Statistical Science
2016, Vol. 31, No. 2, 245–258
DOI: 10.1214/16-STS551

Trace-Contrast Models for Capture–Recapture Without Capture Histories

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

Abstract. Capture–recapture studies increasingly rely upon natural tags that allow animals to be identified by features such as coat markings, DNA profiles, acoustic profiles, or spatial locations. These innovations greatly increase the number of capture samples achievable and enable capture–recapture estimation for many inaccessible and elusive species. However, natural features are invariably imperfect as indicators of identity. Drawing on the recently developed Palm likelihood approach to parameter estimation in clustered point processes, we propose a new estimation framework based on comparing pairs of detections, which we term the trace-contrast framework. Importantly, no reconstruction of capture histories is needed. We show that we can achieve accurate, precise, and computationally fast inference. We illustrate the methods with a camera-trap study of a partially marked population of ship rats (*Rattus rattus*) in New Zealand.

Key words and phrases: Camera-traps, mark recapture, natural tags, Neyman–Scott process, palm likelihood estimation, *Rattus* species.

1. INTRODUCTION

A technological revolution in methods of wildlife recognition is currently taking place. New technologies enable individual animals to be distinguished by photographic, genetic, acoustic, or location metrics, either alone or in combination (e.g., Kühl and Burghardt, 2013, Carroll et al., 2011, Charlton et al., 2011, Borchers and Efford, 2008). These developments create very different sampling opportunities and data from the traditional mark-recapture studies pioneered by Cormack (1964), Jolly (1965), and Seber (1965), in which animals are physically marked by investigators. Our aim here is to introduce and explore a new way of thinking about capture–recapture data, in preparation for a future in which animal identity might be observed only indirectly through a series of informative metrics.

Natural attributes that allow animals to be individually recognized are known as *natural tags*, in contrast to the marked tags that investigators place upon animals. Many animals are sufficiently distinctive to be recognizable from photographs or acoustic traces, while an animal's spatial location is indicative of its identity if capture–recapture is conducted using a sequence of observations made in quick succession. Genetic profiling is a powerful method of distinguishing individuals, although it is not always easy to obtain genetic samples. The opportunities for monitoring hard-to-sample populations using any or all of these protocols are exciting and are rich in statistical challenges. Large but inaccessible creatures such as Antarctic whales could be monitored by drones or satellites (Fretwell, Staniland and Forcada, 2014), while estimation for elusive species in dense forest could become routine with the deployment of microphone or camera arrays (Borchers et al., 2014, Stevenson et al., 2015).

R. M. Fewster is Associate Professor, Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand (e-mail: r.fewster@auckland.ac.nz). B. C. Stevenson is Research Fellow, D. L. Borchers is Professor, School of Mathematics and Statistics, Centre for Research into Ecological and Environmental Modelling, The Observatory, Buchanan Gardens, University of St Andrews, Fife KY16 9LZ, UK (e-mail: bcs5@st-andrews.ac.uk; dlb@st-andrews.ac.uk).

All methods of individual identification that rely upon natural tags share the same key challenge, that individuals can no longer be identified with certainty. A trade-off is introduced between the number of samples and their quality. Using natural tags can greatly increase the number of samples it is feasible to collect, but sacrifices certainty as to whether two samples do or do not correspond to the same animal. Even genetic profiling, which might be thought conclusive, suffers from a nonignorable level of laboratory error, which is especially pronounced with low-quality genetic samples obtained noninvasively from feathers, hair, or faeces (Wright et al., 2009, Taberlet and Luikart, 1999).

Researchers typically deal with identity uncertainty at the sample-matching stage, reconstructing capture histories for individuals by comparing samples and verifying putative matches. For example, photographs are often matched through laborious examination of the photo catalogue by a panel of experts (e.g. Carroll et al., 2011). Genetic profiles are typically treated as matches if they agree on more than a threshold number of loci, and where there is doubt, the samples might be genotyped repeatedly for verification, a costly and time-consuming process (Vale et al., 2014). A different approach is taken by Wright et al. (2009) and Barker et al. (2014), who incorporate into their models direct estimates of error rates based on multiple genotyping attempts, and treat capture histories as latent variables to be sampled by a MCMC algorithm.

With all these approaches, model fitting proceeds on the basis of reconstructed capture histories, whether they are assumed to be determined without error following a manual matching process or sampled through MCMC. In this article we develop an alternative framework in which we consider capture–recapture estimation without capture histories. Our aim is to sidestep altogether the process of deciding which samples are matches and which are not, and instead to rely upon the information gained from "similar" and "dissimilar" pairwise comparisons.

We use the term *trace* to indicate any type of animal detection record, such as a photograph, footprint, acoustic recording, genetic sample, or location in space or time. We describe our approach as *trace-contrast modeling* because it is based on pairwise comparisons or contrasts between traces. By avoiding making decisions about sample-matching, we aim to accommodate the much greater number of samples that will become available through improved technologies for wildlife recognition, and for which manual matching is likely to become infeasible or error-prone. We also show how the trace-contrast framework allows us to deal with different marking levels in the population, including unmarked or partially marked populations. We give an illustration using camera-trap data from a partially marked population of ship rats (*Rattus rattus*) in New Zealand.

2. A FRAMEWORK FOR ANALYSIS

Our framework is based on ideas from spatial point process analysis, particularly drawing on the work of Tanaka, Ogata and Stoyan (2008). We begin by explaining the context and approach developed by Tanaka, Ogata and Stoyan (2008), and then show how the same ideas can be applied to problems in capture–recapture with uncertain identities.

Suppose we have a field of apple trees, in which the trees are invisible but we can see dropped apples. Our interest is in estimating the number of trees, but our only evidence comes from apples. The situation is pictured in Figure 1, in which the invisible trees are marked by crosses and their detectable apples are marked by points. Although apples cluster around their own trees, there are regions of overlap in which neither the number of trees nor the assignment of apples to trees is clear. The aim is to estimate the number of trees without needing to make judgements about which apples belong to which trees.

More generally, we conceptualize the trees to be unobservable *parent points*, and we conceptualize each parent to produce a number of detected *offspring points*, specifically apples. Offspring who share a parent, in other words, apples that fall from the same tree, are referred to as *siblings*, and the set of offspring of a parent is called its *family*. We shall eventually connect

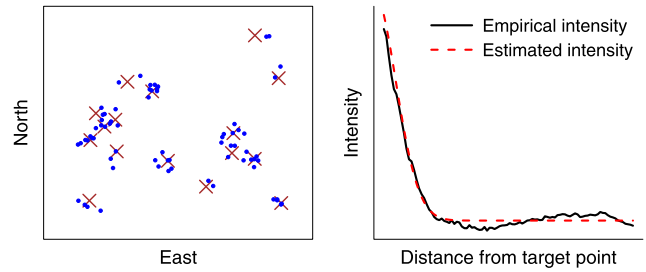

FIG. 1. *Left*: *a field of apple trees* (*crosses*) *with dropped apples* (*points*) *clustering around each tree. Right*: *intensity function for the difference process*, $\lambda_0(r)$, *describing the expected number of points per unit area at distance r from a single target point in the left panel. The solid line is the empirical intensity* $\tilde{\lambda}_0(r)$ (*see text*), *and the dashed line is the parametric curve* $\lambda_0(r)$ *evaluated at the maximum Palm likelihood estimates of* μ, ν, *and* σ (*see text*).

this framework to capture–recapture studies by regarding individual animals as the unobservable parents, and the detected samples or traces of the animal to be the offspring points.

The Neyman–Scott process (Neyman and Scott, 1958) is a popular way of modeling the parent–offspring cluster process. The invisible parent points arise from a homogeneous Poisson process with intensity μ. Each parent generates a random number of observable offspring, scattered about the parent location. The number of offspring of a parent, K, is a random variable with mean ν: for example, $K \sim$ Poisson(ν). Offspring positions are generated from a spatial probability density, for example, the Thomas process (Thomas, 1949) uses a bivariate Gaussian centred on the parent location with covariance matrix $\sigma^2 I_2$, where I_2 is the 2×2 identity matrix. The parameters in this example are (μ, ν, σ), where μ specifies the density or abundance of the unseen parent points. These spatial cluster processes are also known as contagion processes (Diggle, 2003).

