



Variance Propagation for Density Surface Models

Mark V. BRAVINGTON, David L. MILLER , and Sharon L. HEDLEY

Spatially explicit estimates of population density, together with appropriate estimates of uncertainty, are required in many management contexts. Density surface models (DSMs) are a two-stage approach for estimating spatially varying density from distance sampling data. First, detection probabilities—perhaps depending on covariates—are estimated based on details of individual encounters; next, local densities are estimated using a GAM, by fitting local encounter rates to location and/or spatially varying covariates while allowing for the estimated detectabilities. One criticism of DSMs has been that uncertainty from the two stages is not usually propagated correctly into the final variance estimates. We show how to reformulate a DSM so that the uncertainty in detection probability from the distance sampling stage (regardless of its complexity) is captured as an extra random effect in the GAM stage. In effect, we refit an approximation to the detection function model at the same time as fitting the spatial model. This allows straightforward computation of the overall variance via exactly the same software already needed to fit the GAM. A further extension allows for spatial variation in group size, which can be an important covariate for detectability as well as directly affecting abundance. We illustrate these models using point transect survey data of Island Scrub-Jays on Santa Cruz Island, CA, and harbour porpoise from the SCANS-II line transect survey of European waters.

Supplementary materials accompanying this paper appear on-line.

Key Words: Abundance estimation; Distance sampling; Generalized additive models; Line transect sampling; Point transect sampling; Spatial modelling.

1. INTRODUCTION

Distance sampling is a widely used method for estimating abundance when detection is imperfect (Buckland et al. 2001), based on encounters along line or point transects. Detection probability (detectability) is estimated using within-encounter data (e.g. perpendicular

M. V. Bravington and D. L. Miller: Joint first author.

M. V. Bravington, Commonwealth Scientific and Industrial Research Organisation Marine Laboratory, Hobart, Australia. D. L. Miller (✉), Centre for Research into Ecological and Environmental Modelling and School of Mathematics and Statistics, University of St Andrews, St Andrews, Fife, Scotland. S. L. Hedley, St Andrews, Fife, Scotland
(E-mail: dave@ninepointeightone.net).

© 2021 The Author(s)

Journal of Agricultural, Biological, and Environmental Statistics
<https://doi.org/10.1007/s13253-021-00438-2>

distance from trackline), by fitting “detection functions” that may involve environmental covariates (e.g. local weather conditions). In traditional stratified distance sampling, an average animal density is then estimated within each survey stratum—i.e. some region within which survey coverage is supposed to be uniform—based on the observed encounter rate within that stratum divided by the detectability, and then scaled by the stratum area. Since the abundance estimate is a simple function of statistically independent quantities (encounter rate and detectability), its variance can be estimated straightforwardly.

Instead of using strata, with modern statistical tools it is possible to fit spatially explicit models of density, where local density is assumed to vary gradually in space (and perhaps also in response to specific environmental covariates, which we here include under the general heading of “spatially explicit”). Spatially explicit estimates are advantageous in many situations: when abundance estimates are required across arbitrary sub-regions that do not coincide with survey strata; to reduce bias when coverage is uneven; or when identifying particularly important habitat for conservation, for example.

There are various approaches to actually fitting spatially explicit models. The general idea, as in the stratified case, is that the expected local encounter rate is the product of local detectability and local density, but with both factors now potentially depending on local spatial and/or environmental covariates. Here, we consider specifically density surface models (DSMs; Hedley and Buckland 2004; Miller et al. 2013), which take a two-stage approach. The first stage is to estimate detectability using a detection function model; any standard or bespoke model could be used (see Sect. 2). In the second stage, the encounter rate data are fitted to location/environmental covariates using a GAM, specifically the “basis-and-penalty” formulation of GAMs in Wood (2017) in which smoothers are represented via random effects. The estimated detectabilities for each segment of search effort are easily accommodated in the GAM (technically, as offsets to the linear predictor; see below), and the range of smoothers and interactions that can be fitted in is very wide.

Splitting the analysis into two stages is appealing partly because existing domain-specific software and diagnostic expertise can be applied as-is to each stage separately, and partly because it avoids any need to write inevitably complicated code that incorporates two individually complex aspects. It is also straightforward to produce a point estimate of abundance for any desired sub-region straight from the fitted GAM. However, when detectability and density both vary spatially, the problem is what to do about variance given that GAMs do not intrinsically “understand” the notion of uncertainty in their offset.

In this paper, we show how statistical uncertainty about detectability can in fact be accommodated painlessly within standard GAM software. Our approach is to first fit the detection function as usual, but then to rewrite the fitted detection function log-likelihood as a quadratic approximation centred on its point estimates, and to incorporate the uncertainty about the detection function parameters via random effects in the second-stage GAM. This fits directly and automatically into the Wood/Wahba formulation of a GAM, whereby a smooth surface is described by a set of coefficients treated as random effects; thus, the machinery for handling random effects in general is already built into the `mgcv` software used in Miller et al. (2013)’s DSM code. This amounts to refitting the detection function model (or a good approximation to it) at the same time as fitting the GAM. We retain the benefits of two-stage modelling, but all the uncertainty about detectability as well as density

is now captured in the usual GAM outputs. The refitted detection function model should not differ greatly from the original fit, and we use this idea to propose a diagnostic for overall model specification issues.

Following a summary of DSMs and notation in Sect. 2, we present our new formulation in Sect. 3, including variance computation and diagnostics. Section 4 comments on problems with existing approaches to variance propagation in DSMs. In Sect. 5, we extend the formulation to cover DSMs where group size varies spatially and affects detectability (a common situation with whales and dolphins). Section 6 gives examples of the variance propagation method and the group size model. Some discussion is given in Sect. 7, including possible generalizations.

2. DENSITY SURFACE MODELS

In distance sampling, observers move along a set of survey lines or between points, counting (groups of) animals, recording distances from the centre line or centre point to the observed groups (or their cues, such as blows for cetaceans or calls for birds), the size of each detected group and potentially other covariates that may affect detectability.

To fully describe the DSMs in this paper, we distinguish four different classes of variable.

