A hierarchical distance sampling approach to estimating mortality rates from opportunistic carcass surveillance data

Steve E. Bellan[1]*, Olivier Gimenez[2], Rémi Choquet[2] and Wayne M. Getz[1,3]

[1]*Department of Environmental Science, Policy and Management, University of California, 137 Mulford Hall, Berkeley, CA, 94720, USA;* [2]*Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175, Campus CNRS, 1919 Route de Mende, Montpellier Cedex 5, 34293, France; and* [3]*School of Mathematical Sciences, University of KwaZulu-Natal, Durban, South Africa*

Summary

1. Distance sampling is widely used to estimate the abundance or density of wildlife populations. Methods to estimate wildlife mortality rates have developed largely independently from distance sampling, despite the conceptual similarities between estimation of cumulative mortality and the population density of living animals. Conventional distance sampling analyses rely on the assumption that animals are distributed uniformly with respect to transects and thus require randomised placement of transects during survey design. Because mortality events are rare, however, it is often not possible to obtain precise estimates in this way without infeasible levels of effort. A great deal of wildlife data, including mortality data, are available via road-based surveys. Interpreting these data in a distance sampling framework requires accounting for the non-uniformity of sampling. In addition, analyses of opportunistic mortality data must account for the decline in carcass detectability through time. We develop several extensions to distance sampling theory to address these problems.

2. We build mortality estimators in a hierarchical framework that integrates animal movement data, surveillance effort data and motion-sensor camera trap data, respectively, to relax the uniformity assumption, account for spatiotemporal variation in surveillance effort and explicitly model carcass detection and disappearance as competing ongoing processes.

3. Analysis of simulated data showed that our estimators were unbiased and that their confidence intervals had good coverage.

4. We also illustrate our approach on opportunistic carcass surveillance data acquired in 2010 during an anthrax outbreak in the plains zebra of Etosha National Park, Namibia.

5. The methods developed here will allow researchers and managers to infer mortality rates from opportunistic surveillance data.

Key-words: carcass, cue, disease, distance sampling, hierarchical model, mortality, opportunistic surveillance

Introduction

Distance sampling is a common class of methods used to estimate abundance of wildlife populations (Buckland, Goudie & Borchers 2000). In conventional distance sampling (CDS), a region is sampled from randomly placed lines (line transect sampling) or points (point transect sampling) with n detected animals counted and their respective distances, y_i, $i = 1, \ldots n$, to the traversed line or point recorded. The distribution of these distances is then used to estimate the decline in an animal's detection probability as a function of increasing distance from the observer. The fitted detection function then facilitates estimation of animal abundance or density in the surveyed region, and the precision thereof (Buckland *et al.* 2001).

One of the strengths of CDS is that the random placement of transects in the study area (i.e. design-based surveys) supports two assumptions. First, the surveyed area is assumed to be a random (i.e. unbiased) sample of the larger study area (between-transect scale), allowing extrapolation of density estimates from the former to the latter. Second, the perpendicular distance of animal locations to the survey transects is assumed to be uniform (within-transect scale). We use $\pi(y)$ to denote the probability density distribution of an independent and identically distributed random variable y and $\pi(y_i)$ to denote the probability density at a particular distance y_i. If $\pi(y)$ is uniform, then any drop-off in the expected number of animals detected at greater distances is due to declining detectability, rather than changes in animal density.

Both assumptions above may not hold for distance sampling-type data collected from non-randomly located transects or paths. Violations of the assumption that the surveyed area is

*Correspondence author. E-mail: steve.bellan@gmail.com

not a random sample of the study area can be accounted for by modelling population density as a spatially explicit function of habitat covariates (Hedley & Buckland 2004; Williams, Hedley & Hammond 2006; Johnson, Laake & Ver Hoef 2010). Without transect randomisation, however, accounting for bias caused by violations of the assumption that $\pi(y)$ is uniform can pose a challenge (Johnson, Laake & Ver Hoef 2010; Marques *et al.* 2010). Violation of the uniformity assumption will commonly occur in surveys conducted from easily navigable permanent paths that either indirectly (e.g. via association with habitat variables) or directly (e.g. road avoidance) affect animal behaviour. Without knowing how a species' utilisation varies with distance from transects, results will be biased (Johnson, Laake & Ver Hoef 2010); although see Marques *et al.* (2010) for an alternative approach to disentangling detectability from $\pi(y)$. Therefore, from an analytical point of view, most CDS literature holds that randomisation of transect location is indeed necessary (Buckland *et al.* 2001, 2004).

Opportunistic surveys – i.e. those in which detections are recorded while observers are performing other tasks – have some important advantages over design-based surveys, suggesting a need for methods that can account for the bias associated with non-uniform $\pi(y)$. These advantages include: (1) the relatively low-cost of acquiring distance data from paths already being traversed for other reasons (Williams, Hedley & Hammond 2006; Kiszka *et al.* 2007) or from easily navigable paths (Walsh & White 1999), (2) the relative ease of collecting opportunistic data long-term (Kiszka *et al.* 2007; Himes Boor & Small 2012), and (3) avoidance of the problem that, when studying extremely rare or elusive animals, resource limitations may prevent feasibly sized design-based studies from detecting enough animals for informative inference.

Opportunistic surveys may also be preferable when counting rare events, such as sighting rare or elusive animals, or finding carcasses of even common animals. In the latter case, carcasses may be detectable for such short durations that an unfeasibly large design-based survey would be necessary to detect enough carcasses for adequate precision. Yet opportunistically sighted carcasses are often recorded in long-term data sets, often along with cause of death, thereby enhancing our understanding of a species' mortality dynamics. In this manuscript, we extend CDS through the development of new methods to incorporate estimates of non-uniform $\pi(y)$ from auxiliary Global Positioning System (GPS) movement data in distance sampling estimators. In addition, we also extend CDS to incorporate data on rates at which scavengers dispose of carcasses into estimates of mortality rates. If carcasses are quickly removed from the environment (i.e. by decay or consumption), a smaller proportion of carcasses will be detected. Consequently, researchers have paid much attention to carcass removal rates when estimating wildlife mortality due to wind farms (Smallwood *et al.* 2010), roads (Santos, Carvalho & Mira 2011), pesticides (Rivera-Milán, Zaccagnini & Canavelli 2009) and power lines (Ponce *et al.* 2010). Using the 'multiplier' approach (Buckland *et al.* 2001; Buckland *et al.* 2004), one can then estimate the mortality rate by dividing the estimated carcass abundance (from distance sampling analysis) by

the estimated duration for which a carcass is detectable (equivalent to multiplying by the estimated removal rate), taking care to incorporate variance due to the latter in the former (Plumptre 2000).