There are various ways of estimating the parameters (μ, ν, σ), but a full maximum likelihood analysis is generally considered computationally impracticable (Guan, 2006, Waagepetersen, 2007, Tanaka, Ogata and Stoyan, 2008). Tanaka, Ogata and Stoyan (2008) propose two key steps for a pseudo-likelihood estimation procedure:

1. Construct the *contrast process* consisting of all pairwise distances or *contrasts* between points, and derive the intensity function of this process in terms of the parameters (μ, ν, σ).
2. Assume that the contrast process can be approximated by an inhomogeneous Poisson process with the intensity function derived in step 1. Maximize the corresponding inhomogeneous Poisson likelihood to estimate (μ, ν, σ).

This method is described as *Palm likelihood estimation*, and is one of three methods in the `spatstat` R package (Baddeley and Turner, 2005) for fitting clustered point processes, the other options being due to Guan (2006) and Waagepetersen (2007). The idea of modeling contrasts is common to all three methods. The Palm likelihood method is attractive for our purposes because the Poisson formulation can easily be extended to include additional model components.

2.1 The Contrast Process

We refer to the process of offspring points described above as the *offspring process* for short. The contrast process consists of pairwise distances of the form $r_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$, where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the spatial positions of offspring points i and j. Due to the proximity of siblings, the intensity of the contrast process peaks at short distances, as seen in Figure 1. Beyond the range of sibling distances, it reaches an asymptote corresponding to the background intensity of the offspring process. Our interest lies in deriving the precise parametric form of this intensity function in terms of (μ, ν, σ).

Some care must be taken when defining exactly what is meant by the contrast process and its intensity. We assume that our observations take place through a finite window onto a vast, stationary, isotropic point process with parameters (μ, ν, σ). In our example, our data are the apples in a single field, but we imagine the trees and apples extending across the landscape in all directions. We examine pairwise distances of points up to a maximum distance R, where R must be pre-selected such that it is large enough to capture the asymptote shown in Figure 1, but not so large that it creates problems with edge effects and computability.

For creating distance comparisons, or contrasts, we select *target* points to act as foci, such that we compile distances from the target points one by one. The finite observation window can be dealt with either by creating a buffer zone around the perimeter of the window or by employing periodic boundary conditions. In the buffer-zone treatment, the target points are interior points whose distance from the window boundaries is greater than R. In the periodic boundary scenario, we imagine that the window picks up at one edge where it leaves off at the opposite edge. This means that the right-hand edge of a rectangular window is treated as if it is glued to the left-hand edge, such that points on the right edge are considered close to those on the left edge, and similarly for the top and bottom edges. The buffer-zone treatment is more generally defensible, although it leads to a reduced number of target points (Tanaka, Ogata and Stoyan, 2008, Diggle, 2003, page 13). We define T to be the number of target points. In the buffer-zone treatment, T is the number of interior offspring points, whereas in the periodic boundary treatment, T is the total number of offspring points.

Before defining the one-dimensional contrast process of distances $r_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$, we first define the two-dimensional *difference process* of points $\mathbf{x}_i - \mathbf{x}_j$ to clarify the link with other literature and to provide an extendable development. The difference process is constructed in a disc of radius R centered at the origin as shown in Figure 2. Consider taking just one of the T

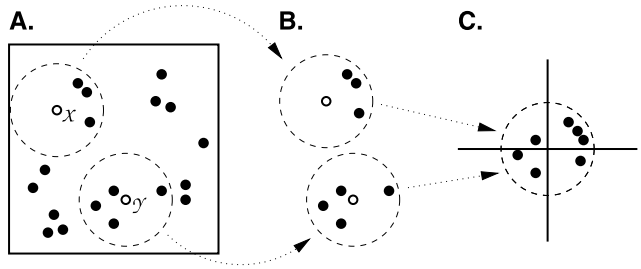

FIG. 2. (A) *The original process of offspring points. The two white points denote offspring points* $\mathcal{X}$ *and* $\mathcal{Y}$ *that are temporarily selected as target points*; *circles of radius* R *are centered on each target*. (B) *Two realizations of the difference process*, *consisting of the filled points inside the two circles of radius* R. (C) *The superposition of the two realizations of the difference process*, *gained by placing the two circles from panel* B *at the same origin and deleting the two white target points. Eventually*, *all* T *eligible points in panel* A *will act as targets*, *and the plot in panel* C *will become the superposition of the corresponding* T *circles of radius* R, *known as a Fry plot*.

target points, say, point $\mathcal{X}$ in Figure 2(A). We draw a circle of radius R centered on point $\mathcal{X}$, then relocate it to the origin, including all filled points captured within the circle, but not including the target point $\mathcal{X}$ itself [Figures 2(B) to 2(C)]. This is a single realization of the difference process. We then repeat the same steps for another target point $\mathcal{Y}$, superposing the two realizations. Eventually we have T realizations of the difference process all superposed at the origin, corresponding to each of the T target points in turn. The norms of all points in these T superposed realizations constitute the *contrast process*. Thus, Figure 2(A) shows the original offspring process; Figure 2(B) shows two realizations of the difference process; and Figure 2(C) shows the first two of the T superposed realizations of the difference process, the norms of which will create the contrast process. We describe the norm $r_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$ as a *contrast*.

The intensity function of the difference process is defined as $\lambda_0(r)$ for $0 \le r \le R$. In terms of the offspring process, $\lambda_0(r)$ is the expected number of points per unit area located in a thin ring of radius r centered on a single target point. Focusing only on two-dimensional processes at present, the most convenient definition of λ_0 is

$$\lambda_0(r) = \frac{1}{2\pi r}\frac{d}{dr}\Lambda_0(r),$$

where Λ_0 is defined with respect to the offspring process as

$$\Lambda_0(r) = \mathbb{E}(\text{number of further points located distance} \le r \text{ from one target point}).$$

Equivalently, we can define Λ_0 with respect to the contrast process as

$$\Lambda_0(r) = \frac{\mathbb{E}(\text{number of contrasts} \le r)}{T},$$

where we divide by T because the contrast process is gained from superposing T realizations of the difference process. Our final definition for $\lambda_0(r)$ in terms of the contrast process is then

$$\lambda_0(r) = \frac{1}{2\pi r T}\frac{d}{dr}\{\mathbb{E}(\text{number of contrasts} \le r)\}. \tag{1}$$

The function $\lambda_0(r)$ is called the *Palm intensity function* of the offspring process, named after the work of Palm (1943). We note in passing that $\lambda_0(r)$ is related to the pair correlation function $g(\cdot)$ and Ripley's K-function $K(\cdot)$ of the offspring process, via the expressions $\lambda_0(r) = \mu\nu g(r)$ and $\Lambda_0(r) = \mu\nu K(r)$ (Tanaka, Ogata and Stoyan, 2008). However, we shall only work with the definitions given above.

We can now explain the intensity curves shown in Figure 1. The empirical intensity, shown in the solid line, is $\tilde{\lambda}_0(r)$, where

$$\tilde{\lambda}_0(r) = \{(\text{number of contrasts} \le r + \varepsilon/2) - (\text{number of contrasts} \le r - \varepsilon/2)\}/(2\pi r T \varepsilon),$$

for suitably small ε. The dashed line is the parametric form of the Palm intensity $\lambda_0(r)$, parametrized in terms of (μ, ν, σ), and evaluated at the maximum Palm likelihood estimates (see below).

It is intuitively clear that the Palm intensity should be informative about μ, ν, and σ. Broadly speaking:

- For large r, beyond the range of siblings, the curve $\lambda_0(r)$ converges to the overall intensity of the process, $\mu\nu$.
- The peak in $\lambda_0(r)$ for small r is due to short distances between siblings of the same family, so the width of the peak is informative about the parameter for offspring dispersal around the parent, σ.
- The height of the peak reflects the concentration of siblings and is informative about the average number of offspring per parent, ν.

