Density Surface Modeling

WMMC Distance Workshop
7 December 2019

Jason J. Roberts¹
David L. Miller²

¹ Marine Geospatial Ecology Lab, Duke University
² Centre for Research into Ecological and Environmental Modelling, University of St. Andrews

Photo: NOAA/NEFSC, Permit #17355
Münster (1552)
Meerwunder und seltzame Thier

“Here be dragons!”
Olaus Magnus
(1539) *Carta Marina*
Maury (1851) Whale Chart

Lieut. Matthew Fontaine Maury, U.S. Navy
Maury (1851) Whale Chart

Multiple species:
- Sperm Whales
- Right Whales
- Straggling Sperm
- Right

Relative density:
Two Whales of the same species in a square denote that square to be much frequented by that species.

Seasonality:
The best fishing season in each square denoted by the letters:
- W = Winter
- V = Spring
- S = Summer
- A = Autumn
- All = All months

Map showing whale distribution with annotations.
Ready to shoot the Whale—Harpoon shown in position.
THE COST OF NOISE
A right whale calling to another faces the twin challenges—intensity and frequency—that noise poses to many marine animals.

INJURY
Intense noises, such as air gun blasts that ricochet off the seafloor, drown out animal sounds and may cause hearing loss and other damage.

INTERFERENCE
Sounds close in frequency interfere, canceling each other. A ship's propeller miles away can mask a right whale's call.
The North Atlantic Right Whale Major Habitat Areas

- Canadian Critical Habitat Areas
- United States Critical Habitat Areas
- Primary right whale sighting locations
- Occasional right whale sighting locations

Gulf of St. Lawrence
Gulf of Maine

Pershing et al. (2015)
North Atlantic Right Whale Abundance (Pace, 2019)

Nest(2018) = 412
We need some maps:

• Where are they?
• How many are there?
• How do they move seasonally?
• Where will they go in the future?
Density surface modeling

Stage 1

Line transects → Detectability → Spatial model (GAM) → Model checking/criticism → Predictions

(or point transects)

Stage 2

Spatial covariates

Adapted from figure by David L. Miller
The seminal paper (2004)

Spatial Models for Line Transect Sampling

Sharon L. Hedley and Stephen T. Buckland

This article develops methods for fitting spatial models to line transect data. These allow animal density to be related to topographical, environmental, habitat, and other spatial variables, helping wildlife managers to identify the factors that affect abundance. They also enable estimation of abundance for any subarea of interest within the surveyed region, and potentially yield estimates of abundance from sightings surveys for which the survey design could not be randomized, such as surveys conducted from platforms of opportunity. The methods are illustrated through analyses of data from a shipboard sightings survey of minke whales in the Antarctic.

Journal of Agricultural, Biological, and Environmental Statistics, Volume 9, Number 2, Pages 181–199
DOI: 10.1198/1085711043578
Spatial models for distance sampling data: recent developments and future directions

David L. Miller¹*, M. Louise Burt², Eric A. Rexstad² and Len Thomas²

¹Department of Natural Resources Science, University of Rhode Island, Kingston, RI 02881, USA; and ²Centre for Research into Ecological and Environmental Modelling, The Observatory, University of St Andrews, St Andrews KY16 9LZ, UK
Stage 2 of traditional DSM line transect analysis

1. Split transects into segments

2. Model count per segment with spatial covariates

\[ \mathbb{E}(n_j) = \hat{p}_j A_j \exp \left[ \beta_0 + \sum_k f_k(z_{jk}) \right] \]

\[ A_j = 2wl_j \]

Fig. 4.4. Illustration of notation for the count model, using three hypothetical segments of lengths \( l_1, l_2, \) and \( l_3 \).

