar than locations of different animals.

We describe the Tanaka et al. (2008) estimation framework in Section 2, generalize this approach in Section 3, and show how it applies to two-camera aerial surveys in Section 4. We apply our method to data in Section 5, and present simulation studies in Section 6.

2. Estimation framework

The framework of Tanaka et al. (2008) estimates parameters of Neyman-Scott point processes (NSPPs; see Baddeley et al., 2015, pp. 459–469). NSPPs generate clustered point patterns. Clusters are formed by 'parents', which are produced by a homogeneous Poisson process with intensity D, the parameter of interest. Parents themselves are not observed, but each independently spawns a cluster of 'children', the number of which comes from some distribution identical across parents. In this section, we consider processes where the number of children is Poisson with expectation ν. The children are independently scattered around their parents—it is these locations that are observed. Specific names for NSPPs indicate the distribution of the children around their parents; for example, two-dimensional Thomas

processes have locations of children given by a bivariate normal distribution centered on their parents, with variance matrix $\sigma^2 \boldsymbol{I}_2$; see Figure 1(i) for an example. Conceptually, NSPPs may adequately describe spatial locations of apples around the trees from which they fell (Fewster et al., 2016), locations of plants around their ancestors due to seed dispersal (Baddeley et al., 2015, p. 469), and galaxy clusters in the universe (Neyman and Scott, 1952).

Parameter estimation for NSPPs is notoriously difficult. Due to the parent process remaining a latent feature, it is not known how many parent points there are, where the parent points are located, or which groups of children are 'siblings' (i.e., observed points that share a common parent, such as two apples that fell from the same tree). These must be modeled as latent variables, and so, under a maximum-likelihood framework, evaluation of the likelihood function must consist of summing over all possible numbers of parents, integrating over the possible locations of each parent, and summing over all possible matchings of children to parents. Even with small sample sizes, the number of possible children-to-parent matchings is extremely large, and so the likelihood is widely held to be intractable (Baudin, 1981; Guan, 2006; Tanaka et al., 2008; Waagepetersen, 2007). Adopting a Bayesian approach may appear more tractable—this involves sampling instead of enumerating numbers of parents, locations of parents, and children-to-parent matchings. However, creating an MCMC scheme that adequately samples over this complex latent structure while achieving acceptable mixing and convergence rates is problematic (Waagepetersen, 2007).

[Figure 1 about here.]

Tanaka et al. (2008) provided an alternative NSPP estimator that is efficient to compute. It relies on the Palm intensity function, $\lambda_0(r)$, defined as the expected intensity of the process at a location distance r from a randomly selected child point. For NSPPs, this is a decreasing function. The intensity is high when r is small, because the presence of a child suggests nearby siblings. The Palm intensity declines to an asymptote as r increases, because if r is

larger than the spread of a cluster, then the Palm intensity is only due to points from other clusters—the local influence of nearby siblings no longer applies (see Figure 2).

The Palm intensity is $\lambda_0(r) = D\nu + \nu/(4\pi\sigma^2)\exp\{-r^2/(4\sigma^2)\}$ for two-dimensional Thomas processes (Tanaka et al., 2008), and is based on the Chi distribution of intersibling distances (Fewster et al., 2016). One can estimate $\boldsymbol{\theta} = (D, \nu, \sigma)$ by finding the parameter values whose Palm intensity best fits the observed data, according to some objective function (Figure 2). Tanaka et al. (2008) proposed the 'Palm likelihood' as an objective function for this purpose.

[Figure 2 about here.]

The Palm likelihood involves constructing the 'difference process' for each of the n observed points by taking the differences between a selected point and all others (Figures 1(ii)–(v)). Comparing a reference point at (x_1, y_1) to another point at (x_2, y_2) provides the point $(x_2 - x_1, y_2 - y_1)$ in its difference process. All n difference processes are then superposed (Figure 1(vii)). The expected intensity of the n superposed difference processes at a location distance r from the origin is the Palm intensity multiplied by n.

The estimator $\hat{\boldsymbol{\theta}}$ is gained by fitting an inhomogeneous Poisson point process with intensity function $n\lambda_0(r)$ to the superposed difference processes in $\mathbb{R}^2$ (see Figure 1(vi)). The Palm likelihood is defined as the likelihood for this inhomogeneous Poisson process. Tanaka et al. (2008) showed that, for two-dimensional Thomas processes, this is

$$L(\boldsymbol{\theta}) = \left\{\prod_{\{i,j:i\neq j, r_{ij}<t\}} n\lambda_0(r_{ij})\right\} \times \exp\left[-n\nu\left\{\pi Dt^2 + 1 - \exp\left(\frac{-t^2}{4\sigma^2}\right)\right\}\right], \qquad (1)$$

where $r_{ij} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$ gives the Euclidean distance between the ith and jth observed child points, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. The constant t is a truncation distance for dealing with edge effects, which we explain below. Numerical maximization provides an estimator for $\boldsymbol{\theta}$, which is computationally efficient to obtain because the Palm likelihood is in closed form.

As with any spatial modeling process, attention must be given to edge effects. There are many ways of dealing with these (see Baddeley et al., 2015, pp. 212–220). Tanaka et al.

(2008) calculated distances subject to periodic boundary conditions and set a truncation distance, t, to the superposed difference processes. Any points further than t from the origin were ignored. For this approach, all edges of a rectangular surveyed area must be at least $2t$ in length, and t should be larger than any plausible distance between two siblings.

An inhomogeneous Poisson point process with intensity $n\lambda_0(r)$ is not the correct model for the superposed difference processes; it is used solely to provide an objective function to maximize, giving an estimator of $\boldsymbol{\theta}$. While this provides straightforward estimation, it is an *ad hoc* model. Nevertheless, Tanaka et al. (2008) showed via simulation that the Palm likelihood method gives estimators with negligible bias, and Prokešová and Jensen (2013) proved that they are consistent and have asymptotic normality. See Baddeley et al. (2015, pp. 292–295, 484–485) for further background on Palm intensity and likelihood.