For the Neyman–Scott process described above, with $K \sim \text{Poisson}(\nu)$ offspring per parent and offspring dispersal governed by Gaussian$(\mathbf{0}, \sigma^2 I_2)$, we show in the Appendix that

$$\lambda_0(r) = \mu\nu + \frac{\nu}{4\pi\sigma^2}\exp\left(-\frac{r^2}{4\sigma^2}\right) \quad (r \ge 0), \tag{2}$$

from which the comments above about the roles of μ, ν, and σ can be verified.

Much of the challenge of our proposed framework is in deriving the correct form for the Palm intensity $\lambda_0(r)$, given the specification of the offspring process. The Appendix details the derivation of (2) for the two-dimensional Neyman–Scott process, and derives the equivalent expression for $\lambda_0(r)$ for the one-dimensional Neyman–Scott process.

2.2 Maximum Palm Likelihood Estimation

The discussion in the previous section, as well as the pictorial representation in Figure 1, demonstrates that the parameters (μ, ν, σ) are estimable by fitting the Palm intensity $\lambda_0(r)$ to the contrast data. In order to fit the curve, we need an objective function that can be optimized with respect to (μ, ν, σ). While there are many choices for an objective function, Tanaka, Ogata and Stoyan (2008) propose the likelihood function corresponding to an inhomogeneous Poisson process with intensity $\phi(r) = 2\pi r T \lambda_0(r)$. Thus, $\phi(r)$ denotes the expected number of contrasts in a small interval $[r, r+\delta r]$, divided by the interval width δr. Tanaka, Ogata and Stoyan (2008) describe the estimators resulting from maximizing this objective function as *maximum Palm likelihood estimators* (MPLEs).

In reality, the contrast process is not an inhomogeneous Poisson process, so properties of maximum likelihood estimators do not apply. The MPLE procedure was proved by Prokešová and Jensen (2013) to yield consistent estimators, but there is no theoretical support for variance estimation using the inverse Hessian matrix (Tanaka, Ogata and Stoyan, 2008, Prokešová and Jensen, 2013), so we recommend that variance is estimated by a bootstrap procedure. A key advantage of the MPLE formulation is that it can readily be extended to incorporate supplementary information about samples by using the marked point-process formulation of the inhomogeneous Poisson process likelihood. For these reasons we focus entirely on the MPLE formulation, while noting that it is only one way of fitting the curve $\lambda_0(r)$ to the contrast data.

The likelihood for an inhomogeneous Poisson process, with data $\mathbf{r} = (r_1, \ldots, r_n)$, and with rate function $\phi(r)$ parametrized by (μ, ν, σ) for $0 \le r \le R$, is

$$\mathcal{L}(\mu, \nu, \sigma; \mathbf{r}) = \frac{(\int_0^R \phi(r)\,dr)^n}{n!} \exp\Big(-\int_0^R \phi(r)\,dr\Big) \cdot \prod_{i=1}^n \frac{\phi(r_i)}{\int_0^R \phi(r)\,dr} \tag{3}$$

$$= \frac{1}{n!} \exp\Big(-\int_0^R \phi(r)\,dr\Big) \prod_{i=1}^n \phi(r_i).$$

Disregarding an additive constant, the objective function for estimating (μ, ν, σ) is given by the corresponding logarithm:

$$\ell(\mu, \nu, \sigma; \mathbf{r}) = -\int_0^R \phi(r)\,dr + \sum_{i=1}^n \log\{\phi(r_i)\}. \tag{4}$$

In the case of the Neyman–Scott process above, with offspring distribution $K \sim \text{Poisson}(\nu)$ and offspring positions generated from Gaussian$(\mathbf{0}, \sigma^2 I_2)$, the MPLE process is completed by maximizing (4) with respect to (μ, ν, σ), where from equation (2) we have

$$\phi(r) = 2\pi r T \Big\{\mu\nu + \frac{\nu}{4\pi\sigma^2} \exp\Big(-\frac{r^2}{4\sigma^2}\Big)\Big\}, \tag{5}$$

and the data $(r_1, \ldots, r_n)$ consist of all contrasts generated by the T selected target points. Equation (16) in Tanaka, Ogata and Stoyan (2008) is gained by combining (4) and (5), and using the closed-form $\int_0^R \phi(r)\,dr = \nu T[\pi\mu R^2 + 1 - \exp\{-R^2/(4\sigma^2)\}]$.

The formulation above is capable of yielding accurate and precise estimates of (μ, ν, σ), as long as clusters in the offspring point-process are reasonably distinguishable. Fundamentally, we succeed in estimating parameters of the original offspring process, including the density of parent points μ, without making any judgements about which offspring belong to which parent, and despite the overlap in family clusters.

2.3 Auxiliary Information

Recall that a contrast is $r = \|\mathbf{x} - \mathbf{y}\|$ for a pair of points $\mathbf{x}$ and $\mathbf{y}$ in the original offspring process. Suppose that each contrast r_i has associated with it some additional observation z_i that might be informative about μ, ν, σ, or some new parameters of interest. We aim to model the data $\mathbf{z} = (z_1, \ldots, z_n)$ along with the contrast data $\mathbf{r} = (r_1, \ldots, r_n)$. In the parlance of point processes, z_i is called a *mark*, and the contrast point process with the inclusion of this additional information is described as a *marked point process*.

For a single contrast, let $f(z|r)$ be the probability density of the observation z given the contrast distance r. Let θ be a vector of extra parameters needed for this model, where θ is empty if $f(z|r)$ involves only the parameters μ, ν, and σ.

A key advantage of the Palm likelihood approach to model fitting is the easy way in which the Poisson likelihood formulation (3) extends to incorporate the additional information $\mathbf{z}$. The joint density of the number of contrasts n, the contrast positions $r_1, \ldots, r_n$, and

the additional contrast information $z_1, \ldots, z_n$ factorizes as $f(n) f(r_1, \ldots, r_n|n) f(z_1, \ldots, z_n|r_1, \ldots, r_n; n)$. Assuming that each z_i depends only upon r_i and marks are mutually independent under the approximation of the contrast process by an inhomogenous Poisson process, we only need to append the factor $\prod_{i=1}^{n} f(z_i|r_i)$ to the likelihood in (3) and take logs to give the new objective function to be maximized in the presence of auxiliary information:

$$\ell(\mu, \nu, \sigma, \theta; \mathbf{r}) = -\int_0^R \phi(r)\, dr + \sum_{i=1}^{n} \log\{\phi(r_i)\} + \sum_{i=1}^{n} \log\{f(z_i|r_i)\}, \tag{6}$$

where $\phi(r) = 2\pi r T \lambda_0(r)$ for a two-dimensional process, and $\lambda_0(r)$ is the Palm intensity function as before.

It is important to note that the auxiliary information z is associated with a contrast $r = \|\mathbf{x} - \mathbf{y}\|$, so it typically comprises a couplet of information about the original offspring points $\mathbf{x}$ and $\mathbf{y}$. It might therefore be tempting to assume that z is independent of the contrast distance r, in other words, that $f(z|r) = f(z)$. However, care is needed with this assumption. If the observations in z reflect some property of the parents of the two points $\mathbf{x}$ and $\mathbf{y}$, then z is likely to depend upon r due to the influence of r on the probability that $\mathbf{x}$ and $\mathbf{y}$ share a common parent. If $\mathbf{x}$ and $\mathbf{y}$ are close (small r), they are more likely to have the same parent than if they are distant (large r). For the purpose of capture–recapture studies, the "parent" of a point $\mathbf{x}$ corresponds to the animal who deposited the sample at $\mathbf{x}$, so auxiliary information will typically be connected with the parent and we need to derive the influence of r on $f(z|r)$.

