

1 **Running title:** Spatial models for distance sampling
2 **Number of words:** ~4158
3 **Number of tables:** 0
4 **Number of figures:** 6
5 **Number of references:** 27

6 **Spatial models for distance sampling data:**
7 **recent developments and future directions**

8 **David L. Miller^{1*}, M. Louise Burt²,**
9 **Eric A. Rexstad², Len Thomas².**

- 10 *1. Department of Natural Resources Science, University of Rhode Island,*
11 *Kingston, Rhode Island 02881, USA*
12 *2. Centre for Research into Ecological and Environmental Modelling,*
13 *The Observatory, University of St. Andrews, St. Andrews KY16 9LZ, UK*

14 ***Correspondence author. dave@ninepointeightone.net**

Summary

1. Our understanding of a biological population can be greatly enhanced by modelling their distribution in space and as a function of environmental covariates. Model-based inference may also be used to obtain abundance estimates from non-randomly designed surveys.
2. Density surface modelling achieves both of the above aims. DSMs combine distance sampling to account for uncertain detection and a spatial model for the effects of environmental covariates.
3. We offer a comparison of recent advances in the field and consider the likely directions of future research. In particular we consider recent advances in the spatial modelling literature that may be advantageous to applied ecologists.
4. The methods discussed are freely available in R packages developed by the authors.

Keywords: distance sampling, line transect sampling, point transect sampling, population abundance, population density, spatial modelling, wildlife surveys

33 Introduction

34 When surveying biological populations it is increasingly common to record
35 spatially referenced data; for example: coordinates of observations, habitat
36 type, altitude or (if at sea) bathymetry. Spatial models allow for the vast
37 databases of spatially-referenced data to be harnessed, allowing for interac-
38 tions between environmental covariates and population densities to be invest-
39 igated. Mapping the spatial distribution of a population can be extremely
40 useful, especially when communicating results to non-experts. Recent ad-
41 vances in both methodology and software have made spatial modelling read-
42 ily available to the non-specialist (e.g., Wood, 2006; Rue *et al.*, 2009). Here
43 we use the term “spatial model” to include any model that includes spatially
44 referenced covariates, not just smooths of location. This article concerns
45 combining spatial modelling techniques with distance sampling (Buckland
46 *et al.*, 2001, 2004).

Distance sampling takes plot sampling (counting the individuals or groups of objects in a strip or circle) and extends it to the case where detection is not certain. Observers travel along transect centre lines or stand at points and record the distance from the centre line or point to the object of interest (y). These distances are used to estimate the *detection function*, $g(y)$ (bottom left panel, figure 1), by modelling the decrease in detectability with increasing distance from the line or point (conventional distance sampling, CDS). The detection function may also include animal/observer specific covariates (multiple covariate distance sampling, MCDS; Marques *et al.*, 2007). From the fitted detection function, the probability of detection can be calculated.

The estimated probability that an animal is detected, \hat{p}_i , can then be used to calculate abundance as

$$\hat{N} = \frac{A}{a} \sum_{i=1}^n \frac{1}{\hat{p}_i}, \quad (1)$$

where A is the area of the study region, a is the area covered by the survey (i.e., the sum of the areas of all of the strips/circles) and the summation takes place over the n observed individuals (Buckland *et al.*, 2001, Chapter 3). In general distance sampling is more efficient than plot sampling since all objects observed are recorded and only later discard observations deemed to far away (outside of the *truncation distance*).

When fitting the detection function in a distance sampling analysis, one assumes that the objects of interest are distributed according to some process (Buckland *et al.*, 2001, Section 2.1). It is usually possible to design surveys such that a homogenous process can be assumed so, with respect to the line, objects are distributed uniformly. This can be achieved by e.g., ensuring that transect lines run perpendicular to geographical features that would attract or repel animals or by post-stratification (Buckland *et al.*, 2001, Section 3.7).

Estimators such as eqn. 1 are referred to as *design-based* since they rely on the design of the study to ensure inference is valid. This article focusses on *model-based* inference. Using spatially explicit models one can investigate the response of biological populations to biotic and abiotic covariates which vary over the survey area. Modelling the spatial process also enables the use data from badly designed or opportunistic surveys, for example “incidental” data arising from “ecotourism” cruises can be included in analyses (Williams *et al.*, 2006).

68 Our aims in a DSM analysis are usually two-fold: estimating overall
69 abundance and investigating the relationship between abundance and envir-
70 onmental covariates. As with any predictions which are outside of the range
71 of the data, one should heed the usual warnings regarding extrapolation. For
72 example, in an terrestrial study, habitat may cause significant issues if there
73 was not search effort in all habitats. Frequently, maps of abundance or dens-
74 ity are required and any spurious predictions can be visually assessed, as well
75 as by plotting a histogram of the predicted values. A sensible definition of
76 the region of interest avoids prediction outside the range of the data.

77 The article focuses on those recent advances in spatial modelling of dis-
78 tance sampling data which are of most utility to applied ecologists. These
79 new methods are available in the R packages `Distance` and `dsm`, and will soon
80 be available in the popular Windows application Distance (Thomas *et al.*,
81 2010).