3. Generalization of the Palm likelihood estimation framework

In principle, Palm likelihood estimation could be applied to any clustered point process, but first it is necessary to derive the Palm intensity and Palm likelihood functions of the process. Tanaka et al. (2008) did so for five Poisson cluster processes, four of which were variants of the Thomas process. All five processes were two-dimensional and used a Poisson distribution for the number of children spawned by each parent. In this section, we derive the Palm intensity and likelihood for NSPPs that are of general dimension d, and for which the number of children spawned by each parent is not necessarily a Poisson random variable.

Additionally, for some applications, we may partially observe sibling relationships, such that it is known with certainty that particular pairs of points are either siblings or nonsiblings. For example, two apples in a field may be known to be nonsiblings if they are of different varieties, but this still leaves many pairs of apples of the same variety that might or might not be siblings. We also extend the methodology to incorporate this information.

Hereafter, functions with the dimensionality of the point process, d, appearing in their

definitions are followed by the superscript $^{(d)}$ to denote this dependence; for example, the Palm intensity function is now given by $\lambda_0^{(d)}(r)$.

3.1 *The Palm intensity and Palm likelihood functions*

In Web Appendices A and B, we derive the Palm intensity for d-dimensional NSPPs as

$$\lambda_0^{(d)}(r) = D\,E(C) + \frac{f_Q^{(d)}(r)E\{C(C-1)\}}{s^{(d)}(r)E(C)}. \tag{2}$$

The random variable Q is the distance between two randomly selected siblings, with probability density function $f_Q^{(d)}(r)$ and cumulative distribution function $F_Q^{(d)}(r)$. Both are fully specified for Thomas processes in Web Appendix C. The random variable C is the number of children generated by a randomly selected parent, with expectation $E(C)$ and variance $V(C)$; it need not be Poisson. From Web Appendix A, $E\{C(C-1)\}/E(C) = \{V(C) + E(C)^2\}/E(C) - 1$, and so the distribution of C only affects the Palm intensity through its expectation and variance. Functions $s^{(d)}(r)$ and $v^{(d)}(r)$, used below, are the boundary volume and the volume of a d-dimensional hypersphere of radius r (Web Appendix B).

The Palm likelihood is derived by treating the superposed difference processes as an inhomogeneous Poisson point process. The intensity of this process is $n\lambda_0^{(d)}(r)$ at a point in $\mathbb{R}^2$ distance r from the origin. This gives our objective function, the Palm likelihood:

$$L^{(d)}(\boldsymbol{\theta}) = \left\{ \prod_{\{i,j:i\neq j;r_{ij}<t\}} n\,\lambda_0^{(d)}(r_{ij}) \right\} \times \exp\left\{ -n\int_0^t \lambda_0^{(d)}(r)s^{(d)}(r)\,\mathrm{d}r \right\}, \tag{3}$$

where $r_{ij} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$ is the Euclidean distance between the ith and the jth child point. As per Tanaka et al. (2008), we calculate distances subject to periodic boundary conditions and truncate them at t to deal with edge effects. The integral is available in closed form:

$$\int_0^t \lambda_0^{(d)}(r)s^{(d)}(r)\,\mathrm{d}r = D\,E(C)v^{(d)}(t) + \frac{E\{C(C-1)\}F_Q^{(d)}(t)}{E(C)}. \tag{4}$$

3.2 *Incorporating known sibling information*

In some survey contexts, additional data may provide information about relationships between some pairs of points. We partition the Palm intensity into separate components

due to (i) comparisons between known siblings, $\lambda_{0s}(r, \alpha)$; (ii) comparisons between known nonsiblings, $\lambda_{0n}(r, \beta)$; and (iii) comparisons between points whose relationship is unknown, $\lambda_{0u}(r, \alpha, \beta)$. These three functions sum to $\lambda_0^{(d)}(r)$. In Web Appendix D, we show that this provides the Palm likelihood function

$$
L^{(d)}(\boldsymbol{\theta}) = \left\{ \prod_{\{i,j; i \neq j, r_{ij} < t, h_{ij}=0\}} n \, \lambda_{0s}(r_{ij}, \alpha) \right\} \times \left\{ \prod_{\{i,j; i \neq j, r_{ij} < t, h_{ij}=1\}} n \, \lambda_{0n}(r_{ij}, \beta) \right\}
$$
$$
\times \left\{ \prod_{\{i,j; i \neq j, r_{ij} < t, h_{ij}=2\}} n \, \lambda_{0u}(r_{ij}, \alpha, \beta) \right\} \times \exp \left\{ -n \int_0^t \lambda_0^{(d)}(r) s^{(d)}(r) \, \mathrm{d}r \right\}, \quad (5)
$$

where α is the probability that two randomly selected siblings are known siblings; β is the probability that two randomly selected nonsiblings are known nonsiblings. Here, $h_{ij} = 0$ if points i and j are known siblings, $h_{ij} = 1$ if they are are known nonsiblings, and $h_{ij} = 2$ if their relationship is unknown. We assume that α and β are known; see Web Appendix D.

3.3 *Variance estimation*

We follow Fewster et al. (2016) and recommend variance estimation via a parametric bootstrap. This is computationally efficient because simulation from NSPPs is straightforward, and our estimation procedure maximises an easily computed function.

4. Cluster capture-recapture for two-camera aerial surveys

In this section we show how data collected on two-camera aerial surveys can be modeled using an NSPP, and that estimation of animal density, D, is possible using the generalizations we have made to the approach of Tanaka et al. (2008) in Section 3.