1. Density covariates, x , vary in space and potentially affect local animal abundance: e.g. latitude and depth. They are required for prediction and fitting, and are assumed known across the entire region of interest.
2. Effort covariate(s), z , affect detection probability: e.g. sea conditions measured on the Beaufort scale, or observer identity. They are assumed known along each transect, but not necessarily in unsurveyed areas.
3. Individual covariates, g , that affect detection probability and are a persistent property of each group (independent of whether the group is observed or not) during its window of observability: e.g. size (number of animals), and perhaps behaviour. Here, g is assumed known for each observed group (see Discussion). The random variable G varies from one group to the next, and its statistical distribution $F_G(g; x)$ may vary spatially. $F_G(g; x)$ may have a direct effect on abundance (via the mean group size), as well as on detection probability, in which case it is also necessary to estimate certain properties of $F_G(g; x)$ such as its local mean.
4. Observation variables, y , which are random properties of one observation on one group: e.g. perpendicular distance between the group and the sampler. In certain settings, y may contain other elements. For example, in a multi-observer platform survey (e.g. MRDS; Borchers et al. 1998), y might also include which of the active observers saw the group; in a cue-based setting, y might include the bearing between sighting and observer.

These classes are assumed to be mutually exclusive; overlaps can lead to fundamental problems for distance sampling which we do not address here (e.g. non-uniform animal distribution within the sample unit; Marques et al. 2012). The distinction between *individual*

and *effort* covariates is often glossed over but they have rather different implications for abundance estimation (see below).

In the first stage of DSM, the detection function $\pi(y|\theta, z, g)$, which involves unknown parameters θ as well as z and g , describes the probability of making an observation at y . The parameters θ are usually estimated by maximizing this log-likelihood across observations s :

$$l(\boldsymbol{\theta}) = \sum_s \log_e \left(\frac{\pi(y_s|\boldsymbol{\theta}, z_{t_s}, g_s)}{p(\boldsymbol{\theta}; z_{t_s}, g_s)} \right), \quad (1)$$

where t_s is the transect containing sighting s . Here, p is the overall detection probability for a group, defined by

$$p(\boldsymbol{\theta}; z, g) = \int \pi(y; \boldsymbol{\theta}, z, g) dF_Y(y),$$

where F_Y is the distribution function of y . In standard distance sampling where y consists only of perpendicular distance, F_Y is uniform between 0 and some fixed truncation distance, beyond which observations are discarded. This formulation encompasses a wide range of models, including multiple covariate distance sampling (MCDS; Marques and Buckland 2003) with z and g , multi-observer mark–recapture distance sampling (MRDS; Borchers et al. 1998), and cue-based “hazard probability models” (Skaug and Schweder 1999).

The second part of DSM models the local count of observations via a GAM to capture spatial variation in animal density. This allows us both to estimate abundance within any sub-region of interest, and to compensate as far as possible for uneven survey coverage (whether by design, or by virtue of field logistics and weather conditions). Since line transects are generally very long in comparison with their width and therefore contain a range of density and density covariate values, we divide transects into smaller *segments*, which are the sample units for GAM (in which case the subscript t_s above refers to segments rather than transects). Point transects are left as-is; we use the term “segments” from now on to refer to both points and line segments, without loss of generality. Environmental covariates are assumed not to change much within each segment. The relationship between counts n_i per segment i and density covariates x_{ik} is modelled as an additive combination of smooth functions with a log link:

$$\mathbb{E} \left[n_i | \boldsymbol{\beta}, \boldsymbol{\lambda}, p(\hat{\boldsymbol{\theta}}; \mathbf{z}_i) \right] = a_i p(\hat{\boldsymbol{\theta}}; \mathbf{z}_i) \exp \left(\beta_0 + \sum_K f_k(x_{ik}) \right), \quad (2)$$

where each segment is of area a_i , and n_i follows some count distribution such as quasi-Poisson, Tweedie, or negative binomial. The f_k are smooth functions, represented by a basis expansion ($f_k(x) = \sum_j \beta_j b_j(x)$, for some basis functions b_j); β_0 is an intercept term, included in parameter vector $\boldsymbol{\beta}$; $\boldsymbol{\lambda}$ is a vector of smoothing (hyper)parameters which control the wiggleness of the f_k . We take a Bayesian interpretation of GAMs, in which $\boldsymbol{\lambda}$ controls the variance of a multivariate improper Gaussian prior (Wood 2017):

$$\boldsymbol{\beta} \sim N \left(\mathbf{0}, \phi \left(\sum_k \lambda_k \mathbf{S}_k \right)^{-} \right),$$

with scale parameter ϕ , smoothing parameters λ_k and penalty matrices \mathbf{S}_k ($^{-}$ indicates pseudoinverse). This leads to a quadratic penalty on beta during fitting. We estimate $\boldsymbol{\lambda}$ itself via REML (Wood 2011), an empirical Bayes procedure. Fully Bayesian approaches, placing hyperpriors on $\boldsymbol{\lambda}$ are also possible.

We are interested in the uncertainty of a predicted abundance estimate, \hat{N} . We assume below that we have created some prediction grid with all density covariates available for each cell in the grid. Abundance is predicted for each cell, and summed for an overall abundance, \hat{N} , over some region of interest which may not be the entire surveyed area. Although $p(\hat{\boldsymbol{\theta}})$ does not appear explicitly in the prediction, which is

$$\hat{N} = \sum_j a_j \exp \left(\hat{\beta}_0 + \sum_k \hat{f}_k(x_{jk}) \right),$$

the GAM offsets $p(\hat{\boldsymbol{\theta}})$ clearly do affect $\hat{\beta}$, so it is important to account somehow for detection probability uncertainty. (2) assumes the offset is fixed, so extra steps are required.

3. VARIANCE PROPAGATION FOR DENSITY SURFACE MODELS

Let $p(\boldsymbol{\theta}_0, z_i)$ be the true probability of detection in segment i and for now omit g , thereby assuming that there are no individual-level covariates (e.g. that group size is always 1) for now (see Sect. 5). If $\boldsymbol{\theta}_0$ is the true (unknown) value of $\boldsymbol{\theta}$, and $\hat{\boldsymbol{\theta}}$ is its MLE, we use the shorthand $p_i = p(\boldsymbol{\theta}_0, z_i)$ and $\hat{p}_i = p(\hat{\boldsymbol{\theta}}, z_i)$ when the dependence is clear. The expected number of encounters in segment i is $a_i p_i \rho_i$ where ρ_i is the underlying density, given by the exponential term in 2.