82 Throughout this article a motivating data set is used to illustrate the
83 methods. These data are from a combination of several shipboard surveys
84 conducted on pan-tropical spotted dolphins in the Gulf of Mexico. 47 ob-
85 servations of groups of dolphins The group size was recorded, as well as the
86 Beaufort sea state at the time of the observation. Coordinates for each obser-
87 vation and bathymetry data were also available as covariates for the analysis.
88 A complete example analysis is provided as an online appendix.

89 The rest of the article is structured as follows: we first describe the dens-
90 ity surface modelling approach of Hedley & Buckland (2004), explain how
91 to estimate abundance and uncertainty. We then describe recent advances,
92 practical advice regarding the model fitting, formulation and checking. Be-

fore concluding, we look at two recent methods which take a rather more direct approach to modelling spatial distance sampling data.

Density surface modelling

This section focuses on modelling the abundance/density estimation stage of distance sampling, using the “count model” of Hedley & Buckland (2004) which we refer to as *density surface modelling* (DSM). Both line and point transects can be used but if lines are used then they are split into contiguous *segments* (indexed by j), which are of length l_j ; small enough such that the density does not vary appreciably within a segment. The general idea is to model the count or estimated abundance as a smooth function of covariates using a generalized additive model (GAM; Wood, 2006). For each segment or point, the response is modelled as a function of *covariates at the environmental level* (the z_{jk} with k indexing the covariates, e.g., location, sea surface temperature, weather conditions). The covered area enters the model as an offset: the area surveyed at segment j is $A_j = 2wl_j$ and at point j is $A_j = w\pi^2$ (where w is the truncation distance).

COUNT AS RESPONSE

The model for the count per segment is:

$$\mathbb{E}(n_j) = \exp \left[\log_e(p_j A_j) + \beta_0 + \sum_k f_k(z_{jk}) \right],$$

110 where the f_k s are smooth functions of the covariates in the GAM case and
 111 β_0 is an intercept term. Multiplying the covered area (A_j) by the probability
 112 of detection (p_j) gives the *effective area* of the survey at that point. The
 113 distribution of n_j can then be modelled as overdispersed Poisson, negative
 114 binomial, or Tweedie (see *Recent developments*, below) distribution.

115 Figure 1 (top panel) shows the raw observations from the dolphin data,
 116 along with the transect lines, overlaid on the depth data. Figure 2 shows a
 117 GAM fitted to the dolphin data, the top panel shows predictions from a model
 118 where depth was the only covariate, the bottom panel shows predictions
 119 where a (bivariate) smooth of spatial location was also included.

120 Abundance estimation is not the only information contained in these mod-
 121 els. Plots of marginal smooths of the spatially referenced covariates show the
 122 relationships between the covariates and abundance. The effect of depth on
 123 abundance for the dolphin data can be seen in Figure 3. Between 0 and
 124 500m there is a depth effect which then seems to level off (a straight line
 125 could be drawn inside the confidence band). This may indicate that the dol-
 126 phins prefer water deeper than 500m, however the usual caveats inherent in
 127 interpreting results from observational studies apply.

128 ESTIMATED ABUNDANCE AS RESPONSE

An alternative to modelling counts would be to use the per-segment/circle
 abundance can be estimated using distance sampling methods and the es-

timated counts used as the response. In this case we replace n_j by:

$$\hat{N}_j = \sum_{r=1}^{R_j} \frac{s_{jr}}{\hat{p}_j},$$

129 where R_j is the number observations in segment j and s_{jr} is the size of the
 130 r^{th} group in segment j (if the animals occur individually then $s_{jr} = 1, \forall j, r$).
 131 If there are no covariates other than distance in the detection function then
 132 the probability of detection is constant (i.e., $\hat{p}_j = \hat{p}, \forall j$).

The following model is then fitted:

$$\mathbb{E}(\hat{N}_j) = \exp \left[\log_e (A_j) + \beta_0 + \sum_k f_k (z_{jk}) \right],$$

133 where \hat{N}_j , as with n_j , is assumed to follow an overdispersed Poisson, negative
 134 binomial, or Tweedie distribution.

135 *DSM with covariates at the observation level*

136 The above models only consider the case where the covariates are measured
 137 only at the segment/point level. Often covariates (z_{ij} , for individual/group
 138 i , segment/point j) are collected on the level of observations; for example
 139 sex, length or observer identity. In this case the probability of detection is
 140 a function of the individual level covariates $\hat{p}(z_i)$. Individual level covariates
 141 can be incorporated into the model by adopting the following estimator of
 142 the per-segment abundance:

$$\hat{N}_j = \sum_{r=1}^{R_j} \frac{s_{jr}}{\hat{p}(z_{ij})}.$$

143 It is possible that bias is incurred by larger groups and therefore more
144 visible groups. Including group size as a covariate in the detection function
145 and fitting the above model is one solution. The section *Practical advice*,
146 gives another method that can be used to deal with size bias in grouped
147 populations.

148 By not including an offset, but instead dividing the count (or estimated
149 abundance) by the area, we can also model density rather than abundance.
150 We concentrate on abundance here, see Hedley & Buckland (2004) for further
151 details.

