Reliable variance propagation for spatial density surface models

Mark V. Bravington
Commonwealth Scientific and Industrial Research Organisation, Hobart, Australia

and

David L. Miller
Centre for Research into Ecological & Environmental Modelling and School of Mathematics and Statistics, University of St Andrews, St Andrews, Scotland
email: dave@ninepointeightone.net

and

Sharon L. Hedley
The Schoolhouse, Denhead, St Andrews, Scotland

Summary: Density Surface Models (DSMs) are two-stage models for estimating animal density from line-transect data. First, detectability is estimated by distance sampling, and then a Generalized Additive Model (GAM) is used to estimate animal density as a function of location and/or other spatially-varying environmental covariates. 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 a random effect in the GAM stage. This random effect acts as a quadratic approximation to the detection probability centred at the point estimate of detectability, effectively fitting the detection function at the same time as the spatial model. This allows straightforward computation of the overall variance via standard GAM machinery. A further extension allows for spatial variation in group size, which can be an important covariate for detectability. We illustrate these models using line transect survey data of minke whales and harbour porpoise from a visual line transect survey of European waters.

Key words: abundance estimation; distance sampling; generalized additive models; line transect sampling; spatial modelling; wildlife surveys; group size.

1. Introduction

Line transect surveys are a standard method to survey both marine and terrestrial populations and estimate their abundance (Buckland et al., 2001). Data from these surveys can be used to build spatially-explicit estimates of abundance. These estimates need to take into account the spatial distribution of the population, as well as detectability. Here we consider density surface models (DSMs; Hedley and Buckland, 2004; Miller et al., 2013). DSMs are two-stage models: first a detection function is used to model detectability, followed by spatial model (in our case a generalized additive model; GAM; Wood, 2017) to model the effect of spatially-referenced environmental covariates on density. To describe density, the GAM usually includes the coordinates of transect locations and often other spatially-referenced environmental covariates such as sea surface temperature or bathymetry. Each transect is broken into short segments, and the response variable is the count of observed animals per segment. Probability of detection for each segment estimated from the fitted detection function enters the GAM as part of the offset to adjust the effort, giving the effective area surveyed per segment.

This two-stage approach is convenient, because it uses well-established software and diagnostic tests for each part of the problem. However, it leads to complications when attempting to estimate uncertainty in abundance estimates, as uncertainty from the detection function must be propagated through to the spatial model. It is awkward to jointly fit the detection function and GAM simultaneously: this would require special purpose fitting routines. Conceptually however, we are fitting both models at once, except we replace the detection function log-likelihood by a quadratic approximation centred on the point estimate, which enters the GAM directly as an additional penalty term.

Following a summary of DSMs and notation in Section 2, we present our new formulation in Section 3, including variance computation and diagnostics. Section 4 comments on problems with existing approaches to variance propagation in DSMs. In Section 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 presents two examples, with and without group size. Some discussion is given in Section 7.

2. Density surface models

We focus on modelling line transect distance sampling data: observers move along a set of survey lines, counting (groups of) animals, recording distances from the line to the observed groups (or their cues, such as blows for cetaceans), and the size of each detected group. Using the distances of each group from the trackline, and perhaps other data from each sighting such as the bearing, we first estimate the overall probability of detection, which may also depend on covariates of the group itself (e.g., the number of animals in it) and of the conditions for observing at that time. In the second stage of a DSM, we fit a GAM to the number of groups seen within each block of observation effort, as a smooth function of spatial covariates; the estimated detection probabilities from the first stage become offsets to allow for incomplete detection.

Without loss of generality, we use the term GAM throughout to refer to the spatial part of the DSM, though the model could be a generalized linear mixed model (GLMM), or generalized additive mixed model (GAMM). Any model term that can be written as the product of a design matrix and a coefficient vector on the linear predictor scale, penalised by a sum of quadratic terms may be represented this way.

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. \( x \) is required for prediction as well as fitting, and is assumed known across the entire region of interest. (2) Effort covariate(s), \( z \) (or \( z \)), affect detection probability: e.g., sea conditions measured on the Beaufort scale, observer name; 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. \( g \) 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 \( G \) 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 trackline. In certain settings, \( y \) may contain other elements. For example, in a multi-observer-platform survey (e.g., MRDS; Buckland et al. [2004]), \( y \) might also include which of the active observers saw the group; in a cue-based setting, \( y \) might include the bearing between sighting and trackline.