The cameras monitor the same strip of ocean, the 'detection region', of length k km and width $2w$ km, with some lag in time, l, between them. See Figure 3(i) for a depiction of example simulated data, where the detection region is given by the grey strip.

Let $\boldsymbol{x}_{ij} = (x_{ij}, y_{ij})$ be the location of the ith individual at the passing of the jth camera, where x_{ij} is the distance along the transect and y_{ij} is the perpendicular distance from the

transect's center line. If the second camera passes shortly after the first, then we expect $\boldsymbol{x}_{i1}$ and $\boldsymbol{x}_{i2}$ to be close to one another. We model this as follows: The ith individual is associated with a parent location, $\boldsymbol{p}_i$, representing the central point about which its locations at the passing of the cameras are distributed. We assume that parent locations are a realization of a homogeneous Poisson point process, and that $(\boldsymbol{X}_{ij} \mid \boldsymbol{p}_i) \sim \mathrm{N}_2(\boldsymbol{p}_i, \sigma^2 \boldsymbol{I}_2)$, where $\boldsymbol{X}_{ij}$ is the ith individual's location at the passing of the jth camera.

The survey records the locations of detected animals within the detection region. We consider that an animal's location is observed by a camera at a discrete point in time—it is not possible for a single camera to detect the same individual more than once. For now, we assume that a passing camera always detects an individual that is both within the detection region and on the ocean surface. A passing camera does not detect an individual that is either diving or outside the detection region. We use a continuous-time Markov chain to explain how an individual's dive-states at the passing of the two cameras are related (Section 4.4.1).

[Figure 3 about here.]

4.1 *Link with spatial capture-recapture*

This two-camera survey is similar to a spatial capture-recapture (SCR) area-search survey (Efford, 2011), in which a detection region is searched across a number of occasions (in our case, the two cameras). Individuals' parent locations are analogous to the latent activity centers in SCR, used to model spatial similarity in detection locations. The most obvious distinction is that animal identities are observed on SCR surveys. Chandler and Royle (2012) proposed an SCR method for unidentified animals, but they considered an array of detectors at discrete points, rather than a continuous detection region. Another distinction is that, on a two-camera survey, detections on different occasions are correlated at small time lags because animals tend to stay at the surface for a period of time.

4.2 *Link with Thomas processes*

Detection locations from two-camera aerial surveys share many features with Thomas processes (Section 2; Figure 1). Expected cetacean locations (parents) form a Poisson point pattern, and detection locations (children) are clustered around these, subject to Gaussian dispersion. There are two distinctions, however. First, the number of times each individual is detected (the number of children generated by each parent), C, is not a Poisson random variable; we have $c = 2$ for individuals detected by both cameras, $c = 1$ for individuals detected by only one of the cameras, and $c = 0$ for individuals that evaded detection entirely. Second, we sometimes observe that two children are nonsiblings; detections made by the same camera cannot be of the same individual, even if they are in close proximity (see Figure 3(ii)).

We can therefore estimate cetacean density (the intensity of parents) via maximization of the Palm likelihood we developed in Section 3, Equation (5). To do so, we must specify the expectation and variance of C as a function of the model parameters, and appropriate values for α and β. We do this in Sections 4.4 and 4.5. We first deal with the additional complication of edge effects in Section 4.3.

4.3 *Reduction to one dimension*

Using periodic boundary conditions to account for edge effects requires the truncation distance, t, to be less than half the length of the shortest side of the detection region, $2w$, and to be greater than the largest plausible distance between two siblings (e.g., 5σ for a Thomas process). If animals travel further than the width of the detection region between the passing of the two cameras (e.g., if $5\sigma > 2w$), then it is impossible to fulfil both constraints.

We project the two-dimensional locations onto a one-dimensional point pattern to deal with these edge effects. The remaining dimension represents the distance along the transect at which detections were made (Figure 3(ii)). We set a buffer distance, b, at a value large enough so that any individual with a parent location beyond distance b from the transect cannot

plausibly generate a child within the detection region. We define the 'survey region' as the region containing all locations within distance b of the transect center line (i.e., both the grey and the white regions in Figure 3(i)). The one-dimensional projection of the two-dimensional process enables us to accommodate movement of animals in and out of the detection region, without having to apply periodic boundary conditions in the second dimension.

Setting $d = 1$ and maximizing Equation (3) or (5) provides a density estimate $\hat{D}$ representing the number of parent locations within distance b of the transect per kilometer flown by the aerial vehicles. We can then obtain animal density per square kilometer $\hat{D}_2 = \hat{D}/(2b)$.

If b is large enough so that parent locations outside the survey region will not spawn points in the detection region, then its precise value is inconsequential. Setting b too small causes estimator bias, but setting it too large only increases computation time. The concept of an 'integration buffer width' in the SCR literature is analogous (see Borchers and Efford, 2008).

4.4 *Distribution of C for two-camera aerial surveys*

In our context, the unobserved random variable C is the number of times (0, 1, or 2) we detect a randomly selected individual with a parent location in the survey region. Both the animal's position relative to the detection region and its depth in the water column affect C. For now we assume animals are detected with certainty if they are on or near the ocean surface and within the detection region at the passing of a camera; we discuss the problem of imperfect detection later. In general, the random variable C does not have a binomial distribution due to dependence in each of these events across the two cameras.

Consider a randomly selected animal located within the survey region. Let A_j be an indicator variable for the event that the animal is detected by the jth camera. Let U_j and Z_j be indicators for the animal being on the surface, and for the animal being within the detection region, when the jth camera passes, so $A_j = U_j Z_j$. If diving is not affected by the passing of the cameras, then U_j and $Z_{j'}$ are independent for any choice of $j, j' \in \{1, 2\}$.