We achieve this by deriving an expression for the probability $b(r)$ that a randomly chosen contrast of magnitude r is generated by a pair of siblings in the offspring process. This is readily seen to be the relative intensity of sibling points to all points at distance r from the target, which is gained by isolating the sibling contribution to $\lambda_0(r)$. For the formulation of $\lambda_0(r)$ in (2), we obtain

$$b(r) = \frac{\lambda_0(r) - \mu\nu}{\lambda_0(r)}.$$

We then formulate $f(z|r)$ by partitioning over the two possibilities that a contrast of size r consists of siblings and nonsiblings. If r has no further influence on z, given the status of the contrast as siblings or nonsiblings, then

$$f(z|r) = f(z|\text{siblings}) b(r) + f(z|\text{nonsiblings})\{1 - b(r)\}. \tag{7}$$

2.4 Connection with Capture–Recapture Studies

The framework outlined above creates an intuitively-reasonable procedure for drawing inference on clustered spatial point processes without assigning cluster membership, and involves tools that are well known in the analysis of spatial point patterns. Our proposal is to explore how these ideas might be adapted to the capture–recapture context.

We conceptualize individual animals to be the invisible parent points described previously. Each animal contributes a random number K of detected samples or traces to the study, where $K = 0$ is permissible. These traces could be records of the animal's spatial location, for example, taken by a quick succession of aerial photographs. In this case we could directly apply the Neyman–Scott formulation derived throughout Section 2 to estimate μ, the density of animals in the area. However, the general conceptual framework embraces a wide spectrum of possible sample types. Any method of animal recognition in which traces from a single animal tend to be more similar than traces from different animals could, in principle, form a clustered process similar to that shown in Figure 1. While Figure 1 portrays a two-dimensional point process, in the general context it might be anything from one-dimensional to multi-dimensional, depending upon the metrics used to assess animal traces. Traces need not be restricted to spatial locations, but might include photographs, DNA samples, acoustic traces, time-stamps, and combinations of these.

In our proposed *trace-contrast framework*, we model properties of the pairwise differences between samples, rather than using samples to construct capture histories. The framework offers a natural treatment for many modern capture–recapture protocols, in which analysts first go through a lengthy process of pairwise comparisons in order to create capture histories—for example, by matching photographs or DNA samples—before applying a separate step to model capture histories. The trace-contrast framework removes the need for the latter step, and in some contexts offers a more accurate description of the way that samples are deposited by animals and processed by researchers. It also enables uncertainty in sample matching to be quantified as part of the final analysis.

The key advantages of the trace-contrast approach are therefore as follows:

(a) a more direct way of modeling the sampling process in some contexts;

(b) removing the need for sample matching, thereby saving the time and expense of verifying matches: for example by repeat genotyping, or, in the case of photographs, scrutiny by a panel of experts;

(c) reducing the impact of matching errors and instead creating a proper acknowledgment of matching uncertainty;

(d) suitability for very large numbers of samples with imperfect identity metrics, as might be obtained by automatic detectors;

(e) easy incorporation of information from partially marked populations, through the auxiliary information component $f(z|r)$ in (7): if partial marking supplies knowledge of sibling status for particular contrasts, then either the sibling or the nonsibling component of the partition in (7) can be omitted;

(f) depending on the model, estimation can be very fast, even when including bootstrapping;

(g) estimation does not involve latent variables for cluster position and membership, as demanded by approaches based on the full point-process likelihood using the EM algorithm or Bayesian MCMC sampling (Waagepetersen and Schweder, 2006, Møller and Waagepetersen, 2007). In particular, unlike the other approaches, fitting a trace-contrast model does not become more computationally intensive as the number of clusters in the data increases.

Counter to these advantages, if capture histories are readily constructed, then greater precision should be expected from traditional capture–recapture models. For example, we can compare the two-dimensional Neyman–Scott process in (4) and (5), parameterized by (μ, ν, σ), against the traditional model M_t parameterized by $(N, p_1, \ldots, p_m)$, where N is the number of animals in the study region, p_t is the detection probability for each animal on capture occasion $t = 1, \ldots, m$, and all animals are identified with certainty. To compare model performance on the same simulated data, we equate $\mu = N/A$, where A is the area of the study region, ascribe each spatial detection to a capture occasion at random, and construct capture histories based on known identities for model M_t, but use the spatial data without identity information for the trace-contrast model. We find that when σ is small enough for there to be negligible spatial overlap between detections of different animals, the trace-contrast model and model M_t have roughly equal precision. As σ increases, precision from trace-contrast modeling declines relative to model M_t, as is natural. This comparison is however rather artificial, as the two approaches are intended for very different sampling scenarios. The ultimate aim of trace-contrast models is to accommodate situations where lack of identity information is offset by a much higher quantity of samples than is usual for traditional capture–recapture studies.

Additional challenges and directions for further development include the following:

(a) creating a framework for model selection and goodness-of-fit testing, bearing in mind that the MPLE framework is not based on a true likelihood;

(b) investigating robustness and the impact of model misspecification;

(c) possible alternatives to bootstrap variance estimation;

(d) creating a suite of exemplars for the equivalent of the Palm intensity function $\lambda_0(r)$ for various models and sample types, including genetic, photographic, and acoustic data.

3. CAMERA-TRAP MODEL FOR NEW ZEALAND SHIP RATS

Our motivating example in this article is a study of camera-trap data for invasive ship rats (*Rattus rattus*) in forest reserves in northern New Zealand (Nathan, 2016). New Zealand has no native land mammals, so introduced mammals such as ship rats create enormous problems for the conservation of native species and habitats. Ship rats eat seeds and fruit, and predate directly on invertebrates, reptiles, and birds' nests, with the result that they have a severely deleterious impact on the health of the forest as well as its native inhabitants. Through forest damage, competition, and direct predation, they have been solely responsible for the global extinction of several endemic bird and reptile species (e.g., Bell, Bell and Merton, 2016, Robins et al., 2016).

There is considerable effort in New Zealand to improve methods of rat control and elimination. One active area of research is to investigate the efficacy of control devices such as traps and bait stations. In our motivating study, researchers mounted motion-sensitive video cameras around a selection of (non-live) control devices to record the rats' behavior when they encountered the device. The primary research aim is to estimate the probability that a rat interacts with the device, having encountered it, where an "interaction" corresponds to the rat entering a trap or taking a pellet of bait from a bait station. The chief problem is that there is no definition of what constitutes

an "encounter." Rats frequently linger for several minutes around a device, repeatedly triggering the motion cameras, so there might be several detections per encounter. However, the biologically relevant unit of assessment is the encounter as a whole, not the individual motion-triggered detections within the encounter.

Here, we show how the camera data can be modeled in the trace-contrast framework. An unobservable parent point corresponds to a single encounter of a single rat with a device. Detections or traces correspond to motion-triggered video recordings, which each have a time-stamp. The offspring points of the original process correspond to the times at which these detections take place. The clustered point process therefore takes place on a one-dimensional time axis. Sibling points correspond to multiple detections of a single encounter. The definition of an "encounter" is made indirectly via the clustering of detections over time. Instead of imposing arbitrary thresholds—for example, defining an encounter to be a period of activity lasting 10 minutes—we allow the clustering patterns in the data to delimit encounters. Sibling points, corresponding to within-encounter detections, generate a peak in the Palm intensity function as shown in Figure 1. By distinguishing the peak from the asymptote we can estimate properties of encounters, such as interaction probability, without needing to define what constitutes an encounter or assign detections to encounters.

Our parameter of interest, $\alpha = \mathbb{P}(\text{interaction}|\ \text{encounter})$, enters the model only through the auxiliary component of the Palm likelihood, outlined in Section 2.3. We define parents (encounters) to be one of two types: an interaction type $\mathcal{I}$ or a noninteraction type $\overline{\mathcal{I}}$. The status of each parent encounter is determined by parameter α, but is nonobservable. However, if the encounter is of type $\mathcal{I}$, the interaction may be *revealed* on some of the video recordings of the encounter. The auxiliary information z corresponding to a pairwise contrast of video recordings i and j is the pair (I_i, I_j), specifying whether each of the recordings i and j did or did not reveal the rat interacting with the device. If the rat is seen interacting with the device during video recording i, then $I_i = 1$, otherwise $I_i = 0$.