These classes are assumed to be mutually exclusive; overlaps lead to fundamental problems for distance sampling which we do not address here (such as non-uniform animal distribution within the sample unit; e.g., 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 \( \theta \) as well as \( z \) and \( g \), describes the probability of making an observation at some distance \( y \). The parameters \( \theta \) are usually estimated by maximizing this log-likelihood across observations \( s \):

\[
\ell (\theta) = \sum_s \log \left( \frac{\pi (y_s; \theta, z_s, g_s)}{p (\theta; z_s, g_s)} \right)
\]

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

\[
p (\theta; z, g) = \int \pi (y; \theta, z, g) dF_Y (y)
\]

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 MCDS (Marques and Buckland, [2003]) with \( z \) and \( g \), multi-observer MRDS (Borchers et al., [1998]), and cue-based "hazard probability models" (Skag and Schweder, [1999]).

The second part of DSM discards the per-observation variables \( y \), and 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 transects are generally very long in comparison to 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; environmental covariates are assumed not to change much within each snippet. A GAM is used to model the relationship between counts per segment and spatial covariates as an additive combination of smooth functions of density covariates (e.g., altitude, salinity, temperature etc):

\[
\mathbb{E} [n_i | \beta, \lambda, p(\theta; z_i)] = a_i p(\theta; z_i) \exp \left( \beta_0 + \sum_k f_k(x_{ik}) \right), \tag{1}
\]

where the number of individuals per segment (of area \( a_i \)), \( n_i \), follows some count distribution such as quasi-Poisson, Tweedie
or negative binomial and we assume a log link. The \( f_k \) are smooth functions of environmental covariates, \( x_{ik} \), represented by a basis expansion (i.e., \( f_k(x) = \sum \beta_j b_j(x) \) for some basis functions \( b_j \)) penalized by a (sum of) quadratic penalty (or penalties); \( \beta_0 \) is an intercept term, included in parameter vector \( \beta \); \( \lambda \) is a vector of smoothing parameters which control the wiggliness of the smooth components of the model. We take a Bayesian interpretation of GAMs (Wood 2017), in which \( \lambda \) controls the variance of a multivariate improper Gaussian prior (Marra and Wood 2011):

\[
\beta \sim N \left( 0, \phi \left( \sum \lambda_i S_i \right)^{-1} \right),
\]

with \( \phi \) the scale parameter in the mean-variance relationship, \( \lambda_i \) the smoothing parameter and \( S_i \) the penalty matrix for the \( i^{th} \) term, and \( - \) indicates pseudoinverse. We estimate \( \lambda \) via REML (Wood 2011), an empirical Bayes procedure. Fully Bayesian approaches, placing a hyperpriors on \( \lambda \) are also possible. Ignoring group size for now we write \( p(\theta; z_i) \) (we return to group size in Section 5). We may think of the product \( a_i p(\theta; z_i) \) as the effective area of segment \( i \), analogous to the effective strip (half) width (Buckland et al. 2001).

The phrase “estimating variance for density surface models” refers to the variance (or standard error or coefficient of variation) of a predicted abundance estimate, \( \hat{N} \). We assume below that we have created some prediction grid with all density covariates \( x \) available for each cell in the grid. Abundance is predicted for each cell and sum the cells for an overall abundance, \( \hat{N} \), over some region of interest which may not be the entire surveyed area. Although \( p(\theta) \) does not appear explicitly in the prediction:

\[
\hat{N} = \sum_i a_i \exp \left( \hat{\beta}_0 + \sum_k \hat{f}_k(x_{ik}) \right),
\]

the GAM offsets \( p(\theta) \) clearly do affect \( \hat{\beta} \), so it is important to account somehow for detection probability uncertainty. The GAM in (1) assumes a fixed offset, so extra steps are required.