\( j \) : segment index

\( k \) : spatial model covariate index

\( n_j \) : count of individuals for segment \( j \)

\( \hat{p}_j \) : probability of detection for segment \( j \)

\( A_j \) : area for segment \( j \)

\( f_k \) : smooth of covariate \( k \) for values \( z_{jk} \)
Model the effects of:
- Distance to sighting
- Group size
- Sea state / glare / etc.
on probability of sighting

Availability bias due to diving

Survey area
Histogram of distances

Detection function fitting
- MCDS
- CDS
Fitted detection function
Calculate response and offset

Spatial model fitting
GAM
Spatially referenced covariates
Choose response distribution
Fit model
Model checking
"Fitted" model

Prediction
Predict
Abundance map
Abundance estimate

Uncertainty estimation
- Moving block bootstrap
- Variance estimate
- Map of coefficient of variation

Line transect surveys

Adapted from Miller et al. (2013)

Figure: Hooker et al. (2011)
Line transect surveys

Perception bias

Correct for $g(0) < 1$

Model the effects of:
- Distance to sighting
- Group size
- Sea state / glare / etc.
on probability of sighting

DSM workflow
Adapted from Miller et al. (2013)
Typical approach to bias correction:

1. Estimate availability bias correction factor from diving data
2. Estimate perception bias correction factor with an independent observer team, or similar methods
3. Multiply correction factors to get $g(0)$
4. Multiply $\hat{p}_j$ by $g(0)$ in second stage:

$$\mathbb{E}(n_j) = \hat{p}_j g(0) A_j \exp \left( \beta_0 + \sum_k f_k(z_{jk}) \right)$$
Line transect surveys

Correct for $g(0) < 1$

Split into segments

Can also split into grid cells

Model the effects of:
- Distance to sighting
- Group size
- Sea state / glare / etc.

on probability of sighting

DSM workflow
Adapted from Miller et al. (2013)
Line transect surveys

Sample environmental covariates at segments
- Physiographic (e.g. depth)
- Physical (e.g. SST)
- Biological (e.g. prey)

Study area extent

Grids

DSM workflow

Adapted from Miller et al. (2013)
Recent research trends

• Integrating multiple surveys to facilitate models with greater spatial, temporal, and taxonomic coverage

• Extrapolating models across unsurveyed areas, times, and the future

• Incorporating multiple sources of uncertainty into density surface predictions
Recent research trends

• Integrating multiple surveys to facilitate models with greater spatial, temporal, and taxonomic coverage

• Extrapolating models across unsurveyed areas, times, and the future

• Incorporating multiple sources of uncertainty into density surface predictions
Common idea: standardize surveys via detection modeling

Stage 1

Survey 1
Line transects ➔ Detectability

Survey 2
Line transects ➔ Detectability

Survey 3
Line transects ➔ Detectability

Stage 2

Spatial model (GAM) ➔ Model checking/criticism ➔ Predictions

Spatial covariates
Habitat-based cetacean density models for the U.S. Atlantic and Gulf of Mexico

Roberts et al. (2016)

Detection hierarchies

Bottlenose dolphin: 5630 sightings, 17 functions

Examples: aerial surveys

Risso’s dolphin: 368 sightings, 7 detection functions
Taxonomic coverage
- 26 single-species models
- 3 guild models

Seasonal coverage
- 11 taxa predicted at monthly resolution
- 18 at year-round resolution
Combining multiple visual surveys to model the habitat of deep-diving cetaceans at the basin scale

Large-scale modelling of deep-diving cetacean habitats
Hierarchical Bayesian model for hazard rate detection:

\[ \text{ESW}_{jks} = \int_0^w g_s(x) \, dx = \int_0^w \left[ 1 - \exp\left( -\left( \frac{x}{e^{\beta_0 + \beta_1 k + \alpha_s}} \right)^{-v_s} \right) \right] \, dx \]

\( s = \text{survey} \)
\( j = \text{platform height} \)
\( k = \text{sea-state} \)

Effects of platform height  
Survey random effects

Virgili et al. (2019)
Recent research trends

• Integrating multiple surveys to facilitate models with greater spatial, temporal, and taxonomic coverage

• Extrapolating models across unsurveyed areas, times, and the future

• Incorporating multiple sources of uncertainty into density surface predictions
Recent research trends

• Integrating multiple surveys to facilitate models with greater spatial, temporal, and taxonomic coverage

• Extrapolating models across unsurveyed areas, times, and the future

• Incorporating multiple sources of uncertainty into density surface predictions
Extrapolating cetacean densities to quantitatively assess human impacts on populations in the high seas

Laura Mannocci, 1* Jason J. Roberts, 1 David L. Miller, 2,3,4 and Patrick N. Halpin 1

1Marine Geospatial Ecology Laboratory, Nicholas School of the Environment, Duke University, Durham, NC 27708, U.S.A.
2Integrated Geosciences, 16 Summer Street, Woods Hole, MA 02543, U.S.A.
3Centre for Research into Ecological and Environmental Modelling and School of Mathematics and Statistics, University of St Andrews, The Observatory, Buchanan Gardens, St Andrews, Fife KY16 9LF, Scotland
4Woods Hole Oceanographic Institution, Woods Hole, MA 02543, U.S.A.