The above 'multiplier' method, however, is invalid when carcasses have multiple chances to be detected, but can only be detected once – i.e. during opportunistic surveillance when multiple trips may be made past a carcass, but communication amongst researchers ensures no double sampling. In such situations, detections are conditional on previous non-detection. The probability of detecting a carcass on one of the several trips is a function of the probability the carcass was available for detection at each trip (and hence on the removal rate), making the detection probability of each carcass a nonlinear function of the number of trips past that carcass, the interval between trips and the removal rate. Thus, our second extension to CDS is the explicit inclusion of detection and removal as competing processes within distance sampling estimators.

Finally, in some systems 'removal rates', which implicitly assume carcass removal to be a discrete event, may not be the relevant concept. For instance, detection of large terrestrial mammal carcasses occurs either by detection of the carcass itself or via detection of various scavenger species, each of which may be more or less detectable depending on size and capability for flight (e.g. large numbers of vultures in flight can be seen at great distances). Thus, carcass availability for detection depends on the probability of scavenger (i.e. sighting cue) presence as a function of time since carcass production (i.e. death). These probabilities will differ between scavenger species based on their abundance, search efficiency and niche partitioning (Hunter, Durant & Caro 2007). Hence, rather than modelling a 'removal' process, we model the sighting cue process itself. Specifically, we estimate $\pi(c|t)$: the probability each sighting cue, c, is available as a function of time since death, t, from motion-sensor camera trap data on scavenger activity at carcasses.

In summary, we address several gaps in methods used to estimate cumulative mortality incidence from opportunistic surveillance data. As an example of the methods developed in this study, we estimate cumulative mortality during outbreaks of seasonally endemic anthrax in the plains zebra (*Equus quagga*) of Etosha National Park (ENP), Namibia. Using a hierarchical modelling framework (Royle & Dorazio 2008), we model carcass production, sighting cue availability and detection as concurrent dynamic processes. Our analysis explicitly accounts for surveillance effort by estimating mortality rates within surveilled space-time windows. We use bootstrap methods to incorporate error associated with estimation of $\pi(y)$ and $\pi(c|t)$ into the final incidence estimate. We first present a technical section extending distance sampling methods, as motivated above. We then introduce the ENP study system, focusing on the observational and ecological processes that play a role in producing the passive surveillance data to be analysed. We continue by using simulated carcass data to assess the accuracy and precision of the developed estimator, before applying it to estimate cumulative mortality in the surveilled region during an anthrax outbreak in ENP in 2010.

Finally, we conclude with a discussion of the utility of these methods as well as suggest future directions.

Materials and methods

A GENERAL LIKELIHOOD FOR CARCASSES

Following the notation in Buckland *et al.* (2001), in CDS the abundance of N animals within half-strip width w of surveyed transects is estimated by the Horvitz-Thompson Aestimator $\widehat{N} = \frac{n}{\int_0^w g(y|\widehat{\theta})\pi(y)dy}$, where n is th number of detected animals, $g(y|\theta)$ is the probability of detecting an animal at a perpendicular distance y from the transect, and θ is the parameter set for this function. To allow robust estimation of N, the detection function $g(y)$ must have a shoulder at the transect ($g'(0) = 0$) and detection must be perfect on the centreline ($g(0) = 1$). Throughout the remainder of the article, we focus on the half-normal detection function, a common such model with these properties,

$$g(y|\theta) = \exp\left(-\left(\frac{y}{\theta}\right)^2\right)$$

The denominator of $\widehat{N}$ is the marginal probability of detecting any animal within a distance w of the transect. We do not know $g(y \mid \theta)$, but rather estimate $\widehat{\theta}$ by maximising the likelihood of the distribution of detected distances y_i given detection: i.e.

$$L(\theta|y) = \prod_{i=1}^{n} \frac{g(y_i|\theta)\pi(y_i)}{\int_0^w g(y|\theta)\pi(y)dy} \qquad \text{eqn 1}$$

Consider an opportunistic surveillance data set in which carcass observations were recorded for $i = 1, \ldots, n$ carcasses that were generated on day d_i at distance y_i from road r_i and sighted with sighting cue c_i. In some systems, there may be only one type of sighting cue – i.e. the carcass itself – but in some systems, such as our motivating example below, there may exist a set of cues with very different detectability-distance drop-offs and durations of time available for detection (e.g. avian scavengers, mammalian scavengers or the carcass itself). In these cases, the likelihood (Eqn 1) can be extended with the probability that carcass i was generated on day d_i at distance y_i from road r_i with cue c_i, given that the carcass was detected. The likelihood for all carcasses found is the product of individual likelihoods of each carcass and thus be written as

$$L_1(\theta|y, c, r, \mathrm{d}) = \prod_{i=1}^{n} \frac{g(y_i, c_i, r_i d_i|\theta)\pi(y_i, c_i, r_i, d_i)}{\Sigma_{(r,d)\in(R,D)} \int_0^w \Sigma_{c\in C} g(y, c, r, d|\theta)\pi(y, c, r, d)dy} \qquad \text{eqn 2}$$

where (R, D) is the spatiotemporal window being considered, C is the set of all detection cues and $\pi(y, c, r, d)$ is the joint probability density function of a carcass being generated y distance from the road on day d, road r, and being observed with cue c. We assume that, d, and c are independent although d and r are likely correlated because particular roads may have elevated carcass densities at certain times as a result of correlated movements of individuals. Thus, we assume that $\pi(y, c, r, d) = \pi(y)\pi(c)\pi(r, d)$. While we do not actually know the day of death d_i, but only the day of detection l_i, for many surveillance systems (i.e. because ageing carcasses is difficult) we assume that the former is known for now, but expand the likelihood with a latent variable formulation below. Noting that the denominator of Eqn 2 is the marginal probability of finding any carcass in (R, D), we can construct a Horvitz-Thompson estimator of carcass abundance in that window.