3.1 Detection Data

The full study described by Nathan (2016) involves various control devices, with two motion-sensitive infrared video cameras and a passive integrated transponder (PIT logger) set around each device to monitor the behavior of the nocturnal ship rats. For simplicity, we pool data from all devices and consider results only from one type of camera, which is mounted horizontally on a post 3 meters from each device. Extending to multiple cameras and the PIT logger is readily done, but involves substantial additional detail, so we do not include data from the vertically mounted camera or the PIT logger in the illustrative analysis here. Our aim is to demonstrate in a simplified context how trace-contrast modeling can deliver inference on the parameters of interest.

Video recordings from the horizontally mounted camera last 60 seconds, so the camera is not available to be triggered again until $\ell = 60$ seconds after an initial trigger. All recordings are later watched to verify that they contain footage of a ship rat, and an observation is made of whether or not the recording reveals the rat interacting with the control device ($I = 1$ or $I = 0$). It is inevitable that some interactions will not be seen on the video recording, and rats can move swiftly enough that an interaction can occur without triggering the camera at all: these possibilities are confirmed by examining the PIT records. Occasionally, two or more ship rats can be seen simultaneously in the same video: we deal with this by entering a new record for each rat at a time randomly chosen between 0 and 60 seconds from the start of the recording. Further resolution was not possible given the speed with which rats enter and leave the video frame during recordings, and given the number of videos to be transcribed.

Before the study, some of the rats were trapped and marked with visually recognizable coat markings. Not all rats are marked, and marks are frequently not readable from the video recordings. We shall use this to demonstrate how our framework can accommodate information from partially marked populations.

A data entry, indexed by d, consists of the following details:

- The time t_d at which the recording was triggered, measured in seconds since the beginning of the study. The data $(t_1, t_2, \ldots, t_D)$ correspond to the observed offspring points in the one-dimensional point process.
- I_d, an indicator that specifies whether or not the recording reveals an interaction of the rat with the control device.
- m_d, the individual rat ID number if the rat has a coat marking and it can be positively identified from the video recording.

The contrast data for a pair of recordings i and j consists of the contrast $r = |t_i - t_j|$ and auxiliary information $z = (I_i, I_j, m_i, m_j)$. From now on we shall

consider primarily the contrast data $(r_1, \ldots, r_n)$ and auxiliary information $z_1, \ldots, z_n$ resulting from n contrasts. We create the contrast data using a 1-hour buffer ($R = 3600$ s) around the beginning and end of each night of the study, such that recordings are only used as target points if they are more than R seconds from sunset or sunrise. Contrasts are only made between pairs of recordings from the same night and the same camera location.

We define a parent point to be an encounter of a single rat with a single device, where the definition of "encounter" will be delimited by the clustering in the data as discussed earlier. Each parent possesses status $\mathcal{I}$ or $\overline{\mathcal{I}}$, where $\mathbb{P}(\mathcal{I}) = \alpha$. Status $\mathcal{I}$ indicates that the encounter involves an interaction, but it does not require the interaction to be seen on any video recording. The parameter of interest is α.

3.2 Video-Lag Model

The Neyman–Scott model described previously is not suitable for the camera data because siblings are not independently and identically distributed about a parent. A single detection makes the camera unavailable for new triggers for $\ell = 60$ seconds, so there is serial dependence between the times of different detections during the same encounter (siblings). In view of this, we propose a sequential model for camera detections, which we call the *video-lag model*, described now:

- Invisible parent points (encounters) occur according to a one-dimensional Poisson process with rate μ encounters per second. Suppose for illustration that the parent point arises at time a.
- Each parent is assigned status $\mathcal{I}$ or $\overline{\mathcal{I}}$, where $\mathbb{P}(\mathcal{I}) = \alpha$.
- Parents are also assigned a status $\mathcal{M}$ or $\overline{\mathcal{M}}$ that specifies whether the encounter involves a rat with identifiable coat markings. The probability of a marked rat is defined by the parameter $\tau = \mathbb{P}(\mathcal{M})$.
- Parents produce K offspring where $K \sim \text{Poisson}(\nu)$. The offspring correspond to detected video recordings of the encounter.
- If $K = 0$, there are no offspring. If $K \geq 1$, the first offspring occurs at time $a + Y_1$, where $Y_1 \sim \text{Exponential}(\psi)$. If there are more offspring, the second occurs at time $a + Y_1 + \ell + Y_2$, where $Y_2 \sim \text{Exponential}(\psi)$ and is independent of Y_1, and where the video-lag time ℓ (60 seconds here) is the time that the camera is unavailable for new triggers because it is recording a previous detection. In general, offspring s occurs at time $a + Y_1 + \cdots + Y_s + (s-1)\ell$.
- If the parent is of type $\mathcal{I}$, the interaction is revealed on each of the k recordings independently with probability η.
- If the parent is of type $\mathcal{M}$, the mark is readable on each of the k recordings independently with probability ω.
- A final parameter γ is the probability that a randomly chosen pair of marked rats from different encounters are different rats, and enables use of information from the partially marked population when two recordings correspond to different rats and therefore cannot be part of the same encounter.

3.3 Palm Intensity for the Video-Lag Model

The Palm intensity $\lambda_0(r)$ for one-dimensional processes is specified in Appendix equation (12). For the video-lag model, we find $\lambda_0(r)$ by first deriving $\Lambda_0(r)$ in terms of siblings and nonsiblings. The nonsibling term gives the expected number of detections within an interval of width $2r$ centered on a target point, and is $2r\mu\nu$. This applies for all r due to our treatment of video recordings with more than one rat visible, ensuring that the data do include nonsibling contrasts within time ℓ of each other. The sibling term takes more derivation, and details are omitted for brevity. It is obtained by partitioning over family size $k = 2, 3, \ldots$, because at least $k = 2$ siblings are required to make a sibling contrast, and then for a family of size k partitioning again over the sibling rank difference within a contrast. The sibling rank difference ranges from $s = 1$ for adjacent siblings, to $s = k - 1$ for the difference between siblings 1 and k. Finally, we use the fact that a sum of s independent Exponential(ψ) random variables has the Gamma(s, ψ) distribution, where s is the shape parameter and ψ is the rate parameter. Some algebra gives

$$\lambda_0(r) = \mu\nu + \frac{1}{\nu}\sum_{k=2}^{\infty}\left\{\sum_{s=1}^{k-1}(k-s)\varphi(r - s\ell, s, \psi)\right\}\frac{\nu^k}{k!} \cdot \exp(-\nu), \tag{8}$$

where $\varphi(r - s\ell, s, \psi)$ is the probability density of the Gamma distribution with shape parameter s and rate ψ, evaluated at $r - s\ell$, and for $r > s\ell$ is given by

$$\varphi(r - s\ell, s, \psi) = \frac{\psi^s}{\Gamma(s)}(r - s\ell)^{s-1}e^{-\psi(r-s\ell)}.$$

For $r \leq s\ell$ we have $\varphi(r - s\ell, s, \psi) = 0$, except for the special case $\varphi(0, 1, \psi) = \psi$.

3.4 Auxiliary Information

The auxiliary information z for a single contrast is the quartet (I_i, I_j, m_i, m_j) specifying whether the target and comparison points in the contrast revealed an interaction and identified marked animals. As in Section 2.3, auxiliary information is formulated via $b(r)$, the probability that a randomly selected contrast with magnitude r corresponds to a pair of siblings, where

$$b(r) = \frac{\lambda_0(r) - \mu v}{\lambda_0(r)}, \tag{9}$$

with $\lambda_0(r)$ specified in (8).