Figure 4. Maps of (a) mean year-round predicted densities (individuals per 100 km²) and (b) mean year-round coefficients of variation from the spatial part of the density model for striped dolphin (black horizontal lines, extrapolations beyond the predictor ranges [these predicted densities should be considered with extreme caution]). An Albers equal area projection is used for the study area.
Combining multiple visual surveys to model the habitat of deep-diving cetaceans at the basin scale

Large-scale modelling of deep-diving cetacean habitats

Also modeled sperm whales and *Kogia* (results not shown).
Virgili et al. (2019)

(a) - Beaked whale model ($D^* = 33.1\%$)

(b) - Predictions with simple interpolation (interpolation 94%)

(c) - Predictions with precautionary interpolation (interpolation 53%)
Predicting cetacean abundance and distribution in a changing climate

Example results for 2 species (8 modeled in total):
 Extrapolation tools developed by Phil Bouchet et al.

- **dsmextra v 1.0.0**
  Extrapolation diagnosis in R

- **New R package** available freely on GitHub
- **Functions for** detecting, summarising and visualising extrapolation (ExDet + %N)
- **Designed to easily integrate** with outputs generated by the dsm package
- **Step-by-step vignette** available, with practical examples on sperm/beaked whale data
- **More features to come ...**

---

**Extrapolation tools for density surface models**

*dsmextra* provides a toolkit for quantifying and visualising extrapolation in density surface models (as implemented in package *dsm*) projected into novel environmental space. Currently, *dsmextra* defines extrapolation on the basis of two metrics: (1) ExDet (Mesgaran et al. 2014), and (2) %N (percentage of data nearby, Mannocci et al. 2018).

*dsmextra* offers a variety of numerical and graphical outputs, including summary plots and interactive maps created as *ggplot2* and *html* objects, respectively. Additional functionality (e.g. assessment methods for dynamic covariates) will be added in future releases.

The idea behind *dsmextra* is to aid ecologists, practitioners, and model end-users in identifying conditions (e.g. areas) under which predicted density surfaces may be prone to errors. In so doing, *dsmextra* may support:

- Better-informed interpretations of density surface model outputs and their associated uncertainties.
- Improvements to model development and selection protocols.
- Cost-effective allocation of future survey effort towards priority, data-poor areas.

**Getting started**

If you are just getting started with *dsmextra*, we recommend reading the tutorial vignette, which provides a quick introduction to the package. You may also find the following technical report useful:

Bouchet et al. (2019). From here and now to there and then: Practical recommendations for extrapolating cetacean density surface models to novel conditions. CREEM technical report 2019-01, Centre for Research into Ecological & Environmental Modelling (CREEM), University of St Andrews, 59 p.

https://densitymodelling.github.io/dsmextra/
Recent research trends

• Integrating multiple surveys to facilitate models with greater spatial, temporal, and taxonomic coverage

• Extrapolating models across unsurveyed areas, times, and the future

• Incorporating multiple sources of uncertainty into density surface predictions
Recent research trends

• Integrating multiple surveys to facilitate models with greater spatial, temporal, and taxonomic coverage

• Extrapolating models across unsurveyed areas, times, and the future

• Incorporating multiple sources of uncertainty into density surface predictions
Variance in density surface models

DSM equation

\[ n_i = \exp (\log_e A_i \hat{p}_i + \beta_0 + f_{\text{space}}(\text{lat}_i, \text{lon}_i) + f_{\text{depth}}(\text{Depth}_i)) \]

Multiple sources of variance

\[ n_i = \exp (\log_e A_i \hat{p}_i + \beta_0 + f_{\text{space}}(\text{lat}_i, \text{lon}_i) + f_{\text{depth}}(\text{Depth}_i)) \]

\(V_\theta\) covariance from distance sampling, \(V_\beta\) covariance from GLM/GAM theory

Material from David L. Miller
How to combine variances?