152 PREDICTION

153 To calculate an abundance estimate for some region of interest, the necessary
154 covariates (those included in the model) must be available for the whole of
155 that region, and they must also be available at the required resolution (using
156 prediction grid cells that are smaller than the resolution of the spatially
157 referenced data will not have an effect on abundance/density estimates).
158 The areas of the segments/points are included as an offset in the model, so
159 predictions must be multiplied by the area of the prediction cell (to move from
160 density to abundance). Predictions can be made for the particular covariate
161 levels and abundance estimates calculated from summing predicted values
162 over the prediction grid cells.

163 VARIANCE ESTIMATION

164 Estimating the variance of abundances calculated using DSM is not straight
165 forward as uncertainty from the estimated parameters of the detection func-

tion must be incorporated into the spatial model. A second consideration is that in a line transect survey, adjacent segments are likely to be correlated; failing to account for this spatial autocorrelation will lead to artificially low variance estimates and hence misleadingly narrow confidence intervals.

Hedley & Buckland (2004) describe a method of calculating the variance in the abundance estimates using a parametric bootstrap, resampling from the residuals of the fitted model. The bootstrap then follows the following steps.

Denote the fitted values for the model to be $\hat{\boldsymbol{\eta}}$. For $b = 1, \dots, B$ (where B is the number of resamples required).

1. Resample (with replacement) the per-segment residuals, store the values in \mathbf{r}_b .
2. Refit the model but with the response set to $\hat{\boldsymbol{\eta}} + \mathbf{r}_b$ (where $\hat{\boldsymbol{\eta}}$ are the fitted values from the original model).
3. Take the predicted values for the new model and store them.

From the predicted values stored in the last step, the per-location and abundance variance can be calculated in the usual manner. The total variance of the abundance estimate can then be found by combining the variance estimate from the bootstrap procedure with the variance of the probability of detection from the detection function model (using the delta method; Seber, 1982). This assumes that the two components of the variance are independent and the method does not take into account spatial autocorrelation (the individual segments are treated as independent).

189 The above procedure assumes that there is no correlation in space between
190 segments and that residuals can be swapped around. If many animals are
191 observed in a segment then we might expect there to be high numbers in the
192 adjacent segments. A moving block bootstrap (MBB; Efron & Tibshirani,
193 1993, Section 8.6) can account for some of this spatial autocorrelation in the
194 variance estimation. The segments are grouped together into overlapping
195 blocks, (so if the block size is 5, block one is segments 1, . . . , 5, block two is
196 segments 2, . . . , 6, and so on). Then, at step (2) above, resamples are taken of
197 the blocks (i.e. groups of segments together) rather than individual segments
198 within the transects. Using blocks should account for some of the autocor-
199 relation between the segments, inflating the variances accordingly. However,
200 since the block size dictates the maximum amount of spatial autocorrelation
201 accounted for, this may not fully account for the autocorrelation. The boot-
202 strap procedures can also be modified to take into account detection function
203 uncertainty by generating new distances from the fitted detection function
204 and then re-calculating the offset by fitting a detection function to the new
205 distances.

206 Recent developments

207 *Variance propagation*

208 Rather than using a bootstrap, Williams *et al.* (2011) calculate the variance
209 without having to refit the model many times. Their method incorporates
210 the uncertainty in the estimation of the detection function into the variance
211 of the spatial model, albeit only with environmental level covariates. Their

212 procedure is as follows:

- 213 1. Fit a density surface model.
- 214 2. Re-fit the model with an additional random effects term. This term
215 characterises the uncertainty in the estimation of the detection function
216 (via the uncertainty of the probability of detection, \hat{P}_a).
- 217 3. Variance estimates of the abundance calculated using standard GAM
218 theory (Wood, 2006, page 245) from the model will include uncertainty
219 from the estimation of the detection function.

220 We consider propagating the uncertainty in this manner not only to be more
221 computationally efficient but also preferable from a technical perspective.
222 The bootstrap does not fully account for spatial autocorrelation, assuming
223 that the residuals are exchangeable when they are not will lead to wider
224 confidence intervals. In simulation the confidence intervals produced are
225 narrower (than their bootstrap equivalents), while maintaining good cover-
226 age.

227 A common way to visualise uncertainty in a DSM is to plot the per-cell
228 coefficient of variation by dividing the standard error for each cell by its
229 predicted abundance. Figure 4 shows a map of the coefficient of variation
230 for the model which includes both location and depth covariates using the
231 variance propagation method.

232 EDGE EFFECTS

233 Recent work (Ramsay, 2002; Wang & Ranalli, 2007; Wood *et al.*, 2008; Scott-
234 Hayward *et al.*; Miller & Wood) has highlighted the need to take care when

smoothing over areas with complicated boundaries; e.g., rivers, peninsulae or islands. If two parts of the domain (either side of a mountain, say) are inappropriately linked by the model (the distance between the points is measured as a straight line, rather taking into account obstacles) then the boundary feature can be “smoothed across” leading to incorrect inference. Ensuring that a realistic spatial model has been fit to the data is essential for valid inference. The soap film smoother of Wood *et al.* (2008) is particularly appealing as the model jointly estimates boundary conditions for a complex study area along with the “interior” smooth. This can be particularly helpful when uncertainty is estimated via a bootstrap as the model helps avoid large, unrealistic predictions which can plague other smoothers (Bravington & Hedley, 2009).