### 3. Variance propagation for Density Surface Models

In this section, for simplicity, we assume that there are no individual-level covariates (e.g., that group size is always 1) so that we omit \( g \) from the equations (it is reintroduced in Section 5). Let \( p(\theta_0, z_i) \) be the true probability of detection in segment \( i \). If \( \theta_0 \) is the true (unknown) value of \( \theta \), and \( \hat{\theta} \) is its MLE, we use the shorthand \( p_i = p(\theta_0; z_i) \) and \( \hat{p}_i = p(\hat{\theta}; z_i) \) when the dependence is clear. The expected number of encounters in the segment is \( a_i p_i \rho_i \) where \( \rho_i \) is the underlying density of groups in the segment.

Given \( p_i \), we can re-write (1) on the log link scale as:

\[
\log E[n_i|\beta, \lambda, p_i] = \eta_i = \log a_i p_i + X_i \beta. \tag{2}
\]

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

\[
\eta_i = \log a_i + \log \hat{p}_i + \log p_i - \log \hat{p}_i + X_i \beta
\]

and then take a Taylor series expansion of \( \log \hat{p}_i \equiv \log p(\hat{\theta}; z_i) \) about \( \theta = \theta_0 \):

\[
\log p(\hat{\theta}; z_i) = \log p(\theta_0; z_i) + (\hat{\theta} - \theta_0)^\top \left[ \frac{d \log p(\theta; z_i)}{d \theta} \right]_{\theta = \theta_0} + O((\hat{\theta} - \theta_0)^2). \tag{3}
\]

By defining the vectors \( \delta \triangleq \hat{\theta} - \theta_0 \) and \( \kappa_i \triangleq \frac{d \log p(\theta; z_i)}{d \theta} \bigg|_{\theta = \theta_0} \), we can rewrite (2) as

\[
\log E[n_i|\beta, \lambda, p_i] = \log a_i p_i + X_i \beta + \kappa_i \delta + O(\delta^2). \tag{4}
\]

Asymptotic arguments suggest that the \( O(\delta^2) \) term can be neglected for reasonable sample sizes.

Taking a “facultative Bayesian” perspective on detection functions inference, we have approximately that

\[
\theta_0 \sim N(\hat{\theta}, V_{\theta}) \implies \delta \sim N(0, V_{\delta}),
\]

where the observation data determine \( \hat{\theta} \) and the covariance matrix \( V_{\theta} \) (the negative inverse Hessian of the log-likelihood for the detection function). In other words, the posterior distribution on \( \theta \) from fitting the detection function now becomes a prior distribution for fitting the density surface. To first order, \( \delta \) then plays the same structural role in (3) as the smoother coefficients \( \beta \). The design matrix for \( \delta \) (\( \kappa \) in (4)) is obtained by differentiating the log-detection probabilities, with respect to \( \theta \) at its estimated value. Derivatives do not need to be accurate, and need only be computed once, so simple 3-point numerical differentiation is perfectly adequate. The only other requirement is the Hessian from the detection function log-likelihood, which should be an automatic output from distance sampling stage regardless of the complexity of the model. This method can be applied automatically to almost any distance sampling setup. A GAM fitting routine can then be used to simultaneously
estimate $\beta$ and $\delta$. Posterior inferences about $\beta$ (and thereby about animal density $\rho$ and abundance) automatically propagate the uncertainty from fitting the detection function. The approximation neglects higher-order terms in $\delta$; we argue below that this will be unimportant when there are enough sightings to estimate the detection functions reliably.

The only technical difference from fitting a standard GAM, is that $\lambda$ is usually unknown and has to be estimated (i.e., the prior on $\beta$ has known covariance, but unknown scale), whereas the prior on $\delta$ is completely determined from the detection function fitting (i.e., in effect $\lambda_\delta = 1/\phi$, where $\phi$ is the scale parameter). This setup 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 $\phi$ is fixed (as opposed to estimated). With the per-segment count data of density surface estimation, a fixed $\phi$ (of 1) would correspond to a purely Poisson response, which might be appropriate if maximum count per segment is only 1 or 2. In our experience, though, better fits can usually be obtained by estimating $\phi$ and using a Tweedie response distribution. In order to implement (4) for a general response distribution using $\text{mgcv}$ (Wood, 2011), we therefore use a one-dimensional search over $\phi$ to maximize the marginal REML. At each iteration, given the working value $\phi^*$, we re-fit the GAM fixing $\phi = \phi^*$ and $\lambda_\delta = 1/\phi^*$. Speed can be improved by re-using some of the setup computations (design matrices, etc) at each iteration. The function $\text{dsm}_\text{varprop}$ in the R package $\text{dsm}$ can be used to propagate uncertainty as above using the delta method and return the variance along with diagnostic checks.