However, the events U_1 and U_2 may be dependent, as may Z_1 and Z_2: an individual on the surface and within the detection region when the first camera passes may be likely to remain as such when the second camera passes. These dependencies are discussed below.

4.4.1 *Dive-state dependence: U_1 and U_2.* Continuous-time Markov chains have previously been applied to cetacean diving behavior (e.g., Hiby and Lovell, 1998), and we use this approach here. We introduce two parameters: the mean dive-cycle duration, τ, and the mean duration of the surface phase, κ. The marginal probability of an individual being on the surface at the passing of either camera is $\Pr(U_j) = \kappa/\tau$ for $j \in \{1,2\}$, and the probability that it is on the surface at the passing of the second given it was on the surface for the first is

$$\Pr(U_2 \mid U_1) = \frac{\kappa}{\tau} + \frac{\tau - \kappa}{\tau} \exp\left\{ -\left(\frac{1}{\kappa} + \frac{1}{\tau - \kappa} \right) l \right\}, \tag{6}$$

where l is the lag in time between the two cameras in seconds; see Hiby and Lovell (1998).

4.4.2 *Detection region dependence: Z_1 and Z_2.* In this section we remove the subscript i for clarity, and so $\boldsymbol{x}_j = (x_j, y_j), j \in \{1,2\}$, contains the Cartesian coordinates of a randomly selected individual within the survey region at the passing of the jth camera. Recall that these coordinates are the distance along the transect and the perpendicular distance from the transect at which the animal was located, respectively. For the latter, let sightings to the left of the plane take a negative value and those to the right take a positive value.

Event Z_j occurs if and only if $-w \leqslant y_j \leqslant w$ (that is, if the individual is in the detection region). Let $\boldsymbol{p} = (p_x, p_y)$ be the coordinates of the individual's parent location, where p_y is the perpendicular distance from the transect. As parent locations are assumed to be generated by a homogeneous Poisson point process throughout the survey region, p_y is a uniform random variable with probability density function $f_{P_y}(p_y) = 1/(2b), p_y \in [-b, b]$.

The probability that an individual is in the detection region at the passing of the second

camera, given that it was in the detection region at the passing of the first, is

$$\Pr(Z_2 \mid Z_1) = \Pr(-w \leqslant Y_2 \leqslant w \mid -w \leqslant Y_1 \leqslant w)$$
$$\approx \frac{1}{2w} \int_{-b}^{+b} \Pr(-w \leqslant Y_1 \leqslant w \mid p_y)^2 \, \mathrm{d}p_y \,; \tag{7}$$

see Web Appendix E. The error associated with this approximation approaches 0 as b increases; the error is negligible for $b \geqslant w + 5\sigma$. Likewise, the marginal probability of being within the detection region is $\Pr(Z_j) \approx w/b$ for $j \in \{1, 2\}$, with the error in this approximation approaching 0 as b increases. The integral can be approximated numerically.

4.4.3 *Expectation and variance of C.* The distribution of C only affects the Palm intensity through its expectation and variance (Web Appendix A); we must derive these in terms of the model parameters. The random variable $C = A_1 + A_2$ is a sum of two dependent Bernoulli random variables, one occurring at the passing of each camera. As A_1 and A_2 depend on both diving and location states, C depends on parameters κ, τ, and σ.

The marginal probability that the jth camera detects an individual with a parent location within the survey region is $\Pr(A_j = 1) = \Pr(U_j \cap Z_j) = \Pr(U_j)\Pr(Z_j) = \kappa w/(\tau b)$ for $j \in \{1, 2\}$, due to independence of U_j and Z_j. As A_j is a Bernoulli random variable, this is equivalent to its expectation. This gives $E(C) = E(A_1) + E(A_2) = 2\kappa w/(\tau b)$. For the variance, $V(C) = E(C^2) - E(C)^2$, the first term of which is $E(C^2) = E(A_1^2) + E(A_2^2) + 2\,E(A_1 A_2) = E(A_1) + E(A_2) + 2\,E(A_1 A_2)$, where $E(A_1 A_2) = \Pr(A_1 A_2 = 1) = \Pr(U_1)\Pr(U_2 \mid U_1)\Pr(Z_1)\Pr(Z_2 \mid Z_1)$, each term of which is also provided above.

4.5 *Sibling information on two-camera aerial surveys: α and β*

In our design, β is the probability that two randomly selected nonsibling detections can be identified as nonsiblings because the two detections were made by the same camera. A randomly selected detection is equally likely to have been by either camera. Half of all randomly chosen nonsibling pairs of detections come from the same camera, in the same

way that half of all pairs of coin flips have the same outcome, and so $\beta = 0.5$. We can never conclude with certainty that detections from different cameras are of the same individual. Hence $\alpha = 0$; we will never identify siblings as siblings.

We estimate $\boldsymbol{\theta}$ via maximization of Equation (5), setting $\alpha = 0$, $\beta = 0.5$, $h_{ij} = 1$ if the ith and jth detections were from the same camera, and $h_{ij} = 2$ otherwise. Unlike maximization of Equation (3), this makes use of all available information about possible sibling relationships.