Given p_i , we can rewrite (2) on the log link scale as:

$$\log \mathbb{E}[n_i | \boldsymbol{\beta}, \boldsymbol{\lambda}, p_i] = \eta_i = \log a_i p_i + X_i \boldsymbol{\beta}. \quad (3)$$

X_i is the (known) i^{th} row of the design matrix, i.e. the values of the basis functions in segment i , so $\log \rho_i = \sum_k f_k(x_{ik}) = X_i \boldsymbol{\beta}$ and $\log a_i p_i$ is an offset. The complication is that we only have an estimate of p_i . To tackle this, we first rewrite the linear predictor η_i as

$$\eta_i = \log a_i + \log \hat{p}_i + \log p_i - \log \hat{p}_i + X_i \boldsymbol{\beta}$$

and then take a Taylor series expansion of $\log \hat{p}_i \equiv \log p(\hat{\boldsymbol{\theta}}, z_i)$ about $\boldsymbol{\theta} = \boldsymbol{\theta}_0$:

$$\log p(\hat{\boldsymbol{\theta}}, z_i) = \log p(\boldsymbol{\theta}_0, z_i) + (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^{\top} \cdot$$

$$\left[\frac{d \log p(\boldsymbol{\theta}, z_i)}{d\boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right] + O(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^2. \quad (4)$$

By defining the vectors $\boldsymbol{\delta} \triangleq \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ and $\kappa_i \triangleq \frac{d \log p(\boldsymbol{\theta}, z_i)}{d\boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$, we can rewrite (3) as

$$\log \mathbb{E}[n_i | \boldsymbol{\beta}, \boldsymbol{\lambda}, \hat{p}_i] = \log a_i \hat{p}_i + X_i \boldsymbol{\beta} + \kappa_i \boldsymbol{\delta} + O(\boldsymbol{\delta}^2). \quad (5)$$

In Supplementary Materials C, we show that the approximations in our approach do not affect the asymptotic order of accuracy. Specifically, the Laplace approximation that underlies REML estimation is accurate to $O(n^{-1})$ and our approximations are of the same order (see Supplementary Materials C for the meaning of n).

We have approximately that $\boldsymbol{\theta}_0 | \mathbf{y} \sim N(\hat{\boldsymbol{\theta}}, \mathbf{V}_\theta) \implies \boldsymbol{\delta} \sim N(\mathbf{0}, \mathbf{V}_\theta)$, where the covariance matrix \mathbf{V}_θ is calculated as the negative inverse Hessian of (1). In other words, the ‘‘posterior distribution’’ of $\boldsymbol{\theta}$ from fitting the detection function now becomes a prior distribution for ρ . To first order, $\boldsymbol{\delta}$ then plays the same structural role in (5) as the basis coefficients $\boldsymbol{\beta}$. The design matrix for $\boldsymbol{\delta}$ (κ in (5)) is obtained by differentiating the log-detection probabilities, with respect to $\boldsymbol{\theta}$ at $\hat{\boldsymbol{\theta}}$. Simple 3-point numerical differentiation is perfectly adequate for calculation of the derivatives. \mathbf{V}_θ should be readily available from detection function fitting (via the Hessian) regardless of the complexity of the model.

This method can be applied automatically to almost any distance sampling set-up provided one can calculate detection probabilities, find their derivatives, and obtain a Hessian for the likelihood. Simultaneous estimates $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ can be obtained from standard GAM fitting software. Posterior inferences about $\boldsymbol{\beta}$ (therefore ρ and abundance) automatically propagate the uncertainty from fitting the detection function.

The only technical difference from fitting a standard GAM is that $\boldsymbol{\lambda}$ is usually unknown and has to be estimated (i.e. the prior on $\boldsymbol{\beta}$ has known covariance, but unknown scale), whereas the prior on $\boldsymbol{\delta}$ is completely determined from the detection function fitting (i.e. in effect $\lambda_\delta = 1/\phi$, where ϕ is the scale parameter). This set-up cannot be specified directly in the R package `mgcv` because of implementation details (at least up to version 1.8; it may be possible within other GAM implementations), unless ϕ is fixed rather than estimated. This is fine for Poisson or negative binomial response, but in our experience, better fits can often be obtained using a Tweedie response distribution, for which ϕ must be estimated. In order to implement (5) for a general response distribution using `mgcv`, we therefore use a one-dimensional search over ϕ to maximize the marginal REML. At each iteration, given the working value ϕ^* , we refit the GAM fixing $\phi = \phi^*$ and $\lambda_\delta = 1/\phi^*$. Speed can be improved by reusing some of the set-up computations (design matrices, etc.) at each iteration.

Diagnostics. If the detection function fits properly and the spatial model has adequate flexibility, then the second-stage model should not lead to much change in the detection function parameters, so that $\hat{\boldsymbol{\delta}}$ should be ‘‘close’’ to 0. Nevertheless, there is scope for interaction if the detection function includes covariates that also vary systematically over space. For example, if weather is systematically worse in some parts of the survey region, then both $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ will contribute to the expected pattern of sightings, and the two sets of parameters

are partially confounded. (That is of course also true for all-in-one models, as well for our two-stage model.)

There are several diagnostics that we have found useful for checking consistency between the two parts of the model. The first is to compare the inferred spatial distribution and abundance from fitting (3) with the “naïve” estimates where detection uncertainty is ignored and the offset $a_i \hat{p}_i$ is treated as exact, ensuring that there are not large differences in the estimated spatial distribution. The second is to check whether the detection probabilities (by covariate level) would be substantially changed by fitting the spatial model; in other words, whether $\hat{\delta}$ is close enough to zero given its prior distribution, or, perhaps more usefully, whether the overall detectability by covariate level has changed. Since the fitted spatial model still includes the information from the first stage, any shift of more than about 1 standard deviations (based on the covariance from the detection function stage) might merit investigation. Third, as a general diagnostic tool for density surface models, we have found it useful to compare total observed and expected numbers of sightings, grouped by detection covariates (e.g. Beaufort). This can be helpful in diagnosing detection function problems, e.g. failure of assumed certain detectability at zero distance under poor weather conditions, as well as failures of the spatial model (e.g. an abrupt change in density). In addition, one could also use standard detection function model checking (e.g. quantile–quantile plots) with the adjusted parameters, $\hat{\theta} + \hat{\delta}$.

Calculating $\text{Var}(\hat{N})$. Once detection function uncertainty has been propagated, we only need to deal with uncertainty in the GAM, which now has an updated covariance matrix. We therefore can rely on two commonly used methods to obtain the variance of model outputs like abundance, \hat{N} .

1. *Delta method* We can calculate:

$$\text{Var}(\hat{N}) = \left(\mathbf{a}_p \frac{\partial \exp \mathbf{X}_p \boldsymbol{\beta}}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right) \mathbf{V}_{\hat{\boldsymbol{\beta}}} \left(\mathbf{a}_p \frac{\partial \exp \mathbf{X}_p \boldsymbol{\beta}}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right)^\top, \quad (6)$$

(the delta method) where $\mathbf{V}_{\hat{\boldsymbol{\beta}}}$ is the covariance matrix for the GAM coefficients Wood (2017, Sects. 5.8 and 6.9.3). We form the prediction matrix, \mathbf{X}_p , which maps model coefficients to values of the linear predictor for the prediction data, so $\hat{\boldsymbol{\eta}}_p = \mathbf{X}_p \hat{\boldsymbol{\beta}}$ (Wood 2017, Sect. 6.10). Derivatives are evaluated at the estimated values of the model parameters.