Even if the study area does not have a complicated boundary, edge effects can still be problematic. Miller *et al.* show that when using global smoothers, smoothing towards the plane can cause the fitted surface to “curl-up” as predictions move further away from the data. They suggest the use of Duchon splines (a generalisation of thin plate regression splines) to alleviate the problem by smoothing toward the intercept.

TWEEDIE DISTRIBUTION

The Tweedie distribution offers a very flexible alternative to the quasi-Poisson distribution as the usual response distribution when modelling count data (Candy, 2004). Through the parameter λ , many common distributions arise; varying λ between 1 (Poisson) and 2 (gamma) leads to a random variable which is a sum of M gamma variables where M is Poisson distributed (Jør-

gensen, 1987). Although it is possible to perform optimization to find λ , this is generally seen as unnecessary as the distribution does not change appreciably when λ is changed by less than 0.1 (therefore trial and error is usually reasonable). Mark Bravington (pers. comm.) suggested plotting the square root of the absolute value of the residuals against fitted values; a “flat” plot (points forming a horizontal line) give an indication of a “good” value for λ . We additionally suggest using the metrics described in the next section for model selection.

Practical advice

Figure 5 shows a flow diagram of the modelling process for creating a density surface model for distance sampling data. The diagram shows which methods are compatible with each other and what the options are for modelling a particular data set.

In our experience, it is sensible to start with a detection function without covariates and a simple smooth of spatial location and then add in more complicated features such as covariates in the detection function, or using a soap film smoother (perhaps afterwards dropping the location term). Model selection can be performed for the detection function using AIC and model checking using goodness-of-fit tests given in Buckland *et al.* (2004). For the spatial model, generalized cross validation (GCV) score (or related metrics such as UBRE or REML) and percentage deviance explained are useful. We also highly recommend the use of standard GAM diagnostic plots (Wood, 2006, Chapter 5).

282 In the dolphin analysis, we include a smooth of location. This not only
283 doubles the percentage deviance explained (27.3% to 52.7%), it also allows
284 us to account for spatial autocorrelation (in a primitive way). One can see
285 this when comparing the two plots in Figure 2 and the plot of the depth in
286 Figure 1, the plot of the smooth of depth alone looks very similar to the raw
287 plot of the depth data. A smooth of an environmental-level covariate such as
288 depth can be very useful for assessing the relationships between abundance
289 and the covariate. Caution should be employed when interpreting smooth
290 relationships and abundance estimates, especially if there is poor coverage of
291 covariate values. For example if there is a large agglomeration of individuals
292 at a high value of depth but no further observations occur at such a high
293 value, then investigators should be skeptical of any relationship. For this
294 reason a smooth of space is recommended for inclusion in candidate models.
295 Limiting the “wigglyness” of smooths of spatial location can be a useful way of
296 restricting their influence whilst still allowing them to “mop up” the residual
297 spatial correlation in the data.

298 In the analysis we have converted from latitude and longitude to kilo-
299 metres from (27.01, -88.3). This is because the bivariate smoother which we
300 use (the thin plate spline; Wood, 2003) is isotropic: it treats the wigglyness
301 of the smoother in each direction as equal: moving 1 degree in latitude is not
302 the same as moving 1 degree in longitude, using kilometres from the centre
303 of the study area is sensible (using SI units removes the need for conversion
304 later).

305 If animals occur in groups rather than individually, bias can be incurred
306 due to larger groups being more visible than smaller groups. Bias due to

group size can be assessed by regressing evaluations of the fitted detection function onto the logarithm of group size, then comparing the expected and observed values of the group size, if there is a large difference then it may be necessary to include size as a covariate in the detection function. The bottom right panel of figure 1 shows a such a plot with the regression line overlaid.

Direct modelling of the spatial point process

Rather than use a GAM to model the spatially explicit part of the model, two recent articles have modelled the process using point processes (Cox & Isham, 1980). In both cases the density of object is governed by a spatially-varying *intensity function*, which can include covariates in a similar manner to the GAM.

Johnson *et al.* (2010) propose a point process-based model for distance sampling data (known as DSpat). They first assume that the locations of all individuals in the survey area (not just those observed) form a realisation of a Poisson process. Parameters of the intensity function are then estimated via standard maximum likelihood methods for point processes (Baddeley & Turner, 2000). In contrast to Hedley & Buckland (2004), all parameters are estimated jointly so uncertainty from both the spatial pattern and the detection function is incorporated into variance estimates for the abundance. This also ensures that correlations between the detection function and underlying point process are estimated correctly (and do not falsely inflate or deflate variance estimates). The authors also address the issue of overdispersion

330 (commonly a symptom of animals or groups clustering), unmodelled by spa-
331 tial covariates using a post-hoc correction factor (see *Recent Developments*,
332 below, for another approach).

333 Niemi & Fernández (2010) also use Poisson processes but incorporate
334 them into a fully Bayesian approach. Unlike Johnson *et al.* (2010) model
335 fitting proceeds in two stages: first the detection function is fitted, then the
336 spatial model (via MCMC) assuming the detection function parameters are
337 known, so detection function uncertainty is not incorporated in the spatial
338 model.