$$\widehat{N}_{ST} = \frac{n}{\Sigma_{(r,d)\in(R,D)} \int_0^w \Sigma_{c\in C} g(y, c, r, d|\widehat{\theta})\pi(y)\pi(c)\pi(r, d)dy} \qquad \text{eqn 3}$$

In the above formulation, $\pi(r, d)$ explicitly accounts for the spatiotemporal (ST) distribution of carcasses. In some ways, this formulation is attractive. In theory it allows us to fit $\pi(r, d)$ and thereby estimate not only the abundance of carcasses but also their spatiotemporal distribution, given a sufficiently large sample size. However, our initial exploratory simulations demonstrated that this formulation has the disadvantage that surveillance effort that has not resulted in the detection of carcasses affects the likelihood and subsequently the estimation of θ. When the majority of surveillance effort resulted in zero detections (often the case for carcass surveillance), including this effort in the analysis biased $\widehat{\theta}$ and subsequently $\widehat{N}_{ST}$. We therefore decided to develop an estimator that was conditional on the particular road r_i where the carcass was found, the date d_i it was generated (i.e. date of death) and the distance y_i. The likelihood is then the conditional probability of y_i and c_i given detection of a carcass at (r_i, d_i):

$$L_2(\theta|y, c, r, d) = \prod_{i=1}^{n} \frac{g(y_i, c_i, r_i, d_i|\theta)\pi(y_i)\pi(c_i)}{\int_0^w \Sigma_{c\in C} g(y, c, r_i, d_i|\theta)\pi(y)\pi(c)dy} \qquad \text{eqn 4}$$

The denominator of (3) is no longer the marginal detection probability, but rather the conditional probability of detecting a carcass given r_i and d_i and is thus specific to each detected carcass. Thus, following Borchers *et al.* (1998), rather than dividing n by the marginal probability of detecting, we formulate a Horvitz-Thompson-like estimator which sums the inverse detection probabilities of each carcass detected $\widehat{N}_C = \Sigma_{i=1}^{n} \frac{1}{\hat{p}_i}$, where $\hat{p}_i$ is the average probability of detecting a carcass on road r_i that was detected on day d_i. The detection cue changes over time so $\hat{p}_i$ encompasses the probability of detection with any cue, yielding:

$$\widehat{N}_{HT} = \Sigma_{i=1}^{n} \frac{1}{\int_0^w \Sigma_{c\in C} g(y, c, r_i, d_i|\theta)\pi(y)\pi(c)dy} \qquad \text{eqn 5}$$

Our simulations also demonstrated that this estimator also was biased due to integration over both y and the latent variable d_i when surveillance effort was sparse (Supplementary Table S1). We found that a slightly modified Horvitz-Thompson estimator that also conditioned on y_i was approximately unbiased, so we thus use the following estimator throughout the rest of the manuscript

$$\widehat{N}_C = \Sigma_{i=1}^{n} \frac{1}{\Sigma_{c\in C} g(y_i, c, r_i, d_i|\widehat{\theta})\pi(c)} \qquad \text{eqn 6}$$

THE DETECTION FUNCTION

The numerator of Eqn 4 gives the probability a given carcass was detected at y_i from the road with cue c_i given death on road r_i on day d_i. For a known date of death d_i

$$g(y_i, c_i, r_i, d_i)\pi(y_i)\pi(c_i) = \pi(y_i)\Sigma_{q\in Q} g^*(y_i, c_i, r_i, d_i, t_{q,i})\pi(c_i|t_{q,i}) \qquad \text{eqn 7}$$

where $t_{q,i} \equiv l_{q,i} - d_i$ is the number of days between death and the q-th trip past the i-th carcass and $l_{q,i}$ is the calendar day of this trip. We define $g^*(y_i, c_i, r_i, d_i, t_{q,i})$ as the probability of detection on exactly the q-th trip with cue c_i, conditional on previous non-detection. The cumulative probability of detection within t_{max} days (i.e. the maximum number of days after death that we believe a carcass is detectable, chosen based on knowledge of the study system) of d_i is calculated by summing over this function for all trips in Q_i, where Q_i is defined such that $q \in Q_i$

if $t_{q,i} \in [0, t_{max}]$. The probability of detection exactly $t_{q,i}$ days after death with cue c_i is the probability of missing the carcass on all trips to road r_i after the carcass was generated, but prior to the trip of detection multiplied by the probability of detecting it on trip q: i.e.

$$g^*(y_i, c_i, r_i, d_i, t_{q,i})\pi(c_i|t_{q,i}) = h(y_i, c_i)\pi(c_i|t_{q,i}) \prod_{s \in S_{q,i}}$$
$$(1 - \Sigma_{c \in C} h(y_i, c)\pi(c|t_s)) \qquad \text{eqn 8}$$

where $h(y_i, c_i)$ is the distance detectability function (equivalent to $g(y, z)$ used in the CDS notation above) giving the probability of detecting a carcass displaying sighting cue c_i given the carcass is y_i distance from the road. The product occurs over all trips in $S_{q,i}$, where $S_{q,i}$ is defined such that $s \in S_{q,i}$ if $t_s \in [0, t_{q,i}]$ and trip s occurred before trip q (when they were on the same day).

The function $h(y_i, c_i)$ is the detection function in the CDS sense in that, it is the probability of detecting a carcass y_i metres from the road given that the carcass is presenting sighting cue c_i. We use the half-normal detection function with a scale parameter for each cue type

$$h(y_i, c_i) = \exp\left[-\left(\frac{y_i}{\sigma_{c_i}}\right)^2\right].$$

While for simplicity, we do not consider covariates other than cue type here, other covariates can easily be included in the model using the following modification for M covariates $(Z_i, ..., Z_M)$ thought to influence detection: $h(y_i, c_i) = \exp\left(-\left(\frac{y_i}{\sigma_i}\right)^2\right)$ where $\sigma_i = \sigma_{c_i} \times \exp(\Sigma_{m=1}^M \beta_m z_{m,i})$, and β_m is the parameter determining the effect of the covariate z_m. We could imagine increased precision in the estimator if, for example, we included covariates such as driver, number of passengers in the car or habitat type.