2. *Posterior simulation* The posterior for $\boldsymbol{\beta}$ given data \mathbf{y} and smoothing parameters $\boldsymbol{\lambda}$ are approximately distributed as $\boldsymbol{\beta} | \mathbf{y}, \boldsymbol{\lambda} \sim N(\hat{\boldsymbol{\beta}}, \mathbf{V}_{\hat{\boldsymbol{\beta}}})$. The following algorithm then can be used:

1. For $b = 1, \dots, B$:
 - (a) Simulate from $N(\hat{\boldsymbol{\beta}}, \mathbf{V}_{\hat{\boldsymbol{\beta}}})$, to obtain $\hat{\boldsymbol{\beta}}_b$.
 - (b) Calculate predicted abundance, $\hat{N}_b = \mathbf{a}_p \exp(\mathbf{X}_p \hat{\boldsymbol{\beta}}_b)$ (where \mathbf{a}_p is a row vector of areas for the prediction cells).
2. Calculate the empirical variance or percentiles of the \hat{N}_{bs} .

In practice, B in the order of 1000s appears to work well, though there may be some issues when the approximation breaks down. In these cases, we recommend the use of

importance sampling (either using importance weights to calculate weighted summaries or using a second resampling of the \hat{N}_b s) or a Metropolis–Hastings sampler (as implemented in `mgcv: : gam.mh`). Further examples are given in Supplementary Materials.

Software. The procedure given in this section is implemented in the R package `dsm`, available on CRAN. The `dsm_varprop` function in the package allows the user to provide a fitted DSM and a prediction grid. Using the delta method, it will then calculate an uncertainty estimate for the estimated abundance for that prediction grid. The function also returns the refitted GAM so one can extract the full covariance matrix and perform posterior simulation if required. Diagnostics for $\hat{\delta}$ are calculated by a `summary` method for the returned object.

4. PREVIOUS METHODS FOR ESTIMATING UNCERTAINTY IN DENSITY SURFACE MODELS

Several approaches have previously been suggested to combine detection function and spatial model predicted abundance uncertainties; we review them briefly here. We need to estimate the following:

$$\begin{aligned} \text{Var}_P(\log N) &= \mathbb{E}_P[\text{Var}(\log N|P)] + \text{Var}_P[\mathbb{E}(\log N|P)] \\ &\approx \text{Var}(\log N|\{\hat{p}_i; i = 1, \dots, n\}) \\ &\quad + \text{Var}_P[\log \hat{N}(\{\hat{p}_i; i = 1, \dots, n\})], \end{aligned}$$

where P here is a random variable for the (uncertain) probability of detection and the subscripts indicate the expectation/variance taken over that variable. $\hat{N}(\{\hat{p}_i; i = 1, \dots, n\})$ is the estimated abundance as a function of estimated detection probabilities. The first part of this can be derived from GAM theory as shown in the previous section; the second is more tricky.

Assuming independence. When \hat{p}_i is the same for all observations, then $N(\hat{p}) \propto 1/\hat{p}$, so \hat{N} and \hat{p} are independent. The total variance of the abundance estimate can be calculated by combining the GAM variance estimate with the variance of the probability of detection summing the squared coefficients of variation ($\text{CV}(X) = \sqrt{\text{Var}(X)}/\bar{X}$) (Goodman 1960). Hence,

$$\text{Var}_{\text{IND}}(\hat{N}) = \frac{\hat{N}^2}{\text{CV}^2(\hat{N}_{\text{GAM}}) + \text{CV}^2(\hat{p})}.$$

When there are not covariates in the detection function, we calculate:

$$\text{Var}(\hat{p}) = \left(\frac{\partial \hat{p}}{\partial \theta} \Big|_{\theta=\hat{\theta}} \right) \mathbf{V}_{\hat{\theta}} \left(\frac{\partial \hat{p}}{\partial \theta} \Big|_{\theta=\hat{\theta}} \right)^{\text{T}}. \quad (7)$$

This is fine when the detection function does not contain any covariates, as there is no covariance then between the effort and density covariates. (The procedure outlined in Sect. 3

does not yield a different answer.) In the case where detectability is a function of covariates, it is impossible in general to justify the use of the CV decomposition as there are correlations between the spatial distribution and the covariates affect detectability.

The approach taken by Program Distance (Thomas et al. 2010) is to use Horvitz–Thompson-adjusted counts per segment, instead of the observed count, as the response in the GAM. This removes the detectability from the right hand side of (2). Variance is then calculated by taking the probability of detection averaged over the observations by first calculating the Horvitz–Thompson estimate of the abundance in the covered area ($\hat{N} = \sum_i g_i / \hat{p}_i$, where g_i is group size of the i^{th} observation and \hat{p}_i is the probability of detecting that group) then using that \hat{N} to calculate the implied average detectability, had the analysis not contained covariates ($\tilde{p} = \tilde{n} / \hat{N}$, where \tilde{n} is the number of observed groups). The numerical derivatives of \tilde{p} with respect to θ can then be used in (7) to derive a variance for this probability of detection, averaged over the observations.

We do not recommend this approach either. Transforming the response through multiplication by a random variable breaks the mean–variance and independence assumptions of the GAM, so that the computed CV (\hat{N}_{GAM}) is invalid when detection covariates are present. Additionally, there is no coherent way to generalize the formula to small-area predictions—the effort covariates within a small area will not have the same range as those in the larger survey area (e.g. weather conditions will not be homogenous throughout the survey area). Hence, the uncertainty that applies to the overall \tilde{p} is usually not the appropriate uncertainty to apply to a small area where observing conditions may be atypical.

The bootstrap. Bootstraps are sometimes seen as an attractive alternative to deal with all aspects of variance in DSMs. Hedley and Buckland (2004) describe two possible implementations (one parametric, one nonparametric), which are not easy to choose between and which do not necessarily give similar answers. Ignoring computational time issues, the first practical difficulty in setting up a “good” nonparametric bootstrap for a DSM is sampling units, independent of the fitted model.