339 Both of the above Poisson process models do not account for group size,
340 both stating that this could be included by considering a marked point pro-
341 cess (Cox & Isham, 1980, Section 5.5). Both methods offer direct modelling
342 of the point process, although with some drawbacks compared to the meth-
343 odology of Hedley & Buckland (2004). It should be noted that the loss of
344 efficiency from using a two-stage approach is not large (Buckland *et al.*, 2004,
345 p. 313); distances contain little information about spatial variation because
346 transects are very thin compared to their lengths and circles are very small
347 compared with study area.

348 Discussion

349 The use of model-based inference for determining abundance and spatial
350 distribution from distance sampling data presents new opportunities in the
351 field of population assessment. Inference from a sample of sightings to a
352 population in a study area does not depend upon a random sample design,

353 and therefore data from "platforms of opportunity" (Williams *et al.*, 2006)
354 can be used to make inference.

355 Unbiased estimates are dependent upon either a) distribution of sampling
356 effort being random throughout the study area (for design-based inference)
357 or b) the model is correct (for model-based inference). It is easier to have
358 confidence in the former than in the latter because our models are always
359 wrong. Nevertheless model-based inference will play an increasing role in
360 population assessment as we attempt to squeeze more information from the
361 data we gather.

362 The field is quickly evolving to allow modelling of more complex data
363 building on the basic ideas of density surface modelling. We expect to see
364 large advances in two areas: temporal inferences and the handling of spa-
365 tial correlation. These should become more mainstream as modern spatio-
366 temporal modelling techniques are adopted. Petersen *et al.* (2011) provided
367 a very basic framework for temporal modelling; their model included smooth
368 terms both before and after the construction of an offshore windfarm. Spatial
369 autocorrelation can be accounted for via approaches that explicitly introduce
370 correlations such as generalized estimating equations (GEEs; Hardin & Hilbe,
371 2003) or via mechanisms such as that of Skaug (2006), which allowed observa-
372 tions to cluster according to one of several states (e.g. "feeding" or "transit")
373 taking into account short-term agglomerations ("hot spots").

374 Acknowledgments

375 DLM wishes to thank Mark Bravington and Sharon Hedley for their help
376 and patience in explaining and providing code for their variance propagation
377 method.

378 References

- 379 Baddeley, A. & Turner, R. (2000) Practical maximum pseudolikelihood for spatial
380 point patterns. *Australian & New Zealand Journal of Statistics*, **42**, 283–322.
- 381 Bravington, M. & Hedley, S.L. (2009) Antarctic minke whale abundance estimates
382 from the second and third circumpolar IDCR/SOWER surveys using the
383 SPLINTR model.
384 URL [http://www.iwcoffice.org/_documents/sci_com/sc61docs/](http://www.iwcoffice.org/_documents/sci_com/sc61docs/SC-61-IA14.pdf)
385 SC-61-IA14.pdf
- 386 Buckland, S.T., Anderson, D., Burnham, K.P., Laake, J.L., Borchers, D.L. &
387 Thomas, L. (2001) *Introduction to Distance Sampling*. Oxford University Press.
- 388 Buckland, S.T., Anderson, D., Burnham, K.P., Laake, J.L., Borchers, D.L. &
389 Thomas, L. (2004) *Advanced Distance Sampling*. Oxford University Press.
- 390 Candy, S. (2004) Modelling catch and effort data using generalised linear models,
391 the Tweedie distribution, random vessel effects and random stratum-by-year
392 effects. *CCAMLR Science*, **11**, 59–80.
- 393 Cox, D.R. & Isham, V. (1980) *Point Processes*. Monographs on Applied Probability
394 and Statistics. Chapman and Hall. ISBN 9780412219108.
- 395 Efron, B. & Tibshirani, R.J. (1993) *An Introduction to the Bootstrap*. Chapman &
396 Hall/CRC. ISBN 9780412042317.
- 397 Hardin, J. & Hilbe, J. (2003) *Generalized Estimating Equations*. Chapman and
398 Hall/CRC, London, UK.
- 399 Hedley, S.L. & Buckland, S.T. (2004) Spatial models for line transect sampling.
400 *Journal of Agricultural, Biological, and Environmental Statistics*, **9**, 181–199.
- 401 Johnson, D.S., Laake, J.L. & Ver Hoef, J.M. (2010) A model-based approach for
402 making ecological inference from distance sampling data. *Biometrics*, **66**, 310–
403 318.
- 404 Jørgensen, B. (1987) Exponential dispersion models. *Journal of the Royal Statist-*
405 *ical Society. Series B, Statistical Methodology*, **49**, 127–162.
- 406 Marques, T.A., Thomas, L., Fancy, S. & Buckland, S.T. (2007) Improving estimates
407 of bird density using multiple-covariate distance sampling. *The Auk*, **124**, 1229–
408 1243.
- 409 Miller, D.L., Jones, E. & Matthiopoulos, J. (????) Reliable spatial smoothing
410 without edge effects. pp. 1–8.