4.6 *Uncertain detection at the ocean surface*

The method described above assumes that animals on the surface are detected with certainty, but in some surveys they may be detected with probability $\phi < 1$ (e.g., Conn et al., 2014). In this case, let U_j be the event that an individual is both on the ocean surface and detected by the jth camera. This gives $\Pr(U_j) = \phi\kappa/\tau$, and in place of (6) we have

$$\Pr(U_2 \mid U_1) = \phi \left[\frac{\kappa}{\tau} + \frac{\tau - \kappa}{\tau} \exp\left\{ -\left(\frac{1}{\kappa} + \frac{1}{\tau - \kappa} \right) l \right\} \right].$$

4.7 *Dive-cycle duration and uncertain detection of surfacing individuals*

The parameters ϕ, κ, and τ are used to model dependence due to diving behavior, and are necessary to characterise the joint distribution of U_1 and U_2. However, surveys with two lagged capture occasions do not provide the information to estimate both dive-state parameters. This is not unique to our estimation approach, and remains a problem even if identities are observed; see Web Appendix F. Hiby and Lovell (1998) dealt with this by fixing τ at some known value. Additionally, under scenarios where detection at the surface is not certain, the parameter ϕ is not identifiable. If the lag between cameras is short, two of these parameters must be supplied in order to obtain an unbiased estimate of D_2. In good conditions it may be reasonable to fix $\phi = 1$, but otherwise it can be estimated from auxiliary data, as may either τ or κ (e.g., using data from tags). Uncertainty in these estimates can be incorporated into a bootstrap procedure so that it is reflected in the variance estimate of

animal density.

However, when the lag between cameras is large, we may assume independence between U_1 and U_2. Under the diving behavior framework in Section 4.4.1, plotting $\Pr(U_2 \mid U_1)$ against l in Equation (6) demonstrates that $\Pr(U_2 \mid U_1) \approx \Pr(U_2)$ for any $l \geqslant \tau$. The joint distribution of U_1 and U_2 is then characterised by a single parameter, $\gamma = \phi\kappa/\tau$, representing the probability that an individual in the detection region is detected by a passing camera. In this case, we can estimate the identifiable parameter vector (D_2, σ, γ) instead of $(D_2, \sigma, \tau, \kappa, \phi)$. Thus, our method is fully robust to both uncertain detection at the surface and unknown dive-cycle parameters if the lag between capture occasions is sufficiently large.

4.8 *Software implementation*

We have implemented the methods described above in the R package `palm` (Stevenson, 2018), which is available on CRAN; see Web Appendix J for an example of its use.

5. Application

Aerial surveys that monitor cetaceans using two planes, or two cameras mounted on the same plane, are currently still in the planning phase. Part of this planning is to ensure that there is suitable statistical methodology to analyse the resulting data, which is our purpose here. Our focus is therefore to establish an analytical framework that enables evaluation of the likely precision that a two-camera survey can achieve, to aid practitioners in assessing the costs of running two aircraft or engineering a double camera-mount on a single aircraft. This is best done by simulation for the specific taxa and survey conditions under consideration. Due to the lack of real data at present, we provide here an analysis of synthetic data generated by resampling real data from aerial circle-backs to emulate a two-camera survey.

Circle-backs were conducted on a series of surveys that recorded pods of harbor porpoise (*Phocoena phocoena*) in the North Sea and off the coasts of western Denmark, northeastern

Germany, and The Netherlands. A single plane intermittently circled back to retrace its path, with two observers on board who recorded the location of each pod they sighted. Of the stretches of ocean that were covered twice, the average lag between the passes was 248 s. See Web Appendix G for further information about the circle-back data, and how it was used to construct our synthetic two-camera aerial survey data. Our synthetic data represented a survey of length $k = 1\,100$ km with a detection region halfwidth of $w = 0.125$ km.

We estimated porpoise pod density using the method described in Sections 3 and 4. The lag was set at $l = 248$ s and this was longer than any plausible value for τ, the mean dive-cycle duration (Westgate et al., 1995). We therefore only estimated a single parameter, $\gamma = \kappa/\tau$, the proportion of time spent on the surface, for the distribution of C, and used $\Pr(U_2 \mid U_1) = \Pr(U_1) = \gamma$, as justified in Section 4.7. We set $b = 2$ km, a value suitably large to ensure that the error associated with the numerical approximation in (7) was negligible.

Our estimation method produced estimates (with 95% confidence intervals using the percentile bootstrap method) of 1.05 (0.84, 1.60) pods per km^2 for D_2, 0.15 (0.11, 0.19) km for σ, and 0.86 (0.56, 1.00) for γ. These uncertainties correspond to coefficients of variation of 19% for D_2, 16% for σ, and 13% for γ. See Web Appendix G for discussion about these parameter estimates, and how they compare to values obtained in other studies.

6. Simulation studies

We conducted three simulation studies to assess the performance of our animal density estimator for two-camera aerial surveys. For all simulations, we set $\tau = 110$ s, consistent with surfacing data reported by Westgate et al. (1995), and $\kappa = 94$ s to match $\hat{\gamma}$ from the synthetic data analysis. We set σ based on a calculation involving l, which varied across studies, and an animal speed of 0.95 ms^{-1}. Likewise, we matched D_2 to its estimate from the synthetic data. Survey constants w, b, and k were also set to values from Section 5.

First, the synthetic data in Section 5 had a large lag between the passing cameras, and so

we could assume independence between U_1 and U_2 (Section 4.7). This is not appropriate if two cameras are attached to the same plane; we test this scenario below by generating data with a shorter lag of 20 s. This simulation study is described below. Additional simulation studies test the robustness of our estimator to various assumption violations and compare it to existing spatial and nonspatial CR models, see Web Appendices H and I.

In total, we simulated 10 000 data sets using a lag of $l = 20$ s. From each simulated data set, we estimated the parameters by maximizing the expressions in both (3) and (5), allowing comparison between estimators that do and do not use the partial sibling information.

[Figure 4 about here.]

Results are shown in Figure 4. Estimators for all parameters showed negligible bias, especially those that used the partial sibling information (observed biases of 0.5%, 0.1%, and -0.5% for $\hat{D}_2$, $\hat{\sigma}$, and $\hat{\kappa}$). Density estimates for those that used partial sibling information were slightly more precise (coefficient of variation of 7.7%) than those that did not (9.5%).