3.1 Diagnostics and interpretation of $\hat{\delta}$

Posterior modes $\hat{\delta}$ are available after fitting, and may help with diagnosing inconsistencies between the detection function and spatial model stages. For example, if weather is systematically worse in some parts of the survey region, then both $\beta$ and $\delta$ will contribute to the expected pattern of sightings, and the two sets of parameters will be somewhat confounded. Fortunately, because both $\beta$ and $\delta$ are penalized (i.e., partly constrained by their prior distributions), the spatial model has limited ability to affect $\delta$, but it is nevertheless worth checking whether there is any such tension between the two parts of the model. This can be checked in a couple of ways. The first is to compare the inferred spatial distribution and abundance from fitting (2) with the “naive” estimates where detection uncertainty is ignored and the offset $a_i\hat{p}_i$ is treated as exact. 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.

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 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.

3.2 Calculating $\text{Var}(\hat{N})$

We now discuss two methods to calculate the variance of $\hat{N}$.

**Posterior simulation.** We can estimate uncertainty using the distributional properties of the estimated parameters of the GAM (Wood, 2017, Section 6.9.3). The posterior for $\beta$ given data $y$ and smoothing parameters $\lambda$, are distributed as $\beta|y, \lambda \sim N(\beta, V_\beta)$; where $V_\beta$ is the covariance matrix for the GAM coefficients (Wood, 2017, Section 5.8). We also form the prediction matrix, $X_\rho$, which maps model coefficients to values of the linear predictor for the prediction data so $\hat{\eta}_p = X_p\hat{\beta}$ (Wood, 2017, Section 6.10). The algorithm to generate $B$ samples is then:

1. For $b = 1, \ldots, B$ do the following:
   - (a) Simulate from $N(\hat{\beta}, V_\beta)$, to obtain $\hat{\beta}_b$.
   - (b) Calculate new predicted abundance as $\hat{N}_b = a_p \exp(X_p\hat{\beta}_b)$ (where $a_p$ is a row vector of areas for the prediction cells).

2. Calculate the empirical variance of the $\hat{N}_b$s.

In practice $B$ does not have to be particularly large. Marra et al. (2011) achieve reasonable results with $B = 100$.

**The delta method.** One can avoid the computational burden of posterior simulation (which can be large for complex models) by using the delta method:

$$\text{Var}(\hat{N}) = \left(a_p \frac{\partial \exp X_p\beta}{\partial \beta} \bigg|_{\beta = \hat{\beta}} X_p\right) V_\beta \left(a_p \frac{\partial \exp X_p\beta}{\partial \beta} \bigg|_{\beta = \hat{\beta}} X_p\right)^\top,$$

derivatives are evaluated at the estimated values of the model parameters.
4. Previous methods for estimating uncertainty in Density Surface Models

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

\[
\text{Var}_P(\log N) = \text{E}_P[\text{Var}(\log N|P)] + \text{Var}_P[E(\log N|P)]
\]

\[
\approx \text{Var}(\log N|\theta, z_i) + \text{Var}_P[\log N(p(\theta, z_i))],
\]

where \( P \) here is a random variable for the probability of detection and the subscripts indicate the expectation/variance taken over that variable. The first part of this can be derived from GAM theory and is detailed below, the second is more tricky.

Assuming independence. When \( p(\theta, z_i) \) 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 from the detection function model using the fact that squared coefficients of variation (CV) satisfy the property:

\[
\text{CV}^2(\hat{N}) = \text{CV}^2(\hat{N}_{GAM}) + \text{CV}^2(\hat{p}).
\]

Where \( \text{CV}_{IND}(\hat{N}) \) is the total coefficient of variation of the abundance estimate \( \hat{N}_{GAM} \) is the CV of the abundance estimate from the spatial model and \( \text{CV}(\hat{p}) \) is the CV of the probability of detection. Hence we may calculate \( \text{Var}_{IND}(\hat{N}) \) as:

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

To calculate \( \text{CV}(\hat{p}) \) when there are no covariates in the detection function we calculate \( \text{Var}(\hat{p}) \) using:

\[
\text{Var}(\hat{p}) = \left( \frac{\partial \hat{p}}{\partial \theta} \right) \text{V}_{\hat{p}} \left( \frac{\partial \hat{p}}{\partial \theta} \right)^T,
\]

where \( \text{V}_{\hat{p}} \) is the variance-covariance matrix for the detection function parameters.