The information about rat identity, m_i and m_j, is used to identify cases where the two video recordings in the contrast cannot be siblings. When both rats are marked, both marks are readable, and the two marks differ, the two rats involved in the contrast must be different rats and therefore the contrast must correspond to two different encounters. However, if both rats are marked and the two marks are the same, this does not specify that the recordings are siblings because the same rat might be involved in two different encounters. This is the reason for the parameter γ that gives the probability that, if two different encounters of type $\mathcal{M}$ are chosen at random, the two rats involved are different rats. In this study, the partial marking of the population therefore contributes some information about nonsiblings, but does not contribute any information about siblings. We summarize the information by an indicator J, where $J = 1$ if marks m_i and m_j are both present and differ, and $J = 0$ otherwise.

The final model for auxiliary information, $f(z|r) = f(I_i, I_j, J|r)$, requires expressions for each triple $(I_i, I_j, J) \in \{0, 1\}^3$. By way of example, we give just two of the eight expressions. Each expression partitions over the two possibilities that the detections in the contrast are siblings and nonsiblings, where the probability of siblings is $b(r)$, and we also use the fact that $J = 1$ is impossible for siblings:

$$\begin{aligned} &f(0,0,0|r) \\ &\quad = \{\alpha(1-\eta)^2 + (1-\alpha)\}b(r) + (1 - \tau^2\omega^2\gamma) \\ &\quad\quad \cdot \{\alpha^2(1-\eta)^2 + 2\alpha(1-\alpha)(1-\eta) + (1-\alpha)^2\} \\ &\quad\quad \cdot \{1 - b(r)\}; \\ &f(1,1,1|r) = \tau^2\omega^2\gamma\alpha^2\eta^2\{1 - b(r)\}. \end{aligned} \tag{10}$$

The maximum Palm likelihood estimates for the camera-trap model are obtained by maximizing the objective function (6), where the intensity function for a one-dimensional process is $\phi(r) = 2T\lambda_0(r)$, and where the Palm intensity $\lambda_0(r)$ for the video-lag model is given by (8), and auxiliary information is treated by expressions such as those in (10), with $b(r)$ as given in (9).

3.5 Real Data

The data comprise 2374 video recordings made over 8 consecutive nights in February 2014 from a grid of 34 camera stations in Huapai Reserve. We applied a buffer of $R = 3600$ seconds (1 hour) at each end of each night, leaving $T = 2188$ recordings to act as target points. The total number of contrasts using this value of R is 27,837. The model takes about 15 seconds to fit on a 1.73 GHz laptop. For variance estimation, we bootstrap across stations until the number of contrasts in the bootstrap replicate data is at least 90% of that in the real data.

The maximum Palm likelihood estimates, together with bootstrapped 95% confidence intervals from 1000 replicates, are as follows: average encounter rate per hour, $3600\hat{\mu} = 6.1$ $(2.7, 9.3)$; probability of interaction, $\hat{\alpha} = 0.44$ $(0.31, 0.61)$; expected number of detections per encounter, $\hat{v} = 0.97$ $(0.66, 1.64)$; detection lag rate, $\hat{\psi} = 0.0047$ $(0.0018, 0.0112)$; probability that an interaction is revealed on any recording of a type $\mathcal{I}$ encounter, $\hat{\eta} = 0.68$ $(0.59, 0.81)$; probability a rat is marked, $\hat{\tau} = 0.40$ $(0.27, 0.44)$; probability that the mark of a marked rat is readable on any recording, $\hat{\omega} = 0.67$ $(0.45, 0.73)$; probability that two encounters with marked rats correspond to different rats, $\hat{\gamma} = 0.58$ $(0.46, 0.62)$.

Figure 3 shows the estimated Palm intensity function, $\lambda_0(r)$, and the estimated sibling probability function, $b(r)$. Examples of the corresponding empirical curves from simulated data using the same generating values and of about the same sample size as the real data are also shown. The curves show that the estimated sibling probability drops to 0 at time differences of about 1500 seconds (25 minutes), with only a small probability of siblings beyond 15-minute contrasts.

3.6 Simulation Study

To verify the accuracy of inference using Palm likelihood estimation with the video-lag model, Figure 4 shows the results from 1000 fits using data simulated from the video-lag model, from which contrasts are constructed. The generating parameters for the simulation are the MPLEs from the real data, and the sample sizes approximately match those in the real data. All eight parameters of the model are estimated with no

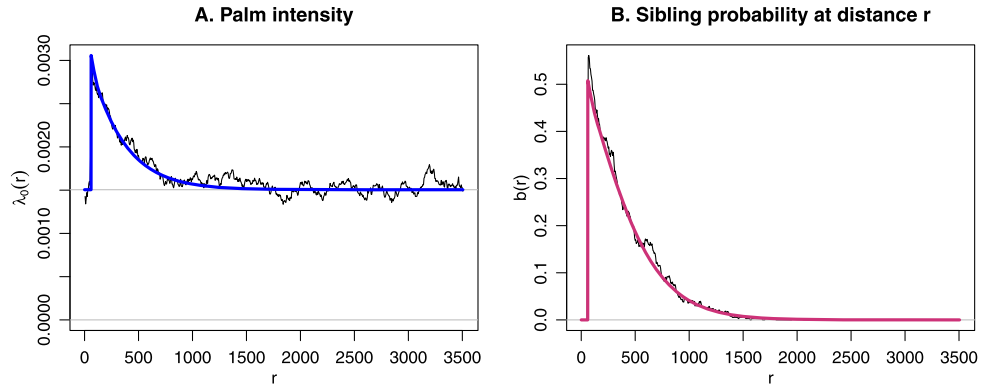

FIG. 3. *Estimated curves for the Palm intensity, $\lambda_0(r)$, and the sibling probability at distance r, $b(r)$, fitted to the real data (bold curves). Empirical results using a simulated data set generated using the same parameter values are shown beneath the curves (thin jagged lines). Empirical results for $b(r)$ can only be calculated using simulated data because sibling knowledge is required.*

discernable bias and with excellent precision. The coefficient of variation is 8.5% for the interest parameter α. In full, the parameters have coefficients of variation 11–12% (μ, ν, ψ); 8–9% (α, η); and 1–2% (τ, ω, γ).

Simulation results are not sensitive to the choice of R, for R ranging from 45 minutes, through 60 mins (shown in Figure 4), to 90 mins, and there is only minor deviation when R is as low as 30 mins. For 100 runs of simulated data, the correlation in $\hat{\alpha}$ between $R = 60$ mins and other choices of R is 0.93 for $R = 30$ mins, and 0.98 to 0.99 for $R = 45, 75$, and 90 mins; and the maximum difference between estimated α values is 0.03, obtained when $R = 30$ mins.

4. DISCUSSION

We have shown that the Palm likelihood estimation framework, first proposed by Tanaka, Ogata and Stoyan (2008) in the context of clustered point processes, opens a promising new direction for drawing inference from capture–recapture studies without needing to construct capture histories. For our example