INCORPORATING A LATENT VARIABLE FOR THE DATE OF DEATH

In many systems we rarely know the date of death d_i for detected carcasses. Consequently, we model d_i as a latent variable, whose distribution was informed by the observed sighting cue, the probability a given sighting cue is available as a function of time since death, and the recent history of surveillance effort. For instance, if a carcass is detected by a cue that is only available for 3 days after death we can be reasonably sure the carcass is less than three days old. More formally, we define $\pi(t_i)$ as the probability that a carcass is t_i days old. We allow t_i to take integer values in the interval $[0, t_{max}]$ and sum over the latent variable d_i

$$g(y_i, c_i, r_i, d_i)\pi(c_i) = \Sigma_{d_i=l_i-t_{max}}^{l_i} \left(\pi(t_i|c_i)\Sigma_{q \in Q}(g(y_i, c_i, r_i, d_i, t_{q,i})\pi(c_i|t_{q,i}))\right)$$
$$\text{eqn 9}$$

where we define $t_i \equiv l_i - d_i$ as the number of days between death and detection (for each potential value of d_i). We can express $\pi(t|c)$ as the probability of detection exactly t_i days after death with cue c_i divided by the marginal probability of detection with cue c_i

$$\pi(t|c_i) = \frac{g(y_i, c_i, r_i, d_i, t)\pi(c_i|t)}{\Sigma_{d_i=l_i-t_{max}}^{l_i} g(y_i, c_i, r_i, d_i, t_i)\pi(c_i|t_i)} \qquad \text{eqn 10}$$

OPTIMISATION AND INTERVAL ESTIMATION

The conditional probability of detecting a carcass at (r_i, d_i) (the denominator in the conditional likelihood given by Eqn 4) requires calculation of an analytically intractable integral. To increase computational efficiency, we use a rectangular quadrature approach to discretize this integral. Likelihood maximisation of Eqn 4 was performed using the R function optim. All scripts needed to reproduce this analysis are included in the Online Supplementary Material.

We formulated both parametric and nonparametric bootstrap confidence intervals and compared their bias and coverage with simulated data (described below). To build parametric bootstrap confidence intervals for the estimated parameters, we invert the Hessian matrix of the likelihood to estimate the covariance-variance matrix $\hat{\Sigma}$. Given the maximum likelihood parameter estimate vector $\hat{\theta}$ and $\hat{\Sigma}$, we draw 10,000 random parameter sets $\hat{\theta}_j$ from the multivariate normal distribution $N(\hat{\theta}, \hat{\Sigma})$ and calculate $\hat{N}_{C,j}$ for each $\hat{\theta}_j$. Confidence intervals are then constructed from the appropriate quantiles of the empirical distribution function of the sample $\hat{N}_{C,j}$.

Nonparametric bootstrapping is commonly used to build robust confidence intervals in line transect methods. Bootstrapping is usually performed by resampling individual transects in a multi-transect survey because transects are generally sufficiently far apart in space or time to assume independence. In the situation of ongoing opportunistic surveillance, there are no well-defined sampling units. Given that no theoretically sound sampling unit was available, for simplicity, we choose to bootstrap over detected carcasses and their corresponding history of surveillance effort. We maximise the likelihood and calculate $\hat{N}_{C,j}$ for each of 1,000 bootstrap samples and construct 95% confidence intervals as above.

To maintain simplicity, we estimate $\pi(y)$ and $\pi(c|t)$ from auxiliary data sets prior to maximising the likelihood rather than simultaneously estimating these distributions along with the detection function parameters. Thus, these distributions were fixed during estimation of $\hat{N}_{C,j}$. To assess how error associated with estimating these distributions affected the interval estimation for the ENP data set, we also resampled the auxiliary data sets used to estimate these distributions during the nonparametric bootstrap resampling above and used the resulting $\pi_j(y)$ and $\pi_j(c|t)$ distributions, respectively, when maximising the likelihood in the j-th bootstrap run. This approach allows uncertainty to percolate through into the cumulative incidence estimates without increasing computational complexity.

INTRODUCTION TO THE STUDY SYSTEM

Anthrax is a fatal disease of mammals caused by the bacterium *Bacillus anthracis* and causes a significant burden of mortality in livestock and wild herbivores World-wide (Hugh-Jones & de Vos 2002). *B. anthracis* is an environmentally transmitted pathogen with animals infected after being exposed to sufficiently large doses of spores in soil, water or food contaminated by a carcass that previously died of infection. In ENP, anthrax is seasonally endemic in plains ungulates and elephants with the highest observed mortalities generally occurring in the plains zebra and during the end of the wet season (Lindeque & Turnbull 1994). Mortalities are generally observed on the central Okaukuejo plains of ENP where large herds of zebra graze during the wet season (Fig. 1, Supplementary Movie 1). These plains are near the Okaukuejo tourist camp where the Etosha Ecological Institute is located and where most surveillance trips on the park road system begin and end. We thus focus on this central region of ENP. Importantly, the habitat across this region is largely open, yielding similar situations for carcass detection. *B. anthracis* is not considered a threat to its hosts in ENP and is not currently managed. However, it remains unknown the extent to which the bacterium regulates its host population, alters competitive interactions or subsidies the scavenger population because the rate of anthrax-generated mortality remains unknown. Accurate estimates of anthrax-related mortality would facilitate better decision-making regarding
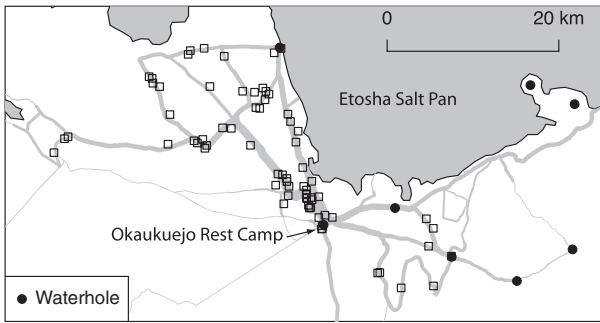
Fig. 1. Map of the central region of ENP showing plains zebra carcasses (squares) detected by passive surveillance in Feb–May 2010. Road (grey lines) width scales with the square root of the number of trips made on that road during the study period.

how (and whether) to manage the disease in the future. Unbiased mortality estimates can also facilitate our understanding of the causes of transmission patterns, as patterns in passive surveillance data reflects not only the transmission process but also the spatiotemporal distribution of surveillance effort.