7. Discussion

While our cluster capture-recapture estimator was motivated by two-camera aerial surveys, our generalization of the method proposed by Tanaka et al. (2008) in Section 3 is of general interest beyond the scope of ecology. It allows NSPP parameter estimation in a broader context, accounting for (i) any distribution for the number of children generated by each parent, (ii) d-dimensional spatial data, and (iii) known information about sibling relationships. Because our estimation framework is an extension of point process methodology, existing tools for model selection and goodness-of-fit are available from the point process literature.

Our method involves a computationally efficient estimation process. On average, computing point estimates from each simulated data set in Section 6 took just 1.6 s on a laptop computer with a 2.00 GHz processor. This indicates cluster capture-recapture has promise

for application to a wide class of wildlife surveys with uncertain animal identification.

7.1 *Clustering of individuals*

Web Appendix H shows some robustness to departures from the assumption that parent locations are a homogeneous Poisson point pattern. However, this assumption is clearly violated if animals travel in groups, potentially causing non-negligible bias. We therefore treated porpoise pods as the detection unit—a common approach in distance sampling, as it is more reasonable to assume independence between groups than independence between animals (Buckland et al., 2001, p. 432). We estimate pod density, but an estimate of the average pod size allows conversion to animal density (e.g., see Hammond et al., 2002, 2013).

If we wish to consider individuals as the detection unit, one option is to consider a multi-generational NSPP, where 'grandparents' are defined as a pod centroid, around which individuals' parent locations are clustered. A Palm likelihood estimator for such processes may provide estimators of both pod and animal densities from the survey data.

7.2 *Concluding remarks*

We have proposed a new method for analysing data from aerial two-camera surveys that accounts for uncertain animal identification, and models the dependence between detections on different occasions due to animal movement into and out of the detection region. Non-spatial CR methods fail to account for animal movement and cannot estimate the effective sampling area (Royle et al., 2013; pp. 167–169), precluding unbiased estimation of animal density (Web Appendix I). Area-search SCR methods do account for animal movement, but both of these existing methods force the surveyor to match detections to identities.

If the lag between cameras is short, then manually allocating identities to individuals could potentially be achieved with reasonable accuracy. If this lag is short relative to the dive-cycle duration, then it is necessary to account for dependence between capture occasions (or cameras) due to cetacean diving behavior via parameters τ and κ; moreover, under a

short-lag scenario, we must account for uncertain detection at the surface, $\phi < 1$, if survey conditions are not conducive to an assumption of perfect detection (Sections 4.6 and 4.7).

Existing methods that estimate between-occasion dependence in standard nonspatial CR and area-search SCR methods (e.g., variants of model M_b; Otis et al., 1978) do not do so with cetacean diving behavior in mind. It is only possible to estimate one of τ, κ, and ϕ directly from the aerial survey, even if identities are observed (Web Appendix F). For short-lag surveys, we recommend estimating two of these parameters from auxiliary data (Section 4.7) and incorporating any uncertainty into the bootstrap variance estimator (Section 3.3).

If the lag between cameras is long, by contrast, then animal density is estimable without fixing any parameters, as only the parameter $\gamma = \phi\kappa/\tau$ is required (Section 4.7).

8. Supplementary materials

Web Appendices and Figures referenced in Sections 3–7 are available with this paper at the Biometrics website on Wiley Online Library, along with the R code used to fit models to the synthetic harbor porpoise data and carry out all simulation studies.

References

Baddeley, A., Rubak, E., and Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*. CRC Press, Boca Raton.

Baudin, M. (1981). Likelihood and nearest-neighbor distance properties of multidimensional Poisson cluster processes. *Journal of Applied Probability* **18,** 879–888.

Borchers, D., Distiller, G., Foster, R., Harmsen, B., and Milazzo, L. (2014). Continuous-time spatially explicit capture-recapture models, with an application to a jaguar camera-trap survey. *Methods in Ecology and Evolution* **5,** 656–665.

Borchers, D. L. and Efford, M. G. (2008). Spatially explicit maximum likelihood methods for capture-recapture studies. *Biometrics* **64,** 377–385.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. (2001). *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*. Oxford University Press, Oxford.

Chandler, R. B. and Royle, J. A. (2012). Spatially explicit models for inference about density in unmarked or partially marked populations. *The Annals of Applied Statistics* **7,** 936–954.

Conn, P. B., Hoef, J. M. V., McClintock, B. T., Moreland, E. E., London, J. M., Cameron, M. F., Dahle, S. P., and Boveng, P. L. (2014). Estimating multispecies abundance using automated detection systems: Ice-associated seals in the bering sea. *Methods in Ecology and Evolution* **5,** 1280–1293.

Dawson, D. K. and Efford, M. G. (2009). Bird population density estimated from acoustic signals. *Journal of Applied Ecology* **46,** 1201–1209.

Efford, M. G. (2011). Estimation of population density by spatially explicit capture-recapture analysis of data from area searches. *Ecology* **92,** 2202–2207.

Fewster, R. M., Stevenson, B. C., and Borchers, D. L. (2016). Trace-contrast models for

capture-recapture without capture histories. *Statistical Science* **31,** 245–258.

Guan, Y. (2006). A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association* **101,** 1502–1512.

Hammond, P. S., Berggren, P., Benke, H., Borchers, D. L., Collet, A., Heide-Jørgensen, M. P., et al. (2002). Abundance of harbour porpoise and other cetaceans in the North Sea and adjacent waters. *Journal of Applied Ecology* **39,** 361–376.

Hammond, P. S., Macleod, K., Berggren, P., Borchers, D. L., Burt, L., Cañadas, A., et al. (2013). Cetacean abundance and distribution in European Atlantic shelf waters to inform conservation and management. *Biological Conservation* **164,** 107–122.