In the case where detectability is a function of covariates, it is hard to justify the use of the CV decomposition as correlations between the spatial distribution and the covariates affect detectability. One method that has been implemented in program Distance ([Thomas et al. 2010]) is to use the Horvitz-Thompson-corrected counts per segment as the response (thus removing the detectability from the right hand side of (1)). To find variance we would then calculate a probability of detection averaged over the observations by first calculating the Horvitz-Thompson estimate of the abundance in the covered area (\( \hat{N} = \sum g_i/\hat{p}_i \), 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 (unless there are no covariates in the detection function). Additionally, there is no coherent way to generalize the formula (which affects both parametric and non-parametric cases) is specific to the use of GAMs and other mixed- or random-effect spatial models. The basic problem is that (most) bootstraps use only the posterior modes of random effects, 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 non-parametrically. 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 about the density in that block. The correct inference would of...
Figure 1. Example comparison of bootstrap and analytical uncertainty for a simple one dimensional Poisson process. The black line is the true intensity function for the Poisson process, put on the response scale, rug plot shows sample locations and points are observations. The blue line is the fitted model (smooth of the space coordinate in a Poisson GAM) and light grey wiggly lines are 500 bootstrap predictions, dashed lines are pointwise upper and lower 95% quantiles from the bootstrap, whereas the dark grey band is the analytical GAM confidence band (using the delta method as given in Section 3.2). The bootstrap appears confident that there is nothing in the unsampled middle area, whereas the analytical estimate illustrates how little we know about the unsampled area.

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 “naive 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 modifications (“type II” and “type III” bootstraps) that are more sensible in the limited situations 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.

The GAM formulation, whereby each random effect $\beta_k$ spreads its influence across a wide area and the posterior distributions of the $\beta_k$ are not independent, disguises the “naive bootstrap” issue but does not fully remove it, especially in line transect settings. 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.

Of course, the above does not imply that simple or indeed complicated bootstraps will never give reliable results in line transect density surface modelling; given plenty of observations and good even 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) Bayes framework of GAMs offers a coherent and general-purpose way to capture uncertainty which we explore next. See Wood (2017, Section 6.10.3) for further comment.
5. A new model for group size

Our variance propagation method addresses when detectability depends only on effort covariates, 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 if the response is “number of groups”. Further, it is not obvious how to combine predictions of different group sizes since average group sizes may vary spatially (Ferguson et al., 2006).

One approach is to use Horvitz-Thompson estimates of abundance per segment as the response in the model, correcting for detectability in the response rather than incorporating it via the offset of the model (Hedley and Buckland, 2004), but this does not allow variance propagation. One could fit a separate spatial model to subsets of the data for each group size, though it seems inefficient to not share information between subsets of the data. To account for group size variation in space and in detectability we propose a new model that allows us to propagate variance in the manner described above.

We form $M$ categories of group sizes, denoted $\{g_m; m = 1, \ldots, M\}$, where groups within each category have similar detectabilities 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”). The explanatory (density) covariates are expanded to include group size category as a factor, 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 Section 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 (1) to include multiple smooths of space which correspond to different categorizations of group size. A separate smooth can be fitted to each group size category and our model is:

$$\begin{align*}
E[n_{i,gm}|\beta, \lambda, p(\hat{\theta}; z, g_m)] &= n_i p(\hat{\theta}; z, g_m) \exp \left( \beta_0 + f_{xy,gm}(x_i, y_i) + \sum_k f_k(w_{ik}) \right).
\end{align*}$$

for $m = 1, \ldots, M$ where $n_{i,gm}$ is the number of observed groups in group class $g_m$ in segment $i$ and $f_{xy,gm}$ is the spatial smooth for group size class $g_m$. $f_k$ are any other terms in the model. For clarity we make the dependence on group size class explicit: $p(\hat{\theta}; z, g_m)$, i.e., the probability of detection given segment-level detection covariates $z$ and group size $g_m$. In practice we duplicate the full segment data as many times as there are group size categories (so our model now has $M \times$ number of segments rows of data) and fill in the response as the number of groups in that category that occur in that segment.