The second, more substantial, issue is the fundamental statistical problem with combining smoothers with bootstraps. The problem does not seem to be well known in the statistical ecology literature, so we give an explanation here. The basic problem is that (most) bootstraps use only the posterior modes of random effects (smoothers), thus omitting a key part of the posterior uncertainty. To see this, consider a simple “spatial model” where the region is divided into blocks, each with its own independent random effect, and a bootstrap that generates new data at each original observation/transect, either parametrically or nonparametrically. If one of the blocks is unsampled in the original data, it will be unsampled in every realization too, and the “spatial model” simply sets the point estimate of that random effect to zero in *every* bootstrap realization; hence, a bootstrap will ascribe zero uncertainty for the density in that block. The correct inference would of course be for the random effect to retain its prior variance.

This phenomenon has been well known in statistics since at least Laird and Louis (1987) (see also the discussants), who coined the term “naïve bootstrap” for such procedures that ignore the point estimate shrinkage inevitable in mixed or random effect models. (Fixed effect models are not susceptible in the same way.) They proposed some parametric modi-

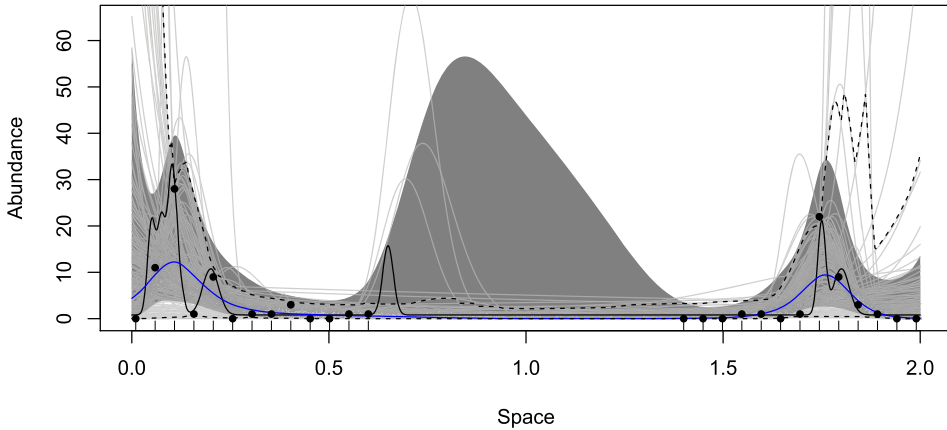


Figure 1. Comparison of bootstrap and analytical uncertainty for a Poisson process. The black line is the true intensity function (on the response scale) and points are observations. Blue line is a smooth of space, light grey wiggly lines are 500 bootstrap predictions, dashed lines are point-wise upper and lower 95% quantiles from the bootstrap, and the dark grey band is the analytical GAM confidence band using (6). The bootstrap appears confident that there is nothing in the unsampled area, but the analytical estimate illustrates how little we know (Color figure online).

fications (“type II” and “type III” bootstraps) that are more effective in the *IID* and block-structured situations that they consider. However, the underlying theory is complex (Carlin and Gelfand 1991; Carlin and Louis 2008) and it is far from clear whether simple yet reliable bootstraps can be devised for complicated multi-stage random effect situations like DSMs. Figure 1 shows a simple unidimensional Poisson process, sampled at either end but not in the middle (rug plot). Bootstrap replicates (shown in light grey, of which there are 500) largely fail to capture our uncertainty in the unsampled middle area. The analytical estimate (dark grey band) illustrates how little we know about the unsampled area.

The above does not imply that simple or indeed complicated bootstraps will *never* give reliable results in DSMs; given plenty of observations and good, uniform coverage, many approaches to inference will give similar and good results. However, it is sometimes not obvious whether this holds for a specific dataset, nor what to do bootstrap-wise if not. Instead, the (empirical) Bayesian framework of GAMs offers a coherent and general-purpose way to capture uncertainty.

5. A NEW MODEL FOR GROUP SIZE

Our variance propagation method so far works if detectability depends only on effort covariates, but not for individual covariates such as group size. Incorporating individual covariates in the detection function is not problematic, but it is not obvious how to allow for these different detection probabilities in the GAM. Further, it is not obvious how to combine predictions of different group sizes since average group sizes may vary spatially.

One approach is to use the Horvitz–Thompson-adjusted response described in the previous section, but as mentioned above this does not allow variance propagation. One could fit separate spatial models to subsets of the data for each group size, but it seems inefficient

to not share information between subsets of the data. Next, we show instead how to extend our variance propagation method to deal with group size.

We form M categories of group sizes, denoted $\{g_m; m = 1, \dots, M\}$, where groups within each category have similar detectability, and fit a detection function incorporating these group size categories. We then fit a GAM to an M -fold replicate of the dataset, with the response in the m^{th} replicate of the i^{th} segment being n_{im} , the number of groups in category m that were seen in that segment. (The total number of observations is unchanged; each observation is allocated to just one of the “replicates”.) Group size category (as a factor) is included as an explanatory variable, and smooths are modified to allow similar variations in density of groups with different sizes. There are no extra assumptions in this formulation from the model in Sect. 3, except to assume that the numbers of groups of different size categories in a given segment are independent, given the underlying density (which is allowed to vary with group size).

Factor-smooth interactions. We extend (2) to include multiple smooths of space which correspond to different categorizations of group size, so our model is:

$$\mathbb{E} \left[n_{i,g_m} | \boldsymbol{\beta}, \boldsymbol{\lambda}, p(\hat{\boldsymbol{\theta}}; z_i, g_m) \right] = a_i p(\hat{\boldsymbol{\theta}}; z_i, g_m) \exp \left(\beta_0 + f_{x_1, g_m}(x_{i1}) + \sum_{k=2}^K f_k(x_{ik}) \right), \quad (8)$$

for $m = 1, \dots, M$ where n_{i,g_m} is the number of observed groups in group class g_m in segment i and f_{x_1, g_m} is the spatial smooth (where x_1 is a spatial coordinate) for group size class g_m . Smoothers like f_{x_1, g_m} are referred to as *factor-smooth interactions* (Wood 2017; Pedersen et al. 2019). f_k are any other smooths (of covariates x_k , for $k > 1$). For clarity, we make the dependence on group size class explicit: $p(\hat{\boldsymbol{\theta}}; z_i, g_m)$, i.e. the probability of detection given segment-level detection covariates z_i and group size g_m .