NON-UNIFORM DISTRIBUTION OF PERPENDICULAR DISTANCES TO ROADS

Roads in ENP often connect waterholes and are built in areas with high game density to facilitate tourism and management (Fig. 1). Thus, concern that $\pi(y)$ is not uniform is warranted. We assume that the distribution of carcass perpendicular distances to the nearest road is an unbiased sample of live animal locations. While large carnivores occasionally move carcasses, the distance moved is rarely great enough to affect this assumption in our experience. Consequently, we can assume that where animals die is a random sample of their movement paths while alive and use movement data from 27 GPS-collared zebra to estimate $\pi(y)$. Specifically, we measured the distance of all GPS fixes to the nearest road and fit a truncated gamma distribution to all points within a maximum strip width of 800 m from the road (Fig. 2; strip width chosen as advised in Buckland *et al.* 2001).
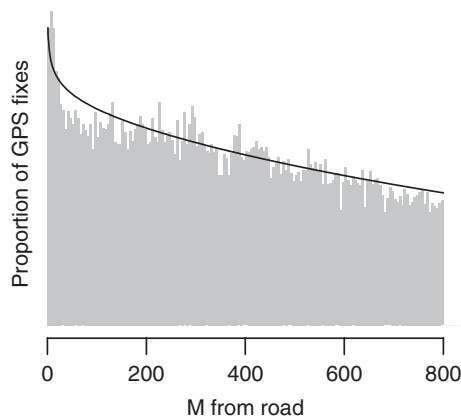

Fig. 2. Distribution of perpendicular distance from road from 52,745 GPS fixes collected from 27 collared plains zebra in the Okaukuejo region of Etosha national park during the late wet season (Feb–May). The black line shows the fitted truncated gamma distribution used as $\pi(y)$ to fit the detectability functions. Data are only showed up to the maximum strip width of 800 m for which the distance sampling analysis is conducted.

TEMPORALLY VARIABLE CUES

Carcasses are observed by detecting cues that we group into three types: (1) avian scavengers (e.g. vultures, marabou storks, crows, raptors), (2) mammalian scavengers (e.g. jackals, hyenas, lions), or (3) the carcass itself. Since detection is intimately linked to these cues, we modelled detectability as a function of cue presence, which itself is modelled as a function of time since death, as estimated from data collected by motion-sensor camera traps deployed at 31 fresh zebra carcasses. Presence of each cue type was abstracted from the photographs for each 15 min interval up until t_{max} = 5 days after death by which time the most detectable cue types are no longer available. In the following analysis, we only included carcasses detected by avian scavengers, mammalian scavengers or a fresh carcass (defined as a carcass with the majority of muscle and internal organs intact) and excluded the few detected carcasses thought to be older than 5 days. Such carcasses are rarely detected far from the road and number too few to robustly estimate detectability functions for these cues. The temporal distribution of cues over time is displayed in Fig. 3. We modelled detection conditional on the dominant cue, where cue dominance was determined by the available cue with the greatest visibility. Thus, mammalian scavengers were the dominant cue when they were present, but avian scavengers were absent, and a fresh carcass was the dominant cue only when neither avian nor mammalian scavengers were present.

OPPORTUNISTIC SAMPLING PLATFORM

In ENP passive carcass surveillance, we consider the 'survey' to be comprised of opportunistic observations of carcasses by researchers while conducting other field work. Surveillance effort is thus highly variable across time and space, depending on the number of individuals working in the park at any given time. To this end, we divided the Okaukuejo road system (Fig. 1) into road segments of length $\leq$ 5 km and asked researchers to record the roads driven for each trip. Consequently, trips are the unit of surveillance effort. Carcasses are only reported once and so that the detection on any given trip is always conditioned on non-detection on all previous trips passing that carcass. Without effort data, we cannot distinguish absence of carcasses observed from absence of effort. Therefore, only carcasses recorded by individuals reporting surveillance effort during road-based passive
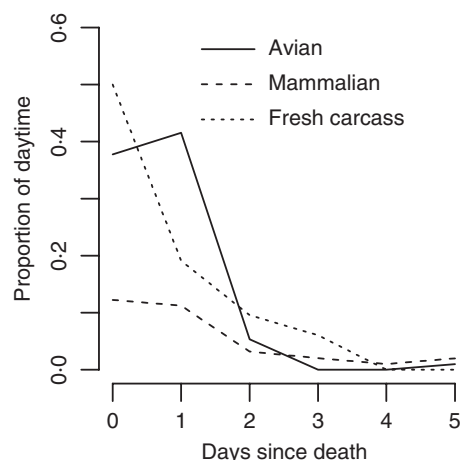

Fig. 3. Proportion of time a sighting cue is the dominant cue at a carcass as a function of day since death as estimated from camera traps placed at fresh zebra carcasses.

surveillance were analysed. The vast majority of carcasses, however, were detected by passive road-based surveillance with reported effort. Ignoring other carcasses will conservatively bias mortality estimates downwards because carcasses reported outside of reported effort may otherwise have been detected during surveillance effort at a later trip.