There are a number of different possible forms for $f_{xy,gm}$. These include completely separate spatial effects for each group size category or smooths for each category that tend toward a “global” smooth that dictates a general spatial effect. We assume here that per-category smooths have the same amount of smoothness but we can also let each per-category smooth have a different smoothing parameter. Though we discuss smooths of space here, this methodology is general and one could construct models with different smooths of any spatial covariate per category of group size, for example letting smooths of depth vary by group size — though one would require some biological justification for such a model. These smooths are referred to as factor-smooth interactions (Baayen et al., 2016; Wood, 2017). Here we adopt the “fa” basis in mgcv which can be thought of as a non-linear version of a random slopes model: smooths are generated for each factor level with some reference level, other levels measure the deviation from that reference level. Combining the reference level smooth with a smooth for a particular level will give the resulting smooth for that level. Here each of the levels has 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).

Abundance and uncertainty estimation with group size smooths

We can estimate abundance by simply summing over the predictions for each group size category ($\bar{N}_m$) and weighting them by the corresponding mean group size ($\bar{g}_m$):

$$\bar{N} = \sum_{m=1}^M \bar{g}_m \bar{N}_m.$$ 

We can find $\text{Var}(\bar{N}|S)$ from the variance propagation procedure above, but we need $\text{Var}(\bar{N})$, which we can obtain from the law of total variance:

$$\begin{align*}
\text{Var}(\bar{N}) &= \text{E}_G[\text{Var}(\bar{N}|G)] + \text{Var}[\text{E}_G(\bar{N}|G)] \\
&= \text{Var}(\bar{N}|G) + \sum_{m=1}^M \text{Var}(\bar{G}_m) \bar{N}_m^2, 
\end{align*}$$

where $\text{Var}(G_m)$ is the variance within category $m$, estimated by its empirical variance. The effect of $\text{Var}(G_m)$ on $\text{Var}(N)$ should be small (because categories are narrow, and mean must lie within category), and also should not vary much spatially,
Table 1

Predicted number of minke whales per linear kilometre of transect by observed Beaufort, as categorised in the detection function. Number of segments at each sea state category is listed in the final column.

| Beaufort | Predicted abundance per km | \(\hat{p}(\text{Beaufort}_i)\) | Segments |
|----------|-----------------------------|-------------------------------|----------|
| [0,1]    | 0.026                       | 0.508                         | 255      |
| (1,2]    | 0.008                       | 0.216                         | 316      |
| (2,4]    | 0.007                       | 0.247                         | 386      |

Table 2

Changes in estimated detectability for the levels of Beaufort between the fitted model denoted as \(p(\hat{\theta}, \text{Beaufort})\), its standard error and the “corrected” detectability from the variance propagation model denoted \(p(\hat{\theta} + \hat{\delta}, \text{Beaufort})\) for the minke whale data. Beaufort is treated as a factor in the analysis, hence detectability is non-monotonic with increasing sea state.

| Beaufort | \(p(\hat{\theta}, \text{Beaufort})\) | se(\(p(\hat{\theta}, \text{Beaufort})\)) | \(p(\hat{\theta} + \hat{\delta}, \text{Beaufort})\) |
|----------|-----------------------------------|------------------------------------------|---------------------------------|
| [0,1]    | 0.508                             | 0.115                                    | 0.482                           |
| (1,2]    | 0.216                             | 0.078                                    | 0.203                           |
| (2,4]    | 0.247                             | 0.088                                    | 0.304                           |

Table 3

Observed versus expected counts from the minke whale DSM at levels of Beaufort used in the detection function.

| Beaufort | Observed | Expected |
|----------|----------|----------|
| [0,1]    | 41.00    | 42.58    |
| (1,2]    | 14.00    | 15.28    |
| (2,4]    | 20.00    | 14.59    |

6. Examples

We re-analyse line transect data for two species from the SCANS-II surveys in European Atlantic waters. For survey details and a serious analysis see [Hammond et al. 2013]. To keep our analysis simple, we assume certain detection on the trackline and do not account for imperfect detection on the trackline (via e.g., mark-recapture distance sampling; [Burt et al. 2014]). We ignore island/coastline effects in the spatial model which could be accounted for by methods such as [Wood et al. 2008]. We computed a diagnostic for both models comparing observed versus expected counts at the observed levels of observation-level covariates (in addition to usual distance sampling/GAM diagnostics). Models were fitted using the dsr package in R, using the dsr_vardprop function to calculate the variance of \(\bar{N}\).