There are a number of different possible forms for f_{x_1, g_m} . These vary in two main ways: (1) do levels share a smoothing parameter, or have separate ones? (2) do smooths tend towards a “global” smooth that dictates a general spatial effect? Here, we adopt the “fs” basis in `mgcv` which can be thought of as a smooth version of a random slopes model: smooths are generated for each factor level with smooths defined as deviations from a reference level, with all smooths sharing the same smoothing parameter. This is appealing as we might expect that the spatial smooths for each group size are similar but there might be some process that generates larger groups in certain places (e.g. large prey aggregations attracting large groups of animals). This approach is easily extended to other density covariates (e.g. x_1 could be bathymetry or vegetation cover).

Abundance and uncertainty estimation with group size smooths. Abundance is estimated by summing over the predictions for each group size category (\hat{N}_m) and weighting them by the corresponding mean group size (\bar{g}_m): $\hat{N} = \sum_{m=1}^M \bar{g}_m \hat{N}_m$. We can find $\text{Var}(\hat{N} | \bar{G})$ (where \bar{G} is the mean group size) from the variance propagation procedure above, but we need $\text{Var}(\hat{N})$, which we can obtain from the law of total variance:

$$\begin{aligned}\text{Var}(\hat{N}) &= \mathbb{E}_{\bar{G}}[\text{Var}(\hat{N}|\bar{G})] + \text{Var}[\mathbb{E}_{\bar{G}}(\hat{N}|\bar{G})] \\ &= \text{Var}(\hat{N}|\bar{G}) + \sum_{m=1}^M \text{Var}(\bar{G}_m) \hat{N}_m^2,\end{aligned}\tag{9}$$

where $\text{Var}(\bar{G}_m)$ reflects the uncertainty about mean group size within a category, to be estimated empirically from all the observed groups in that category. The effect of $\text{Var}(\bar{G}_m)$ on $\text{Var}(N)$ should be small (because categories are narrow, and mean must lie within category), and also should not vary much spatially, so no further spatial adjustment to that variance component is required.

6. EXAMPLES

Island Scrub-Jays. We first apply our variance propagation method in a simple situation where there is covariance between the abundance and detection processes that is the case of a spatially varying detection covariate. Island Scrub-Jays (*Aphelocoma insularis*) are endemic to Santa Cruz Island, California. Jays primarily reside in areas of chaparral and forest, though the density of this foliage also affects detectability. Sillett et al. (2012) model the distribution of Island Scrub-Jays from 307 point transects surveyed in fall 2008 and spring 2009. Distances were binned into three intervals due to responsive movement ($[0m - 100m]$, $(100m - 200m]$, $(200m - 300m]$). Proportion chaparral (`chap`) and proportion forest (`forest`) were available as covariates, as was elevation (`elev`). Sillett et al. fitted a hierarchical model assuming a negative binomial distribution for abundance and a multi-nomial detection process using a half-normal detection function. Their best models (by AIC) were: fall 2008 abundance modelled as $\beta_0 + \beta_1 \text{chap}^2 + \beta_2 \text{chap} + \beta_3 \text{elev}$, with detectability as a function of `chap`; spring 2009 abundance modelled as $\beta_0 + \beta_1 \text{chap}^2 + \beta_2 \text{chap} + \beta_3 \text{elev}^2 + \beta_4 \text{elev}$, detectability as a function of `forest`.

We replicated the analysis of Sillett et al. using our two-stage variance propagation approach to show that our method can be used in such a situation. In summary, final coefficient estimates were very close to those in the original paper, abundance estimates with associated 95% CIs were very similar for both seasons: Fall 2008 DSM $\hat{N} = 2272$ (1625–3175), Sillett et al. $\hat{N} = 2267$ (1613–3007) and Spring 2009 DSM $\hat{N} = 1684$ (1263–2246), Sillett et al. $\hat{N} = 1705$ (1212–2369); Supplementary Material A gives the comparison in full. For the spring model, the value of the `forest` coefficient in the detection function changed effect size from -0.18 (SE = 0.06) to -0.083 (SE = 0.062) after propagation (indicating no issue with our $\hat{\delta}$ diagnostic). By giving the GAM the flexibility to slightly adjust the detection function parameters via $\hat{\delta}$ (as opposed to treating the estimated detection probabilities as certain), the CV of the abundance estimate is actually improved in this case, from 18.4 to 14.8%.

The jay data present a particularly interesting case as the covariates in the GAM are fixed effects; there is therefore no “cost” (in terms of the penalized likelihood) to changing the GAM coefficients. We see minimal changes in the parameters of the fall model (Supplementary Material A, Table 1), as these are already well modelled (no doubt due to the good coverage of the data): the detection function includes `chap` and the GAM includes `chap`

and chap^2 , so any adjustment via $\hat{\delta}$ is a third-order effect. The spring model has different covariates in each model component, making the correction necessary.

The survey design had extremely good coverage over Santa Cruz Island. We decided to see what the effect of “unbalancing” the design would be to test the robustness of our model. We randomly sub-sampled the fall data to contain only 100 sites and then removed those where chaparral cover was greater than the mean chaparral proportion (over all points). Our sub-sample was left with 16 detections at 65 points. Fitting the fall DSM to the reduced data yields $\hat{N} = 26,434$ (95% CI 209-3,349,000; CV=2,100%). Post-variance propagation, we obtain $\hat{N} = 2,831$ (95% CI 39-206,000; CV=1,100%), both detection function and GAM coefficients having changed (see Supplementary Material A, Table 3). While we would expect a high variance for such a small and unbalanced dataset (and indeed we obtain this), our procedure tames the model to an extent, giving a more realistic estimate of abundance. Once information about both model components is allowed to inform the parameter estimation simultaneously, the coefficients are corrected.

Island Scrub-Jays Simulation. To assess performance of our variance propagation method with the delta method and a one-stage fully Bayesian approach, we conducted a simulation using the Island Scrub-Jay data as a starting point. We kept spatial coverage constant throughout the simulation settings but varied the detectability and therefore the number of observations available for the detection function component of the model. Full details of the simulation set-up are given in Supplementary Material B. Here, we note that our variance propagation method performed well in terms of bias in the abundance estimate and its corresponding variance estimate compared to the fully Bayesian model, even when sample size decreased (Supplementary Material B, Figure 2).

Harbour porpoise. To illustrate our new group size model, we reanalyse an aerial line transect survey of harbour porpoise in Irish Sea, coastal Irish waters, and Western coastal Scotland, where we see spatial variation in observed group size of 1 to 5 animals (typical for harbour porpoise; e.g. Siebert et al. 2006; points in Fig. 2). During SCANS-II aerial surveys, two observers recorded cetacean detections (along with sighting conditions) from bubble windows on both sides of a plane flying at 183m. Complete survey details and a comprehensive analysis is given in Hammond et al. (2013). For simplicity we assume certain detection on the trackline, no errors in group size estimation (less likely with aerial than in shipboard surveys for harbour porpoise; Phil Hammond, Debi Palka, *pers. comm.*, November 2017), and negligible island/coastline effects in the spatial model.