- 411 Miller, D.L. & Wood, S.N. (????) Finite area smoothing with generalized distance
412 splines. pp. 1–27.
- 413 Niemi, A. & Fernández, C. (2010) Bayesian Spatial Point Process Modeling of Line
414 Transect Data. *Journal of Agricultural, Biological, and Environmental Statistics*,
415 **15**, 327–345.
- 416 Petersen, I.K., MacKenzie, M.L., Rexstad, E.A., Wisz, M.S. & Fox, A.D. (2011)
417 Comparing pre- and post-construction distributions of long-tailed ducks *Clan-*
418 *gula hyemalis* in and around the Nysted offshore wind farm, Denmark: a quasi-
419 designed experiment accounting for imperfect detection, local surface features
420 and autocorrelation. 2011-1.
- 421 Ramsay, T. (2002) Spline smoothing over difficult regions. *Journal of the Royal*
422 *Statistical Society. Series B, Statistical Methodology*, **64**, 307–319.
- 423 Rue, H., Martino, S. & Chopin, N. (2009) Approximate Bayesian inference for
424 latent Gaussian models by using integrated nested Laplace approximations. *J.*
425 *R. Statist. Soc. B*, **71**, 319–392.
- 426 Scott-Hayward, L.A.S., MacKenzie, M.L., Donovan, C.R., Walker, C.G. & Ashe,
427 E. (????) Complex Region Spatial Smoother (CReSS). pp. 1–31.
428 URL <http://research-repository.st-andrews.ac.uk/handle/10023/2048>
- 429 Seber, G.A.F. (1982) *The Estimation of Animal Abundance and Related Paramet-*
430 *ers*. Blackburn Pr. ISBN 9781930665552.
- 431 Skaug, H.J. (2006) Markov modulated Poisson processes for clustered line transect
432 data. *Environmental and Ecological Statistics*, **13**, 199–211.
- 433 Thomas, L., Buckland, S.T., Rexstad, E.A., Laake, J.L., Strindberg, S., Hedley,
434 S.L., Bishop, J.R., Marques, T.A. & Burnham, K.P. (2010) Distance software:
435 design and analysis of distance sampling surveys for estimating population size.
436 *Journal of Applied Ecology*, **47**, 5–14.
- 437 Wang, H. & Ranalli, M. (2007) Low-rank smoothing splines on complicated do-
438 mains. *Biometrics*, **63**, 209–217.
- 439 Williams, R., Hedley, S.L., Branch, T.A., Bravington, M.V., Zerbini, A.N. & Find-
440 lay, K.P. (2011) Chilean blue whales as a case study to illustrate methods to
441 estimate abundance and evaluate conservation status of rare species. *Conserva-*
442 *tion Biology*, **25**, 526–535.
- 443 Williams, R., Hedley, S.L. & Hammond, P. (2006) Modeling distribution and
444 abundance of Antarctic baleen whales using ships of opportunity. *Ecology and*
445 *Society*, **11**, 1.

- 446 Wood, S.N. (2003) Thin plate regression splines. *Journal of the Royal Statistical*
447 *Society. Series B, Statistical Methodology*, **65**, 95–114.
- 448 Wood, S.N. (2006) *Generalized Additive Models: An introduction with R*. Chapman
449 & Hall/CRC.
- 450 Wood, S.N., Bravington, M.V. & Hedley, S.L. (2008) Soap film smoothing. *Journal*
451 *of the Royal Statistical Society. Series B, Statistical Methodology*, **70**, 931–955.

Fig. 1 Top: the survey area, transect centrelines and observations with size of circle corresponding to the group size overlaid onto depth data; bottom left, histogram of observed distances with fitted detection function; bottom right, plot of evaluations of the fitted detection function at given distances versus the logarithm of group size with linear trend showing the relation between probability of detection (given distance) and group size.