SPACE-TIME WINDOWS AND EXTRAPOLATION

In CDS, abundance can be obtained from density for a closed study area (such as a demarcated habitat) of size A. When estimating the cumulative incidence of events, such as deaths, we are interested in restricting estimation to a given space-time volume. We choose to estimate outbreak size during the last four months of the 2010 wet season in ENP (Feb–May; Supplementary Movie 1) as well as restrict our attention to this period during simulation. Extrapolation from space-time volume inside the surveillance effort (i.e. strips around road-days with effort in the next 5 days) to space-time volume outside the surveillance effort (i.e. areas far from roads, or near roads, but at times when there is no effort) must be made with caution. The validity of this extrapolation relies on the similarity in host utilisation and transmission intensity between the surveilled area and the greater space-time window. In this analysis, we estimate cumulative incidence in the space-time volume defined by cylinders with half-strip widths of 800 m around the road system for days when roads were driven in the next 5 days. Thus, we make the distinction that of the N_{all} carcasses in the study area, N have nonzero detection probabilities, and for now restrict our attention to estimating this quantity. Future methodological development for extrapolation outside this space-time window to a greater temporal (seasonal) and spatial (area encompassing the population's distribution) scale is mentioned in the discussion as potential future work.

SIMULATION

To assess the accuracy and precision of $\widehat{N}_C$, we simulated data based on the actual surveillance effort analysed below. Briefly, we distributed N carcasses within 800 m of the Etosha National Park roads on days when they were driven in the next 5 days as recorded in our surveillance system (i.e. the surveilled space-time volume). In this way, all N carcasses had a nonzero probability of detection. We conducted simulations with both uniform (Scenarios 1 and 2) and gamma distributions

for $\pi(y)$ (Scenarios 3 and 4). Parameters of the latter were simulated using the fit from GPS movement data (Fig. 2). For Scenarios 1 and 3, we distributed carcasses across roads and days using the discrete uniform distribution $\pi_{\text{unif}}(r, d) = \frac{\text{length}(r)}{(\#\text{days})\Sigma\,\text{length}(r_i)}$. To simulate a more realistic spatiotemporally heterogeneous distribution of carcasses, for Scenarios 2 and 4 we used $\pi_{\text{ST}}(r, d) = \frac{\pi'_{ST}(r,d)}{\Sigma\pi'_{ST}(r,d)}$ where $\pi'_{\text{ST}}(r, d) = \pi_{\text{unif}}(r, d) \times S$ and S is a random variable defined by $S \sim \Gamma(1, 0.5)$. For simulations with gamma $\pi(y)$, we also estimated cumulative mortality assuming $\pi(y)$ is uniform to assess how this assumption might have biased analysis of real data. For each of the six combinations of possibilities, we simulated 100 carcass populations, filtered them through the following detection process and then estimated N using the estimator $\widehat{N}_C$.

Each carcass could be detected on trips to the road where it occurred on the day of death and the five following days. For each trip, the available sighting cue was randomly chosen using $\pi(c|t)$ fitted from the camera trap data. The probability of detection given that cue was then calculated using the detection functions and detection function parameter values given in Table 1 and then a Bernoulli trial determined whether the carcass was detected on that trip or not. Bernoulli trials were performed for trips until the carcass was detected or the last trip within the 5-day detection window was evaluated and the carcass was determined to have been undetected on all trips.

Results

SIMULATION RESULTS

Estimation of the detection function parameters through maximum likelihood maximisation performed well and consequently the Horvitz-Thompson-like estimator performed well for all four scenarios (Table 1). Parametric bootstrapping confidence intervals enclosed the true number of carcasses $\geq 94\%$ of the time, but yielded rather high upper boundaries in comparison to the nonparametric bootstrap confidence intervals which had lower coverage, most likely due to the inappropriateness of using carcasses as the bootstrap sampling unit. Thus, we propose that the parametric bootstrap confidence intervals should be used. The estimator and parametric bootstrap

Table 1. Mean of 100 detection function parameters, carcass abundance estimates ($\widehat{N}_C$), confidence intervals and their coverages are given for each of the four simulation scenarios. Standard errors are given in parentheses

True value		Scenario 1	Scenario 2	Scenario 3	Scenario 4
n	–	65 (1)	66 (1)	77 (1)	75 (1)
$\widehat{N}_C$	300	302 (9)	296 (8)	313 (8)	308 (9)
Bias	–	2.2	−3.6	13	7.6
Mean square errror	–	7494	6628	7542	7676
CI 95[a]	–	204 (4)–724 (53)	203 (4)–680 (41)	221 (4)–804 (93)	219 (4)–671 (47)
CI 95[b]	–	212 (4)–497 (30)	206 (4)–482 (26)	227 (3)–500 (35)	220 (4)–474 (24)
CI coverage[a]	–	0.97	0.94	0.97	0.94
CI coverage[b]	–	0.84	0.81	0.85	0.89
σ_{avian}	0.40	0.4 (0.0043)	0.39 (0.004)	0.39 (0.0045)	0.39 (0.0043)
σ_{mamm}	0.12	0.14 (0.0035)	0.14 (0.0032)	0.14 (0.0031)	0.14 (0.0038)
σ_{carcass}	0.10	0.098 (0.0016)	0.098 (0.0019)	0.099 (0.0016)	0.097 (0.0017)

[a]Confidence intervals constructed using the parametric bootstrap with the information matrix estimate of the covariance matrix and the delta method.
[b]Confidence intervals constructed using the nonparametric bootstrap approach.
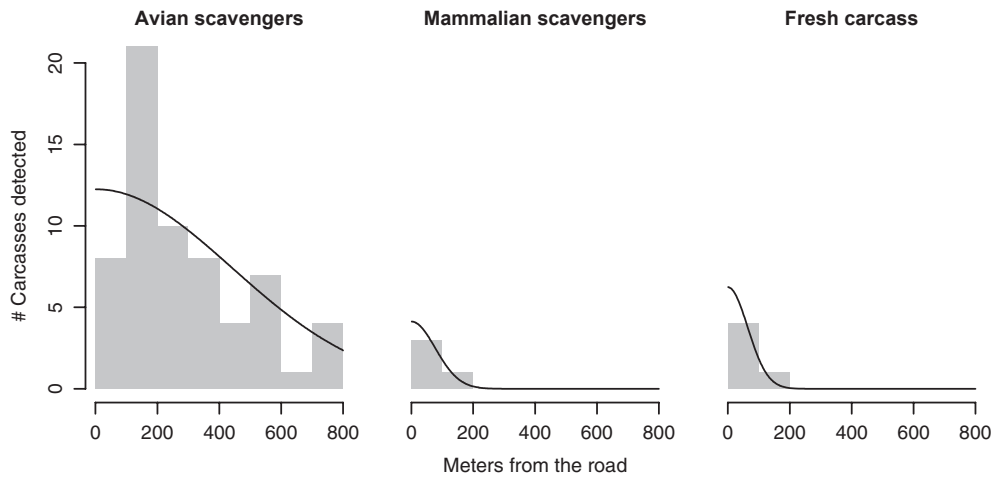
Fig. 4. Distribution of perpendicular distances between sighted carcasses and roads for zebra carcasses detected during passive surveillance in Feb–May 2010 by sighting cue type. Maximum likelihood fitted detectability functions, as estimated with the estimated distribution of $\pi(y)$ modelled as a truncated gamma distribution from GPS movement data from live zebra (Fig. 2), are displayed as a black line, with lines normalised so that the area under the curve matches the area of the histogram bars.

confidence intervals proved relatively robust to spatiotemporal heterogeneity in carcass incidence density (Scenarios 2 and 4).