To fit our DSM, three group size bins were formed: size 1 (131 observations), 2 (35 observations) and 3-5 (14 observations). A hazard rate detection function was fitted to the observed distances (truncated at 300m) with the group size bin (g_m , $m = 1, \dots, 3$) and Beaufort (B_i , binned as 0-1, 2 and 3-5) as factor covariates. Detectability for each segment i per group size factor was then estimated from the detection function: $p(\hat{\theta}; B_i, g_m)$. Following (8), we fitted the DSM:

$$\begin{aligned} & \mathbb{E} \left[n_{i,g_m} \mid \boldsymbol{\beta}, \boldsymbol{\lambda}, p(\hat{\theta}; B_i, g_m) \right] \\ & = a_i p(\hat{\theta}; B_i, g_m) \exp(\beta_0 + f_{E,N,g_m}(E_i, N_i)), \end{aligned}$$

Table 1. Observed versus expected counts from the harbour porpoise DSM (post-variance propagation) at levels of Beaufort used in the detection function

	0-1	2	3-5
Observed	98.00	36.00	30.00
Expected	96.59	35.13	35.69

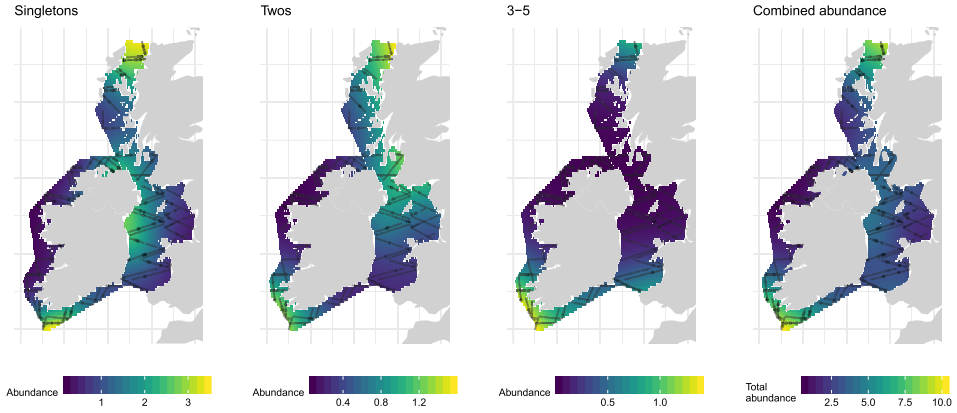


Figure 2. Predicted density surfaces from the new group size model for harbour porpoise. First three plots are density maps for the given group size (i.e. group abundance multiplied by mean group size), right plot shows the combined map, summing the previous three plots per prediction cell. We can see that distribution is roughly similar in all three group size categories though with almost no larger groups in the North, far more animals occurring as singletons than in larger groups.

where in segment i of area a_i the observed number of groups in size category g_m was denoted n_{i,g_m} . It was assumed the response was Tweedie-distributed where the power parameter was constrained to be greater than 1.2 to avoid numerical issues. Each f_{E,N,g_m} was a smooth of space (projected Easting/Northing; E_i, N_i) for group category g_m and had a maximum basis size of 20 (total maximum basis size for f_{E,N,g_m} was therefore 60).

The fitted model had a total effective degrees of freedom of 20.47 for f_{E,N,g_m} . GAM checking showed reasonable fit to the data. Table 1 shows observed vs expected counts by Beaufort—there is some misfit at the highest state, perhaps because detection probability at zero distance changes with Beaufort level (from Hammond et al.: 0.45 for Beaufort 0-1 and 0.31 in Beaufort 2-3). Plots of the per-group size bin predictions and the combined prediction are given in Fig. 2, which show some consistent patterns between group size classes (“hotspot” off Southern Ireland) and some differences (varying distribution in the Irish Sea and Western Scotland), this kind of insight is not possible using a single smooth for all observations and may prove useful in cases where there are occasional very large group sizes (e.g. oceanic dolphins). Using (7), the CV of abundance was estimated to be 2.36%, when our new variance propagation method was used the CV was estimated as 9.65%. The assumption of independence (via (7)) underestimates uncertainty in the case where group size (and detectability) vary in space. Only a small piece of code implementing (9) was required in addition to the `dsm_varprop` function (included in Supplementary

Material A), we then gain the ability to make inferences about group size spatial distribution (traditionally requiring two separate models, one for encounter rate, one for group size; e.g. Becker et al. 2014), as well as improve uncertainty estimation via variance propagation.

7. DISCUSSION

Combining the uncertainty from detection functions with that from spatial models has been a challenging problem for point and line transect analysis, requiring either complicated bespoke software that combines two model components, or ad hoc approaches that lack statistical justification. In this paper, we have demonstrated a simple, flexible, and statistically sound method that can (i) propagate uncertainty from detectability models to the spatial models for a particular class of detection function (i.e. those without individual-level covariates) *and* (ii) include group size as a covariate in the detection function while still being able to propagate uncertainty and address spatial variation in group size. Our methods are implemented in the `dsm` package for R but can be implemented in any standard GAM fitting software.

It is straightforward to apply our factor-smooth approach group size more generally to individual-level covariates which affect detectability and vary in space, but do not directly affect abundance, such as observable behaviour. For example, feeding groups might be more (or less) conspicuous than resting groups, and the proportion feeding/resting may vary across the surveyed region. Unless detectability is included in the analysis, biased abundance estimates could result, especially when survey coverage is non-uniform; and there has been no simple way until now to include such effects in the spatial model. A major advantage of our approach over simple (or complex) stratification schemes is that we are now sharing information between the levels of our categorized variable. This makes the results less sensitive to over-specifying the number of categories, as the model will shrink back towards the simpler model in the absence of strongly informative data. We also note that the factor-smooth approach could be applied to all smooth terms in the GAM, allowing for a very flexible model. This would be appropriate only if it was reasonable that all smooths vary according to the detectability covariate (e.g. feeding behaviour in our harbour porpoise example might depend on both space and depth).

We have assumed that all variables are measured without much error. Measurement error for individual-level covariates such as group size can be a serious problem in distance sampling (Hodgson et al. 2017)—distance between observer and group can affect not just detectability, but also the extent of group size error. If group size varies spatially, it is hard to see how to separate the spatial modelling stage from the distance sampling stage. A full discussion is beyond the scope of this paper, but we suspect that specially designed observation protocols and bespoke analyses may be the only way to tackle such thorny cases.