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

Link, W. A., Yoshizaki, J., Bailey, L. L., and Pollock, K. H. (2010). Uncovering a latent multinomial: Analysis of mark-recapture data with misidentification. *Biometrics* **66,** 178–185.

Neyman, J. and Scott, E. L. (1952). A theory of the spatial distribution of galaxies. *The Astrophysical Journal* **116,** 144–163.

Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs* **62,** 3–135.

Prokešová, M. and Jensen, E. B. V. (2013). Asymptotic Palm likelihood theory for stationary point processes. *Annals of the Institute of Statistical Mathematics* **65,** 387–412.

Royle, J. A., Chandler, R. B., Sollmann, R., and Gardner, B. (2013). *Spatial Capture-Recapture.* Academic Press, Amsterdam.

Stevenson, B. C. (2018). *palm: Fitting Point Process Models via the Palm Likelihood.* R package version 1.1.1.

Tanaka, U., Ogata, Y., and Stoyan, D. (2008). Parameter estimation and model selection

for Neyman-Scott point processes. *Biometrical Journal* **50,** 43–57.

Tancredi, A., Auger-Méthé, M., Marcoux, M., and Liseo, B. (2013). Accounting for matching uncertainty in two stage capture-recapture experiments using photographic measurements of natural marks. *Environmental and Ecological Statistics* **20,** 647–665.

Waagepetersen, R. P. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics* **63,** 252–258.

Westgate, A. J., Read, A. J., Berggren, 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.