6.1 Minke whales

We re-analyse the minke whale (Balaenoptera acutorostrata) shipboard survey data, including Beaufort as a covariate in the hazard-rate detection function (recorded at each segment, but binned for analysis into [0, 1], [1, 2], [2, 4] and restricting data to those segments where sea state was \(\leq 4\)). The following spatial model was then fitted:

\[
\mathbb{E}[n_i | \beta, \lambda, p(\hat{\theta}; \text{Beaufort}_i)] = \alpha(p(\hat{\theta}; \text{Beaufort}_i)) \exp(\beta_0 + f_{\text{space}}(\text{Easting}_i, \text{Northing}_i)),
\]

where \(n_i \sim \text{Tweedie}\), constrained to have a minimum power parameter of 1.2 since values in the range 1–1.2 make little practical difference to variances but do lead to numerical problems. Easting, Northing, are projected coordinates. Distances were truncated at 870m for detection function fitting, as in [Hammond et al. 2013]; we ignore group size as most group sizes were predominantly 1 (total 68 observations, with 3 of size 2 and 1 of size 5, all others singletons). Table 2 shows the number of predicted minke whales per linear kilometre of transect from the model per level of Beaufort, here we see that minke density is very much correlated with weather conditions, so we need to account for this when calculating variance.

Table 2 compares estimated detectability at different Beaufort levels from the fitted detection function with those “corrected” during the fitting of the variance propagation model; the table shows there is little difference in the probability of detection when refitting the variance model. The coefficient of variation for the North Sea area was 0.2465 when using our new variance propagation method and 0.2643 when using (5). Estimates using variance propagation are lower than those using the delta method; this indicates that there is some negative covariance between density and detectability. \(\hat{p}\) increases in the most severe sea state from the middle level (last row of Table 2), so this feature is perhaps driving covariance between the spatial model and detection function.

6.2 Harbour porpoise

Harbour porpoise (Phocoena phocoena) aerial survey data from SCANS-II in the Irish Sea, coastal Irish waters and Western coastal Scotland (strata O, R and N, respectively) was used to illustrate the group size model. Aerial surveys were preferred as error in group size is likely less of an issue than in shipboard surveys for harbour porpoise (Phil Hammond, Debi Palka, personal communication, November 2017). Three group size bins were formed, corresponding to groups of size 1 (131 observations), 2
7. Discussion

We have described a multi-stage modelling approach, where detection functions are first fitted to distance sampling data before being used to correct a spatial model for detectability. Here we show how to (i) propagate uncertainty from the detectability model to the spatial part of the model 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. These methods are implemented in the dsm package for R but can be implemented in any standard GAM fitting software. We also have reiterated the issues with using bootstraps for estimating uncertainty in random effects models, a model class that includes DSMs.

It is straightforward to apply our group size approach to covariates which affect detectability and vary in space, but do not directly affect abundance; e.g., behaviour: if feeding groups are less conspicuous (by being underwater) than resting groups (at the surface) of similar size, and feeding is more prevalent in some parts of the surveyed area.

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). The most fundamental problem is that distance between observer and group can affect not just detection probability, but also the extent of group size error. If group size tends to vary 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., Royle et al., 2004; Johnson et al., 2009; Yuan et al., 2017), and in theory if all models are specified correctly the all-in-one approach could be more efficient. Nevertheless, our own preference is for 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 having to worry about spatial models at the same time); the two-stage process also allows essentially any special-purpose 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 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 (Section 5.3), this 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.

Supplementary Materials

R code implementing the group size model in section 5 along with data and code for the minke whale and harbour porpoise analyses in section 6 are available with this paper at https://github.com/dill/varprop-suppmaterials.
Predicted density surfaces from the new group size model for harbour porpoise. Top row and bottom left are the abundance of animals for the given group size (i.e., group abundance multiplied by mean group size), bottom right 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.