ANTHRAX SURVEILLANCE ANALYSIS

During Feb–May 2010, individuals recording surveillance effort detected 72 zebra carcasses within the 800 m half-strip width of Okaukuejo area roads in ENP. The vast majority of these carcasses were detected by avian scavengers (Fig. 4). Of these carcasses, 50 (69%) were confirmed anthrax positive by selective culture of *B. anthracis* or molecular diagnostics. Using the newly constructed estimator, we estimated that within the surveilled space-time volume there were 272 (208–592) zebra carcasses in total where the parenthetical here and thereafter gives parametric bootstrap confidence intervals (Table 2). While parametric bootstrapping does not include the error associated with estimation of $\pi(y)$ or $\pi(c|t)$ because it is based on optimisation of a single data set, nonparametric bootstrap confidence intervals including error in estimation of these distributions suggests that this error was minor compared to error associated with the estimation of $\widehat{N}_C$. Assuming the prevalence of anthrax amongst observed and unobserved carcasses is equal, we estimated that 189 (145–411) anthrax-related zebra mortalities occurred in the surveilled space-time window – 3.8 (2.9–8.2) times greater than the observed number. Given that this quantity only estimates mortality within the surveilled space-time volume, it (and the associated confidence interval) already serves as a valuable lower bound for the incidence of anthrax during this outbreak. The most recent aerial survey estimate of zebra population size in ENP was 12,982 in 2005 (95% confidence interval: 10,937–15,027) (W. Kilian unpublished data 2011).

Discussion

While CDS provides a solid framework for developing surveys of reasonably abundant and visible animals, opportunistic

Table 2. Estimates of cumulative mortality in the plains zebra of Etosha National Park in the surveilled region during the 2010 anthrax outbreak

	Uniform $\pi(y)$	Gamma $\pi(y)$
n	72	72
$\widehat{N}_C$	366	272
CI[a]	(252, 856)	(208, 592)
CI[b]	(256, 601)	(202, 381)
CI[b]	–	(200, 393)
CI[d]	(242, 590)	(188, 404)
CI[e]	–	(194, 412)
σ_{avian}	0.517	0.624
σ_{mamm}	0.108	0.109
$\sigma_{carcass}$	0.090	0.090

[a]95% Parametric bootstrap confidence intervals.
[b]95% Nonparametric bootstrap confidence intervals.
[c]95% Nonparametric bootstrap confidence intervals with resampling over both observed carcasses (as in footnote) and also the GPS-collared zebras used to fit a gamma distribution for $\pi(y)$.
[d]95% Nonparametric bootstrap confidence intervals with resampling over both observed carcasses (as in footnote) and camera traps used to estimate $\pi(c|t)$.
[e]95% Confidence intervals constructed using the nonparametric bootstrap approach with resampling over observed carcasses, $\pi(y)$ and $\pi(c|t)$ (as in footnotes).

data may be preferable for estimating the abundance of elusive animals or short-lived carcasses. Although opportunistic data are readily available, they are often underused or misused due to biases inherent in the lack of transect randomisation (Hedley & Buckland 2004; Kiszka *et al.* 2007) or a poor understanding of how carcass removal and detection are competing processes (Smallwood *et al.* 2010). Mark–recapture distance sampling (MCDS) may be appropriate for treating distance data from multiple observers in an actively designed survey (Buckland, Laake & Borchers 2010). However, in opportunistic surveillance when each observer may have multiple chances to observe a carcass and communication between observers

ensures each carcass is only recorded once, even the weakest independence assumptions of MCDS are violated. In contrast, temporally explicit modelling of sighting cue distributions allows carcass removal to be treated as a dynamic process operating on the same time-scale as the survey. With a temporal model, the integration of auxiliary data on surveillance effort, sighting cue variation over time and animal movement, we were able to create robust point and interval estimators of cumulative mortality in a distance sampling framework. The general likelihood approach provided here could be used to estimate cumulative mortality in a wide variety of applications, including opportunistic surveillance of mortality due to disease, wind farms, pesticides and road kills (although distance sampling from the road may not be applicable for the latter, our approach to modelling removal processes remains applicable). We feel that where long-term opportunistic data sets already exist, acquiring such auxiliary data (if not already available) will still often be cheaper than active CDS surveys.

To estimate abundance in this framework, we made several assumptions. First, only carcasses detected by individuals recording surveillance effort were included in the analysis. Because carcasses are only detected once (after the first sighting of a carcass, communication amongst vehicles allows all teams to know its location), carcasses detected outside surveillance effort were therefore 'censored' from the data set, biasing cumulative mortality estimates downwards. Second, we assumed that detection functions are not variable across the study area based on the relative homogeneity and openness of the Okaukuejo plains, on which the zebra spend the majority of their wet season. Finally, we estimated $\pi(y)$ using only a limited number of GPS-collared animals. We accounted for the sample size directly by including the error in $\pi(y)$ estimation directly via bootstrapping. However, the choice of functional form for $\pi(y)$ was *ad hoc*. The empirical distribution from the GPS data could have been used itself, although this may be more sensitive to individual animal heterogeneity. We also assumed that the distribution was spatiotemporally homogenous due to the lack of sufficient data to understand whether the way animals act around the road varies in space or time. If animals indeed die closer to the roads than expected from the GPS movement data, then our estimator would be upwards-biased. The goal of this article is to present methods for using auxiliary data in distance sampling analyses of opportunistic data. We caution readers that when applying these methods, they should carefully examine the assumptions regarding the relationship between the available data sets and the true distributions $\pi(y)$ and $\pi(c|t)$.