1. Use the “delta method” (Buckland et al. 2001, p. 52)

\[
\left\{ \text{CV} \left( \hat{N} \right) \right\}^2 = \left\{ \text{CV} \left( \hat{N}_{GAM} \right) \right\}^2 + \left\{ \text{CV} \left( \hat{N}_{DS} \right) \right\}^2
\]

Drawbacks:

• Assumes independence: does not account for covariance between the GAM and the detection function
• There can be only 1 detection function
How to combine variances?

2. Use a moving block bootstrap

Drawbacks:

• Ignores prior information about the GAM smooths

• Assumes covariates have been sampled throughout their ranges, with no gaps
How to combine variances?

3. Propagate variances (Bravington, Miller, and Hedley; in review)

Going back to the DSM formula (on the log scale) and “adding” a random effect, then refitting the model

$$\log e n_i = \log e A_i \hat{p}_i + \beta_0 + f_{\text{space}}(\text{lat}_i, \text{lon}_i) + f_{\text{depth}}(\text{Depth}_i) + \kappa_i \delta.$$  

In the refitted model, we have adjusted the detectability and have a combined variance-covariance matrix:

$$
\begin{pmatrix}
V_{\hat{\beta}} & \text{covariances} \\
\text{Detection function covariance matrix} & V_{\theta}
\end{pmatrix}
$$

Random effect adds the missing covariances but does not add to the mean.
Density Surface Modeling Resources

• Reference books

Buckland et al. (2001)
Buckland et al. (2004)
Buckland et al. (2015)
Density Surface Modeling Resources

• Reference books

• https://distancesampling.org/
  • Free software: program Distance and R packages
  • Upcoming and archived workshops
  • Free online course (26 lectures, 12 exercises)
  • Mailing list and searchable archive (Google Groups)
  • Bibliography (1500 references)

• FAQ for marine mammal density surface modelers:
  https://osf.io/5eza8/wiki/
Questions Frequently Asked by Marine Mammal Density Surface Modelers

In 2018, the U.S. Navy Living Marine Resources (LMR) program initiated the “DenMod” collaboration between the Centre for Research into Ecological and Environmental Modelling (CREEM) at the University of St. Andrews, the Marine Geospatial Ecology Laboratory at Duke University, and the NOAA Fisheries Science Centers— to develop best practices for density surface modeling of marine species and provide practical guidance to the modeling community, particularly those involved in modeling marine mammals for the Navy’s training and testing areas. Pursuant to this goal, we developed this collection of answers to technical modeling questions frequently asked by ourselves and external colleagues. Through this corpus, we offer ready solutions to common modeling problems, distilled from our experiences developing distance sampling and density surface modeling methodology, and applying it in the development of marine mammal density models for the Navy.

How the questions are organized

The questions are organized into wiki pages, with one or more related questions occurring in sequence on each page. The pages are grouped into several “components”. You can drill into the navigation hierarchy on the left or click the component links below:

- DSM Workflows
- Surveys and Survey Data
- Detection Models
- Spatial Models

Scope of the questions

Most of the practitioners within our collaboration have focused on modeling the absolute abundance (number of individuals) and density (individuals per unit area) of marine mammals, usually over large expanses (100s-1000s km) of marine waters of the United States, typically by applying methods derived from distance sampling to visual line transect surveys conducted from ships and aircraft. The questions that appear here reflect this focus.

Our focus

Models that estimate absolute density
Models that use line transect
Models that use methods derived from distance sampling
Spatial Models

This wiki contains questions related to the second stage of traditional density surface modeling, which is the spatial model. Typically, this is a generalized additive model that relates abundance observed on segments of a line transect survey to covariates, which often include an explicit spatial term (a 2D smooth of x and y coordinates) and environmental variables believed to relate to the habitat of the modeled taxon. The kinds of questions that appear here reflect this approach.

The questions and their answers are grouped by topic. Select a topic below or from the alphabetical list in the left navigation bar.

- Splitting transects into segments
- Covariates: which should I use?
- Covariates: steps and tools for acquiring
- Covariates: transformations and outliers
- mgcv: distributions
- mgcv: model selection
- mgcv: soap film smoother
- Predictions: summarizing
- Predictions: validating
- Unidentified and ambiguous sightings
Thank you!

DSM FAQ: https://osf.io/5eza8/wiki/

Ziphius cavirostris

Marine Mammal Program, UNCW