Wright, J. A., Barker, R. J., Schofield, M. R., Frantz, A. C., Byrom, A. E., and Gleeson, M. (2009). Incorporating genotype uncertainty into mark-recapture-type models for estimating abundance using DNA samples. *Biometrics* **65,** 833–840.
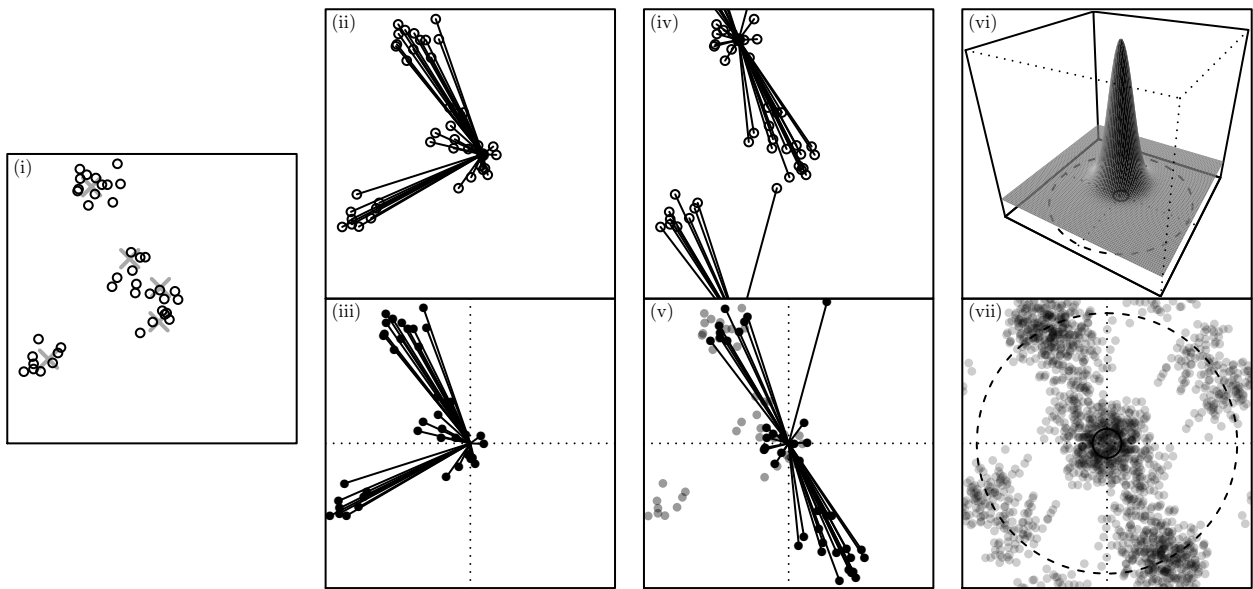
Figure 1. An example of a Thomas point pattern (observed on the unit square) and exposition of the estimation approach due to Tanaka et al. (2008): (i) Children (open circles) and parents (grey crosses) generated from simulation of a Thomas process. Despite there being five parents, due to some being in close proximity the children are loosely aggregated in three groups. (ii) A difference process taken from one particular point. (iii) Shifting this difference process to the origin. (iv) Another difference process taken from a different point; note the use of periodic boundary conditions to deal with edge effects. (v) Superposition of this second difference process onto the first. (vi) The expected intensity of the superposed difference process—this declines radially from the origin, and is given by $n\lambda_0(r)$, where r is the distance from the origin. Thus, intensity of the superposed difference process is high near the origin (solid circle), eventually declining to an asymptote at regions further from the origin (dashed circle). (vii) Superposition of all n difference processes from the small sample $n = 41$ shown in (i).
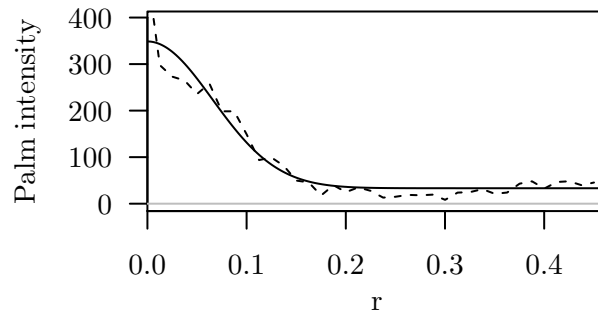
Figure 2. The empirical Palm intensity function (dashed line; see Fewster et al., 2016) and the best-fitting Palm intensity function of the form $\lambda_0(r) = D\nu + \nu \exp\{-r^2/(4\sigma^2)\}/(4\pi\sigma^2)$ (solid line) from the point pattern in Figure 1(i). The latter was estimated using the method of Tanaka et al. (2008).
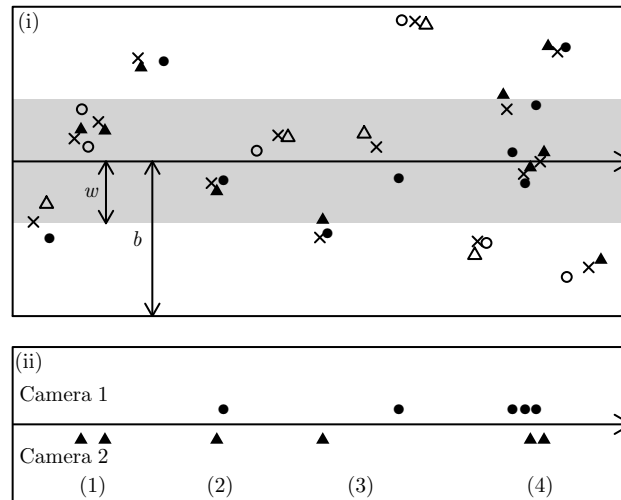
Figure 3. Example simulated data from a two-camera survey. Plot (i): Parent locations (crosses) along with corresponding locations at the passing of the first (circles) and second (triangles) cameras. Solid plotting characters depict individuals on the surface at the passing of a camera, while open plotting characters depict those that were not. The horizontal arrow represents the transect flown by the two cameras; the grey region is the detection region. A detection is made if an individual is both on the surface and in the detection region. Plot (ii): The resulting one-dimensional point pattern that is collected and analyzed here. The identities of the individuals detected are uncertain. Note the four labeled groups of points: Identities associated with (1) are unambiguous—different detections from the same camera cannot be of the same animal. Identities associated with (2) are ambiguous and there are two possible allocations—these points could represent two detections of the same individual, or one detection each of two different individuals. Identities associated with (3) are unambiguous—the distance between the detections is too great for these to correspond to the same individual. Identities associated with (4) are ambiguous—between three and five individuals have been detected, and there are 13 different possible allocations within this group alone. Therefore, in conjunction with the ambiguous group (2), there are a total of 26 different possible allocations in this small example data set.
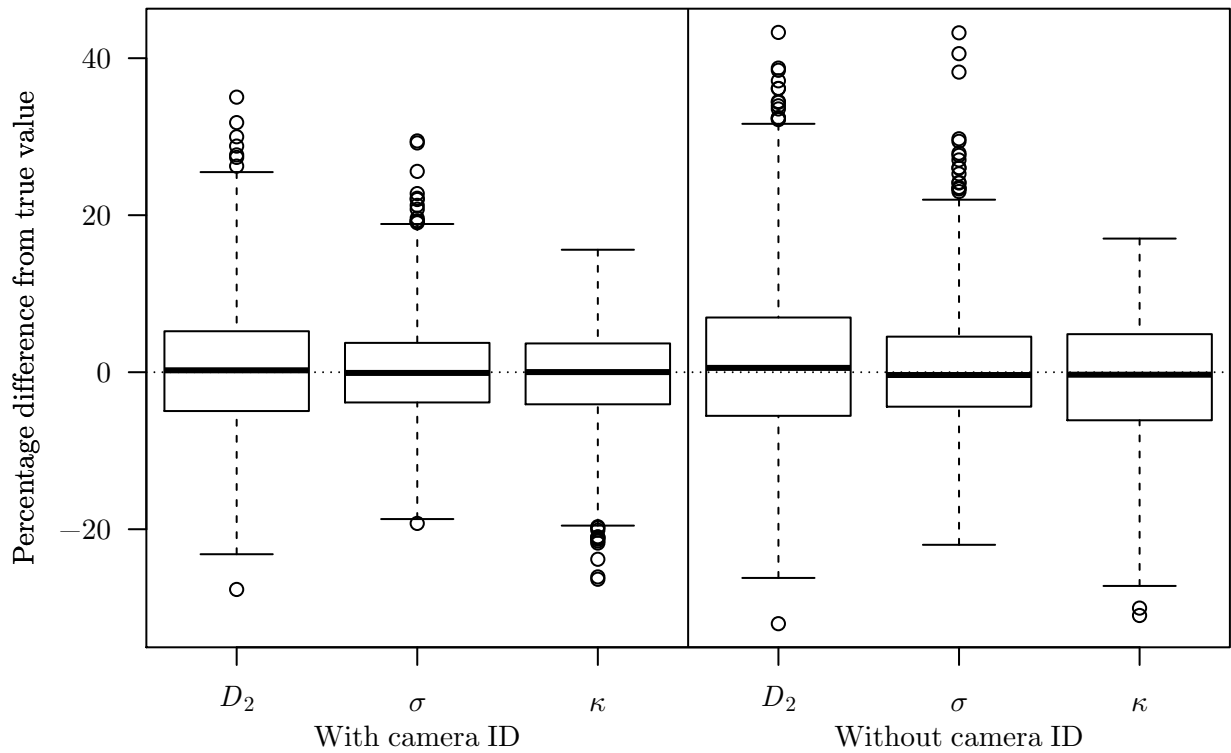
Figure 4. Box plots showing point estimates from the 10 000 simulated data sets for models that both did (via maximization of (5)) and did not (via maximization of (3)) incorporate the partial sibling information via the camera ID. Estimates were obtained with parameter τ fixed at 110 s.