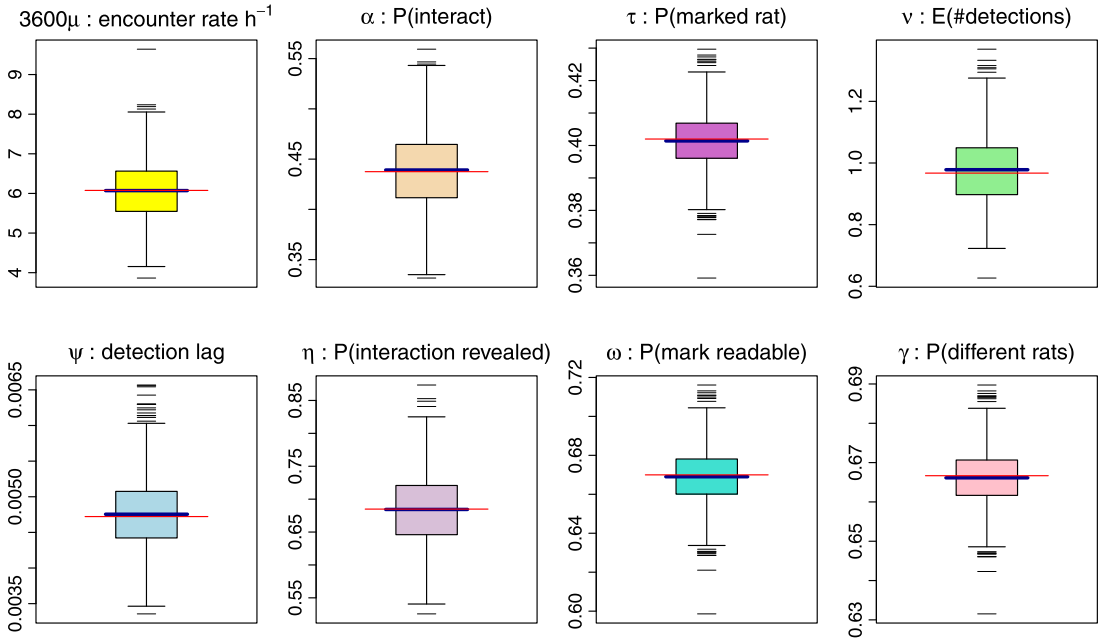

FIG. 4. *Boxplots showing estimates of each of the video-lag model parameters using* 1000 *simulated data sets using the MPLEs from the real data as the generating values. The thin lines across the center of each box show the generating values, and the thick lines show the means of the* 1000 *estimates. Boxes are drawn from the lower to upper quartiles of the* 1000 *estimates.*

settings, inference is accurate and precise, and computationally fast enough that variance estimation by bootstrap is readily achieved. Although there are still many details to explore regarding the robustness of the framework, and how to evaluate it in the absence of a true likelihood function, we propose that the ideas may prove suitable for many kinds of data collected in modern capture–recapture studies. Of additional note is the seamless way in which data on partially marked populations may be accommodated. Partial marking can be used in the framework either to specify that two samples are known matches or, as in our example, known nonmatches.

APPENDIX: DERIVATIONS AND REFERENCE

Here we derive the general Palm intensity formulation for one-dimensional and two-dimensional processes. We give explicit derivations for the one-dimensional and two-dimensional Neyman–Scott processes.

A.1 Palm Intensity Definition in One and Two Dimensions

In any dimension, the Palm intensity is defined as the expected number of points per unit area located at distance r from a single target point. For a process in any dimension, define

$$\Lambda_0(r) = \mathbb{E}(\text{number of further points located distance} \le r \text{ from one target point}).$$

For a two-dimensional process, the area of a thin ring of width δr at distance r around the target is approximately $2\pi r\delta r$, so the Palm intensity $\lambda_0(r)$ is defined as

$$\begin{aligned}\lambda_0(r)^{2\mathrm{D}} &= \lim_{\delta r\to 0}\frac{\Lambda_0(r+\delta r)-\Lambda_0(r)}{2\pi r\delta r}\\ &= \frac{1}{2\pi r}\frac{d}{dr}\Lambda_0(r).\end{aligned} \tag{11}$$

For a one-dimensional process, an interval of width δr located distance r from either side of the target has total width $2\delta r$, so the Palm intensity is

$$\begin{aligned}\lambda_0(r)^{1\mathrm{D}} &= \lim_{\delta r\to 0}\frac{\Lambda_0(r+\delta r)-\Lambda_0(r)}{2\delta r}\\ &= \frac{1}{2}\frac{d}{dr}\Lambda_0(r).\end{aligned} \tag{12}$$

The intensity needed in (4) or (6) is $\phi(r)^{1\mathrm{D}} = 2T\lambda_0(r)^{1\mathrm{D}}$ or $\phi(r)^{2\mathrm{D}} = 2\pi rT\lambda_0(r)^{2\mathrm{D}}$.

A.2 Sibling Contrasts for IID Siblings

Suppose that, for any single family, the locations of the offspring about the parent are independent and identically distributed. Let F be the cumulative distribution function (CDF) for inter-sibling distances:

$$F(r) = \mathbb{P}(\text{distance between a randomly chosen pair of siblings is} \le r).$$

As a preliminary, we derive the expected number of siblings within distance r of a randomly selected target point, which is the component of $\Lambda_0(r)$ for sibling contrasts. The following derivation holds for point processes in any dimension.

We can consider the selection of a random target point to be equivalent to selecting a family at random and then examining one point within the family. Taking the family perspective is an important step because families of size k have k times as many chances to be selected for targets as families of size 1. In practice, we do not know the family membership or family sizes of any of the target points, but it is convenient to conceptualize the selection of a target as operating via its family.

Suppose therefore that we select a single family from all possible families in the point process, where the family size distribution is given by the random variable K with probabilities $\mathbb{P}(K=k)$, and the selection probability for a family of size k is k times the selection probability for a family of size 1, for $k = 0, 1, 2, \ldots$. Over the sample space of families,

$$\begin{aligned}&\mathbb{P}(\text{randomly selected family has size } k)\\ &\quad= \frac{k\mathbb{P}(K=k)}{\sum_{h=0}^{\infty} h\mathbb{P}(K=h)} = \frac{k\mathbb{P}(K=k)}{\mathbb{E}(K)}.\end{aligned} \tag{13}$$

Having selected a family of size $k \ge 1$, we select any of the k identical points to be the target. By assumption, it has $k-1$ siblings, each of which lies within distance r with probability $F(r)$. Thus,

$$\mathbb{E}(\#\text{siblings distance} \le r \text{ from target} \mid \text{family size} = k) = (k-1)F(r). \tag{14}$$

Partitioning over the selected family size, and using (13) and (14), gives

$$\begin{aligned}&\mathbb{E}(\#\text{siblings} \le r \text{ from randomly-selected target})\\ &= \sum_{k=1}^{\infty}(k-1)F(r)\left\{\frac{k\mathbb{P}(K=k)}{\mathbb{E}(K)}\right\}\\ &= \frac{\mathbb{E}\{K(K-1)\}F(r)}{\mathbb{E}(K)}.\end{aligned} \tag{15}$$

A.3 Two-Dimensional Neyman–Scott Process

We have defined the two-dimensional Neyman–Scott process as follows:

- Parent points arise according to a two-dimensional Poisson process with intensity μ;
- Families comprise $K \sim \text{Poisson}(\nu)$ offspring per parent;
- For a parent located at point $\mathbf{a}$, offspring locations are independent Gaussian$(\mathbf{a}, \sigma^2 I_2)$ variates. (This final condition defines a Thomas process.)

Clearly, under these assumptions the expected number of nonsiblings within distance r of any target point is $\pi r^2 \mu\nu$. The expected number of siblings within distance r is given by (15), and $\Lambda_0(r)$ is the sum of the nonsibling and sibling terms.

For $K \sim \text{Poisson}(\nu)$, the computations needed in (15) are $\mathbb{E}(K) = \nu$ and $\mathbb{E}\{K(K-1)\} = \nu^2$. It remains to find $F(r)$, the CDF of distances between siblings.

Given a parent at $\mathbf{0}$, suppose two siblings are placed at positions $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$. By assumption, $x_1 \sim x_2 \sim y_1 \sim y_2 \sim \text{Normal}(0, \sigma^2)$, and all four components are independent. The distance between them is $\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$, where $(x_1 - y_1) \sim (x_2 - y_2) \sim \text{Normal}(0, 2\sigma^2)$. This distance is readily seen to be a variate from the $\sigma\sqrt{2}\,\text{Chi}(2)$ distribution. Transforming the Chi(2) density gives the required expression for $F'(r)$ for $r \geq 0$:

$$\frac{d}{dr}F(r) = \frac{r}{2\sigma^2}\exp\left(-\frac{r^2}{4\sigma^2}\right). \tag{16}$$

Using (11), adding nonsiblings and siblings to obtain $\Lambda_0(r)$, and combining (15) and (16) give

$$\begin{aligned}\lambda_0(r) &= \frac{1}{2\pi r}\frac{d}{dr}\Lambda_0(r)\\ &= \frac{1}{2\pi r}\frac{d}{dr}\left[\pi r^2\mu\nu + \frac{\mathbb{E}\{K(K-1)\}F(r)}{\mathbb{E}(K)}\right]\\ &= \mu\nu + \frac{1}{2\pi r}\left\{\frac{\nu^2 F'(r)}{\nu}\right\}\\ &= \mu\nu + \frac{\nu}{4\pi\sigma^2}\exp\left(-\frac{r^2}{4\sigma^2}\right),\end{aligned}$$

as given previously in equation (2). For a two-dimensional process, the intensity function for Palm likelihood estimation using (4) or (6) is $\phi(r) = 2\pi r T\lambda_0(r)$.