All-in-one fitting of both detection and spatial models is also possible (e.g. Johnson et al. 2009; Sillett et al. 2012; Yuan et al. 2017). If models are specified correctly, then the all-in-one approach could in theory be slightly more efficient, but only insofar as it takes account of third-order changes in the detection function likelihood (since our approach uses

a quadratic approximation). That seems unlikely to make much difference in general—and as is the case for the Island Scrub-jay example. Our own preference is therefore to use the two-stage approach, mainly because in our experience the careful fitting of detection functions is a complicated business which can require substantial model exploration and as few as possible “distractions” (such as simultaneously worrying about the spatial model). The two-stage process allows any form of detection function to be used, without having to make deep modifications to software. In summary, if one *knew* one had the correct model to begin with, one-stage fitting would be slightly more efficient, but this is never the case in practice.

It is valuable to check for any tension or confounding between the detection function and density surface parts of the model, which can occur if there are large-scale variations in sighting conditions across the survey region, and which is readily diagnosed in a two-stage model. Although this does not appear to lead to problems in the datasets we have analysed with the software described in this paper, we have come across it in other variants of line transect-based spatial models with different datasets. It may not be so easy to detect partial confounding when using all-in-one frameworks.

Finally, we note that the approach outlined here (of using a first-stage estimate as a prior for a second estimate, and propagating variance appropriately) is quite general and is comparable to standard sequential Bayesian approaches to the so-called integrated data models. The first-stage model need not be a detection function, but instead could be from another GAM (or other latent Gaussian model). Again, this allows us to ensure that first-stage models are correct before moving to more complex modelling. Modelling need not only be of two stages and could extend to multi-stage models (Hooten et al. 2019).

ACKNOWLEDGEMENTS

The authors thank Natalie Kelly, Jason Roberts, Eric Rexstad, Phil Hammond, Steve Buckland, and Len Thomas for useful discussions, Devin Johnson for the suggestion of this as a general statistical method, and Simon Wood for continued development of *mgcv* and GAM theory. The manuscript was greatly improved by comments from the editor and two anonymous reviewers. Data from the SCANS-II project were supported by the EU LIFE Nature programme (project LIFE04NAT/GB/000245) and governments of range states: Belgium, Denmark, France, Germany, Ireland, Netherlands, Norway, Poland, Portugal, Spain, Sweden, and UK. This work was funded by OPNAV N45 and the SURTASS LFA Settlement Agreement, and being managed by the US Navy’s Living Marine Resources program under Contract No. N39430-17-C-1982, US Navy, Chief of Naval Operations (Code N45), grant number N00244-10-1-0057 and the International Whaling Commission.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

REFERENCES

- Becker EA, Forney KA, Foley DG, Smith RC, Moore TJ, Barlow J (2014) Predicting seasonal density patterns of California cetaceans based on habitat models. *Endanger Species Res* 23(1):1–22
- Borchers DL, Zucchini W, Fewster RM (1998) Mark-recapture models for line transect surveys. *Biometrics* 54(4):1207
- Buckland ST, Anderson DR, Burnham KP, Borchers DL, Thomas L (2001) Introduction to distance sampling. Estimating abundance of biological populations. Oxford University Press, Oxford
- Carlin BP, Gelfand AE (1991) A sample reuse method for accurate parametric empirical Bayes confidence intervals. *J R Stat Soc Ser B*
- Carlin BP, Louis TA (2008) Bayesian methods for data analysis, 3rd edn. CRC Press, Boca Raton
- Goodman LA (1960) On the exact variance of products. *J Am Stat Assoc* 55(292):708
- Hammond PS, Macleod K, Berggren P, Borchers DL, Burt L, Cañadas A et al (2013) Cetacean abundance and distribution in European Atlantic shelf waters to inform conservation and management. *Biol Conserv* 164(C):107–122
- Hedley SL, Buckland ST (2004) Spatial models for line transect sampling. *J Agric Biol Environ Stat* 9(2):181–199
- Hodgson A, Peel D, Kelly N (2017) Unmanned aerial vehicles for surveying marine fauna: assessing detection probability. *Ecol Appl Publ Ecol Soc Am* 27(4):1253–1267
- Hooten MB, Johnson DS, Brost BM (2019) Making recursive Bayesian inference accessible. *The American Statistician*
- Johnson DS, Laake JL, Ver Hoef JM (2009) A model-based approach for making ecological inference from distance sampling data. *Biometrics* 66(1):310–318
- Laird NM, Louis TA (1987) Empirical Bayes Confidence Intervals Based on Bootstrap Samples. *J Am Stat Assoc* 82(399):739–750
- Marques FFC, Buckland ST (2003) Incorporating covariates into standard line transect analyses. *Biometrics* 59(4):924–935
- Marques TA, Buckland ST, Bispo R, Howland B (2012) Accounting for animal density gradients using independent information in distance sampling surveys. *Stat Methods Appl* 22(1):67–80
- Miller DL, Burt ML, Rexstad EA, Thomas L (2013) Spatial models for distance sampling data: recent developments and future directions. *Methods Ecol Evol* 4(11):1001–1010
- Pedersen EJ, Miller DL, Simpson GL, Ross N (2019) Hierarchical generalized additive models: an introduction with mgcv. *PeerJ*, p e6876
- Siebert U, Gilles A, Lucke K, Ludwig M, Benke H, Kock K-H, Scheidat M (2006) A decade of harbour porpoise occurrence in German waters—analyses of aerial surveys, incidental sightings and strandings. *J Sea Res* 56(1):65–80
- Sillett TS, Chandler RB, Royle JA, Kéry M, Morrison SA (2012) Hierarchical distance-sampling models to estimate population size and habitat-specific abundance of an island endemic. *Ecol Appl* 22(7):1997–2006
- Skaug HJ, Schweder T (1999) Hazard models for line transect surveys with independent observers. *Biometrics* 55(1):29–36
- Thomas L, Buckland ST, Rexstad EA, Laake JL, Strindberg S, Hedley SL et al (2010) Distance software: design and analysis of distance sampling surveys for estimating population size. *J Appl Ecol* 47(1):5–14
- Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J R Stat Soc Ser B Stat Methodol* 73(1):3–36

Wood SN, (2017) *Generalized Additive Models: An Introduction with R, Second Edition*. Chapman and Hall/CRC Texts in Statistical Science. CRC Press, Boca Raton

Yuan Y, Bachl FE, Lindgren F, Borchers DL, B IJ, Buckland ST et al (2017) Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. *Ann Appl Stat* 11(4):2270–2297

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.