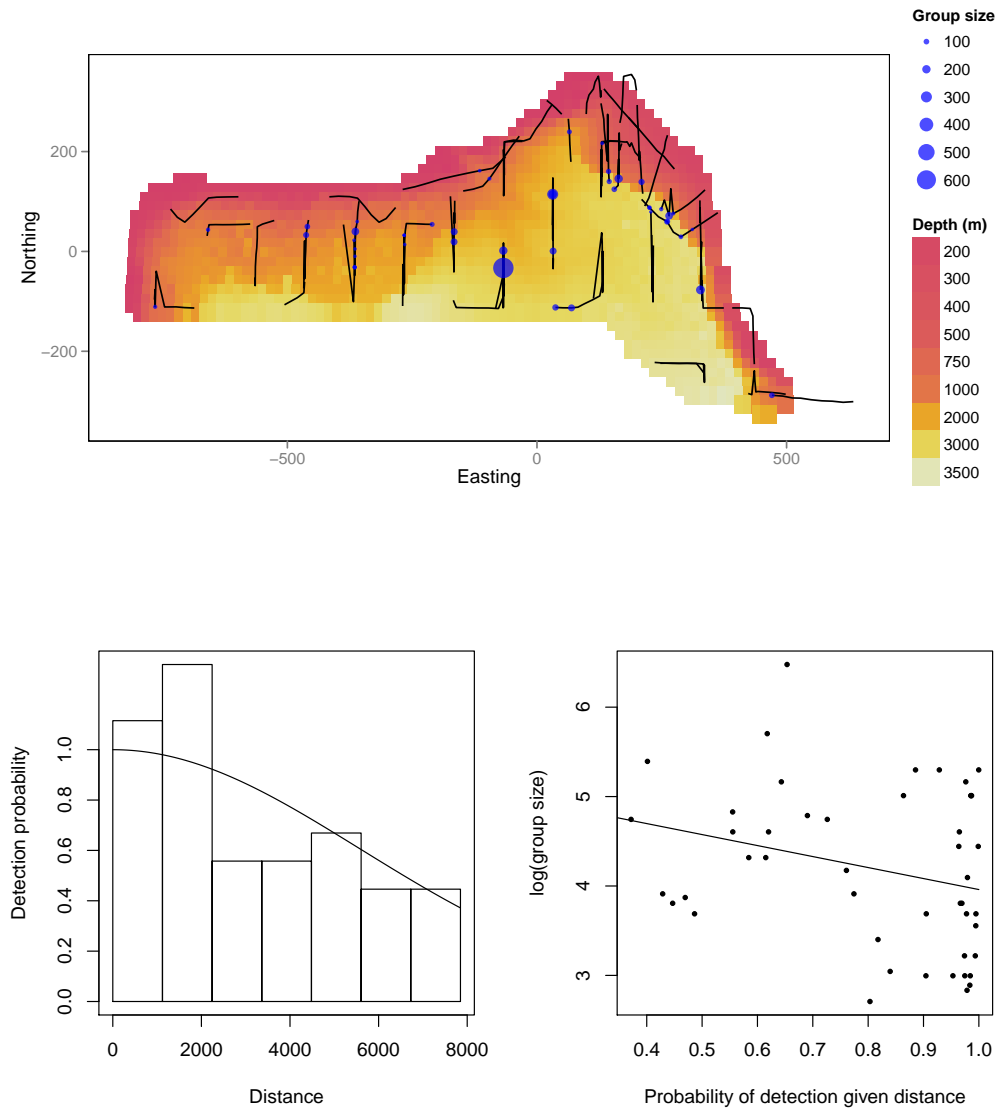


Fig. 2 Predictions for the dolphin data. Top: Predictions from the model using only depth as an explanatory variable, bottom: the model using both depth and location.

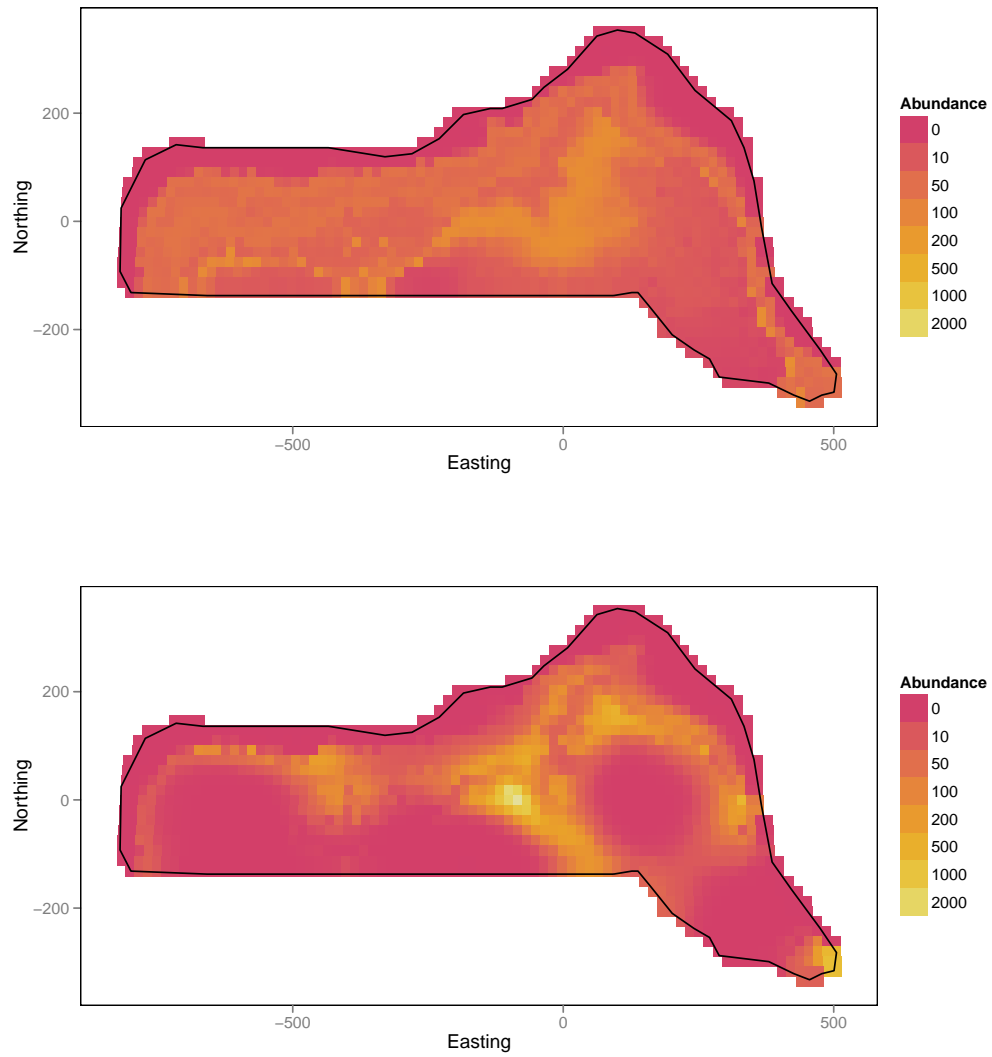


Fig. 3 Plot of the effect on the response of depth, note that it is possible to draw a straight line between 750m and 3000m within the confidence band (between the dashed lines), so the wiggles in the smooth may not be indicative of any relationship. What is clear is that there is some effect up to about 500m. The number in brackets on the y axis indicates the effective degrees of freedom of the smooth term. The rug ticks at the bottom of the plot indicate we have good coverage of the range of depth values in the survey area. Note that the y axis in such plots is on the scale of the link function (log in this case), so care should be taken in their interpretation.