A.4 One-Dimensional Neyman–Scott Process

The one-dimensional Neyman–Scott process is defined like the two-dimensional process in Section A.3, except that the parent points follow a one-dimensional Poisson process with intensity μ, and the offspring locations for a parent at position a are independent Gaussian(a, σ^2) variates. Inter-sibling distances are now $\sigma\sqrt{2}\,\text{Chi}(1)$ variates, giving for $r \geq 0$,

$$\frac{d}{dr}F(r) = \frac{1}{\sqrt{\pi\sigma^2}}\exp\left(-\frac{r^2}{4\sigma^2}\right). \tag{17}$$

Using (12), adding nonsiblings and siblings to obtain $\Lambda_0(r)$, and combining (15) and (17) give

$$\begin{aligned}\lambda_0(r) &= \frac{1}{2}\frac{d}{dr}\Lambda_0(r)\\ &= \frac{1}{2}\frac{d}{dr}\left[2r\mu\nu + \frac{\mathbb{E}\{K(K-1)\}F(r)}{\mathbb{E}(K)}\right]\\ &= \mu\nu + \frac{\nu}{\sqrt{4\pi\sigma^2}}\exp\left(-\frac{r^2}{4\sigma^2}\right).\end{aligned}$$

For a one-dimensional process, the intensity function for Palm likelihood estimation using (4) or (6) is $\phi(r) = 2T\lambda_0(r)$.

ACKNOWLEDGMENTS

We thank Helen Nathan who conducted the ship rat camera study and supplied the data. Thanks to the Editors and two anonymous referees whose constructive comments greatly improved the paper. The ship rat study was funded by a NZ Ministry of Science and Innovation grant to Landcare Research. This work was funded by the Royal Society of New Zealand through Marsden Grant 14-UOA-155. Ben Stevenson was supported by EPSRC/NERC Grant EP/1000917/1.

REFERENCES

BADDELEY, A. and TURNER, R. (2005). Spatstat: An R package for analyzing spatial point patterns. *J. Stat. Softw.* **12** 1–42.

BARKER, R. J., SCHOFIELD, M. R., WRIGHT, J. A., FRANTZ, A. C. and STEVENS, C. (2014). Closed-population capture–recapture modeling of samples drawn one at a time. *Biometrics* **70** 775–782. MR3295738

BELL, E. A., BELL, B. D. and MERTON, D. V. (2016). The legacy of Big South Cape: Rat irruption to rat eradication. *NZ J. Ecol.* **40** 205–211.

BORCHERS, D. L. and EFFORD, M. G. (2008). Spatially explicit maximum likelihood methods for capture–recapture studies. *Biometrics* **64** 377–385, 664. MR2432407

BORCHERS, D. L., 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 Ecol. Evol.* **5** 656–665.

CARROLL, E. L., PATENAUDE, N. J., CHILDERHOUSE, S. J., KRAUS, S. D., FEWSTER, R. M. and BAKER, C. S. (2011). Abundance of the New Zealand subantarctic southern right whale population estimated from photo-identification and genotype mark-recapture. *Mar. Biol.* **158** 2565–2575.

CHARLTON, B. D., ELLIS, W. A. H., MCKINNON, A. J., BRUMM, J., NILSSON, K. and FITCH, W. T. (2011). Perception of male caller identity in koalas (Phascolarctos cinereus): Acoustic analysis and playback experiments. *PLoS ONE* **6** e20329.

CORMACK, R. (1964). Estimates of survival from the sighting of marked animals. *Biometrika* **51** 429–438.

DIGGLE, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*, 2nd ed. Arnold, London.

FRETWELL, P. T., STANILAND, I. J. and FORCADA, J. (2014). Whales from space: Counting southern right whales by satellite. *PLoS ONE* **9** e88655.

GUAN, Y. (2006). A composite likelihood approach in fitting spatial point process models. *J. Amer. Statist. Assoc.* **101** 1502–1512. MR2279475

JOLLY, G. M. (1965). Explicit estimates from capture–recapture data with both death and immigration-stochastic model. *Biometrika* **52** 225–247. MR0210227

KÜHL, H. S. and BURGHARDT, T. (2013). Animal biometrics: Quantifying and detecting phenotypic appearence. *Trends Ecol. Evol.* **28** 432–441.

MØLLER, J. and WAAGEPETERSEN, R. P. (2007). Modern statistics for spatial point processes. *Scand. J. Stat.* **34** 643–684. MR2392447

NATHAN, H. W. (2016). Detection probability of invasive ship rats: Biological causation and management implications. Ph.D. thesis, Univ. Auckland, New Zealand.

NEYMAN, J. and SCOTT, E. L. (1958). Statistical approach to problems of cosmology. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **20** 1–43. MR0105309

PALM, C. (1943). Intensitätsschwankungen im Fernsprechverkehr. *Ericsson Technics* **44** 1–189. MR0011402

PROKEŠOVÁ, M. and JENSEN, E. B. V. (2013). Asymptotic Palm likelihood theory for stationary point processes. *Ann. Inst. Statist. Math.* **65** 387–412. MR3011627

ROBINS, J. H., MILLER, S. D., RUSSELL, J. C., HARPER, G. A. and FEWSTER, R. M. (2016). Where did the rats of Big South Cape Island come from? *NZ J. Ecol.* **40** 229–234.

SEBER, G. A. F. (1965). A note on the multiple-recapture census. *Biometrika* **52** 249–259. MR0210228

STEVENSON, B. C., BORCHERS, D. L., ALTWEGG, R., SWIFT, R. J., GILLESPIE, D. M. and MEASEY, G. J. (2015). A general framework for animal density estimation from acoustic detections across a fixed microphone array. *Methods Ecol. Evol.* **6** 38–48.

TABERLET, P. and LUIKART, G. (1999). Non-invasive genetic sampling and individual identification. *Biol. J. Linn. Soc.* **68** 41–55.

TANAKA, U., OGATA, Y. and STOYAN, D. (2008). Parameter estimation and model selection for Neyman–Scott point processes. *Biom. J.* **50** 43–57. MR2414637

THOMAS, M. (1949). A generalization of Poisson's binomial limit for use in ecology. *Biometrika* **36** 18–25. MR0033999

VALE, R. T. R., FEWSTER, R. M., CARROLL, E. L. and PATENAUDE, N. J. (2014). Maximum likelihood estimation for model $M_{t,\alpha}$ for capture–recapture data with misidentification. *Biometrics* **70** 962–971. MR3295757

WAAGEPETERSEN, R. P. (2007). An estimating function approach to inference for inhomogeneous Neyman–Scott processes. *Biometrics* **63** 252–258, 315. MR2345595

WAAGEPETERSEN, R. and SCHWEDER, T. (2006). Likelihood-based inference for clustered line transect data. *J. Agric. Biol. Environ. Stat.* **11** 264–279.

WRIGHT, J. A., BARKER, R. J., SCHOFIELD, M. R., FRANTZ, A. C., BYROM, A. E. and GLEESON, D. M. (2009). Incorporating genotype uncertainty into mark-recapture-type models for estimating abundance using DNA samples. *Biometrics* **65** 833–840. MR2649856