Acknowledgements
The authors thank Natalie Kelly, Jason Roberts, Eric Rexstad, Phil Hammond, Steve Buckland and Len Thomas for useful discussions, and Simon Wood for continued development of mgcv and GAM theory. Data are from the SCANS-II project 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. DLM was funded by OPNAV N45 and the SURTASS LFA Settlement Agreement, and being managed by the U.S. 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, International Whaling Commission.

References

Baayen, R. H., van Rij, J., de Cat, C., and Wood, S. N. (2016). Autocorrelated errors in experimental data in the language sciences: Some solutions offered by Generalized Additive Mixed Models. arXiv.org.

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

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

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. (2004). Advanced Distance Sampling. Estimating abundance of biological populations. Oxford University Press, Oxford, UK.

Burt, M. L., Borchers, D. L., Jenkins, K. J., and Marques, T. A. (2014). Using mark-recapture distance sampling methods on line transect surveys. Methods in Ecology and Evolution 5, 1180–1191.

Carlin, B. P. and Gelfand, A. E. (1991). A sample reuse method for accurate parametric empirical Bayes confidence intervals. Journal of the Royal Statistical Society: Series B.

Carlin, B. P. and Louis, T. A. (2008). Bayesian Methods for Data Analysis, Third Edition. CRC Press.

Ferguson, M. C., Barlow, J., Fiedler, P., Reilly, S. B., and Gerrodette, T. (2006). Spatial models of delphinid (family Delphinidae) encounter rate and group size in the eastern tropical Pacific Ocean. Ecological Modelling 193, 645–662.

Goodman, L. A. (1960). On the Exact Variance of Products. Journal of the American Statistical Association 55, 708.

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

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

Hodgson, A., Peel, D., and Kelly, N. (2017). Unmanned aerial vehicles for surveying marine fauna: assessing detection probability. Ecological applications : a publication of the Ecological Society of America 27, 1253–1267.

Johnson, D. S., Laake, J. L., and Ver Hoef, J. M. (2009). A Model-Based Approach for Making Ecological Inference from Distance Sampling Data. Biometrics 66, 310–318.

Laird, N. M. and Louis, T. A. (1987). Empirical Bayes Confidence Intervals Based on Bootstrap Samples. Journal of the American Statistical Association 82, 739–750.

Marques, F. F. C. and Buckland, S. T. (2003). Incorporating covariates into standard line transect analyses. Biometrics 59, 924–935.

Marra, T. A., Buckland, S. T., Bispo, R., and Howland, B. (2012). Accounting for animal density gradients using independent information in distance sampling surveys. Statistical Methods & Applications 22, 67–80.

Marra, G., Miller, D. L., and Zanin, L. (2011). Modelling the spatiotemporal distribution of the incidence of resident foreign population. Statistica Neerlandica 66, 133–160.

Marra, G. and Wood, S. N. (2011). Practical variable selection for generalized additive models. Computational Statistics and Data Analysis 55, 2372–2387.

Miller, D. L., Burt, M. L., Rexstad, E. A., and Thomas, L. (2013). Spatial models for distance sampling data: recent developments and future directions. Methods in Ecology and Evolution 4, 1001–1010.

Royle, J. A., Dawson, D. K., and Bates, S. (2004). Modeling abundance effects in distance sampling. Ecology 85, 1591–1597.

Skaug, H. J. and Schweder, T. (1999). Hazard models for line transect surveys with independent observers. Biometrics 55, 29–36.

Thomas, L., Buckland, S. T., Rexstad, E. A., Laake, J. L., Strindberg, S., Hedley, S. L., et al. (2010). Distance software: design and analysis of distance sampling surveys for estimating population size. Journal of Applied Ecology 47, 5–14.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73, 3–36.

Wood, S. N. (2017). Generalized Additive Models: An Introduction with R, Second Edition. Chapman & Hall/CRC Texts in Statistical Science. CRC Press.

Wood, S. N., Bravington, M. V., and Hedley, S. L. (2008). Soap film smoothing. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70, 931–955.

Yuan, Y., Bachl, F. E., Lindgren, F., Borchers, D. L., B., I. J., Buckland, S. T., et al. (2017). Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. The Annals of Applied Statistics 11, 2270–2297.

Received November 2018. Revised .... Accepted ....