The most obvious extension of these methods in future work is to allow for extrapolation to the entire space-time volume of interest (i.e. the study area over an entire season or year). This could be done by modelling overlap between the surveilled space-time volume and the live host animal spatiotemporal distribution, with the latter estimated using movement or other live population survey data. The spatiotemporally explicit formulation of the estimator proposed in this manuscript was biased because the vast majority of surveillance effort did not result in carcass detection, but nonetheless affected the estima-

tion detection function parameters when maximising Eqn 2. Nevertheless, we suggest that future work should incorporate the conditional formulation's (Eqn 4) results into a generalised additive modelling framework to make spatiotemporally explicit estimates of mortality incidence density (Hedley & Buckland 2004; de Segura *et al.* 2007).

References

Borchers, D.L., Zucchini, W. & Fewster, R.M. (1998) Mark-Recapture Models for Line Transect Surveys. *Biometrics*, **54**, 1207–1220.

Buckland, S.T., Goudie, I.B.J. & Borchers, D.L. (2000) Wildlife population assessment: past developments and future directions. *Biometrics*, **56**, 1–12.

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

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

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. & Thomas, L. (2004) *Advanced Distance Sampling*. Oxford University Press, Oxford.

Hedley, S.L. & Buckland, S.T. (2004) Spatial models for line transect sampling. *Journal of Agricultural Biological and Environmental Statistics*, **9**, 181–199.

Himes Boor, G.K. & Small, R.J. (2012) Steller sea lion spatial-use patterns derived from a Bayesian model of opportunistic observations. *Marine Mammal Science*, **28**, E375–E403.

Hugh-Jones, M.E. & de Vos, V. (2002) Anthrax and wildlife. *Revue Scientifique Et Technique De L Office International Des Epizooties*, **21**, 359–383.

Hunter, J.S., Durant, S.M. & Caro, T.M. (2007) Patterns of scavenger arrival at cheetah kills in Serengeti national park Tanzania. *African Journal of Ecology*, **45**, 275–281.

Johnson, D.S., Laake, J.L. & Ver Hoef, J.M. (2010) A model-based approach for making ecological inference from distance sampling data. *Biometrics*, **66**, 310–318.

Kiszka, J., Macleod, K., Van Canneyt, O., Walker, D. & Ridoux, V. (2007) Distribution, encounter rates, and habitat characteristics of toothed cetaceans in the Bay of Biscay and adjacent waters from platform-of-opportunity data. *ICES Journal of Marine Science: Journal du Conseil*, **64**, 1033–1043.

Lindeque, P.M. & Turnbull, P.C.B. (1994) Ecology and epidemiology of anthrax in the Etosha national park, Namibia Onderstepoort. *Journal of Veterinary Research*, **61**, 71–83.

Marques, T.A., Buckland, S.T., Borchers, D.L., Tosh, D. & McDonald, R.A. (2010) Point transect sampling along linear features. *Biometrics*, **66**, 1247–1255.

Plumptre, A.J. (2000) Monitoring mammal populations with line transect techniques in African forests. *Journal of Applied Ecology*, **37**, 356–368.

Ponce, C., Alonso, J.C., Argandona, G., Garcia Fernandez, A. & Carrasco, M. (2010) Carcass removal by scavengers and search accuracy affect bird mortality estimates at power lines. *Animal Conservation*, **13**, 603–612.

Rivera-Milán, F.F., Zaccagnini, M.E. & Canavelli, S.B. (2009) Field trials of line-transect surveys of bird carcasses in agro-ecosystems of Argentina's Pampas region. *Wildlife Society Bulletin*, **32**, 1219–1228.

Royle, J.A. & Dorazio, R.M. (2008) *Hierarchical Modelling and Inference in Ecology: The Analysis of Data From Populations, Metapopulations and Communities*, 1st edn. Academic, Amsterdam, Boston.

Santos, S.M., Carvalho, F. & Mira, A. (2011) how long do the dead survive on the road? Carcass persistence probability and implications for road-kill monitoring surveys. *PLoS ONE*, **6**, e25383.

de Segura, A.G., Hammond, P.S., Canadas, A. & Raga, J.A. (2007) Comparing cetacean abundance estimates derived from spatial models and design-based line transect methods. *Marine Ecology-Progress Series*, **329**, 289–299.

Smallwood, K.S., Bell, D.A., Snyder, S.A. & Didonato, J.E. (2010) Novel scavenger removal trials increase wind turbine-caused avian fatality estimates. *Journal of Wildlife Management*, **74**, 1089–1097.

Walsh, P.D. & White, L.J.T. (1999) What it will take to monitor forest elephant populations. *Conservation Biology*, **13**, 1194–1202.

Williams, R., Hedley, S.L. & Hammond, P.S.. (2006) Modeling distribution and abundance of Antarctic baleen whales using ships of opportunity. *Ecology and Society*, **11**, http://www.ecologyandsociety.org/vol11/iss1/art1.

Supporting Information

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

Data S1. All the R scripts and data files necessary to reproduce both the simulation and Etosha data analyses have been enclosed as Supplementary Data S1. The authors have also made all the code accessible on a public Amazon Machine Image (AMI) with the ID: ami-d82187b1. Interested users are encouraged to run their analysis directly through this image via Amazon EC2 Cloud Computing. A more thorough description of the scripts and data sets are provided in the Online Supporting Information.

Movie S1. The top panel shows daily surveillance effort by showing the roads driven each day as purple lines, where line thickness scales with the square root of the number of trips driven on that road. Squares represent zebra (*Equus quagga*) carcasses with red, blue and black representing confirmed positive, negative and unknown for *Bacillus anthracis* (the causal bacterial agent of anthrax). Dots display the GPS movement fixes (at 20 min intervals) of GPS-collared live zebra. Each colour represents a different animal. The bottom panel shows the cumulative driving effort and carcasses.