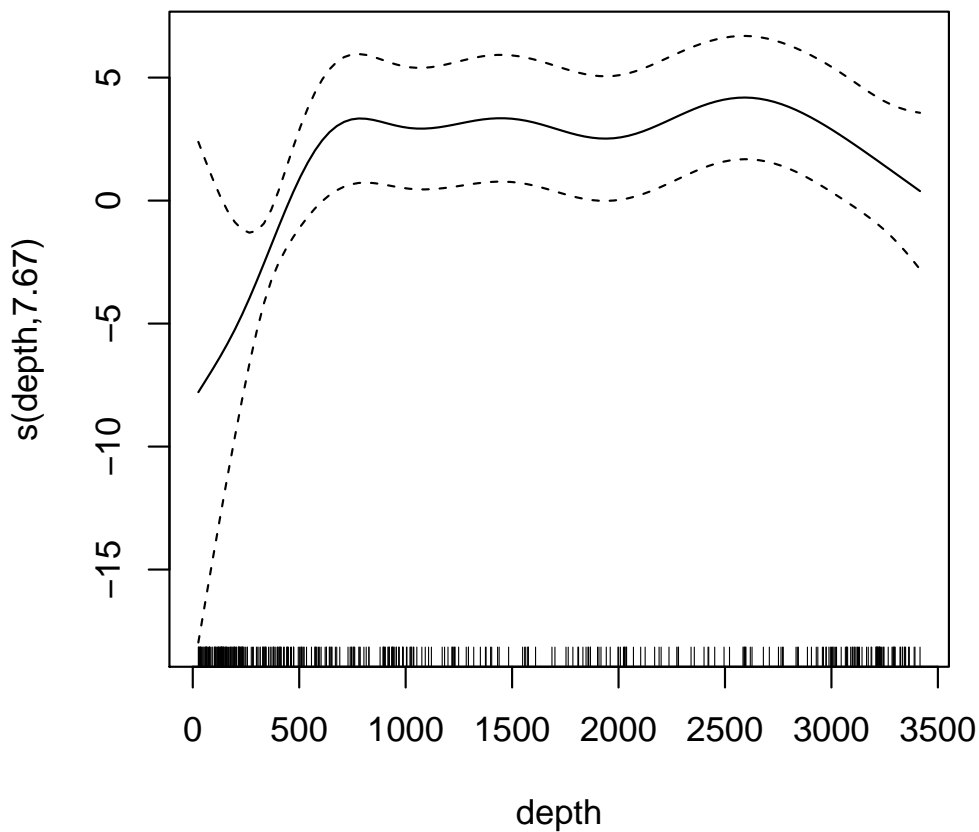


Fig. 4 Plot of coefficient of variation map for the model with smooths of both depth and location. Uncertainty was estimated using the variance propagation method of Williams *et al.* (2011). Overlaid are transect lines and observations.

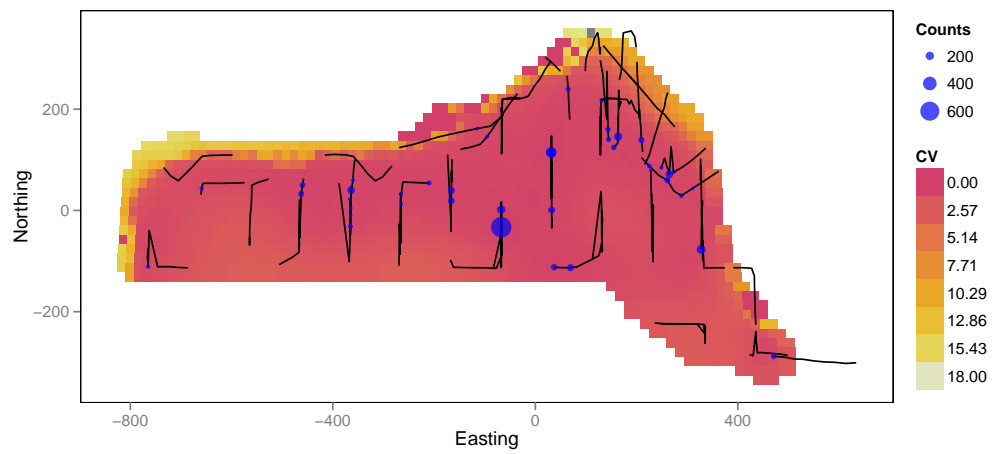


Fig. 5 Flow diagram showing the modelling process for creating a density surface model.

