The Annals of Applied Statistics
2017, Vol. 11, No. 4, 2270–2297
https://doi.org/10.1214/17-AOAS1078
© Institute of Mathematical Statistics, 2017

POINT PROCESS MODELS FOR SPATIO-TEMPORAL DISTANCE SAMPLING DATA FROM A LARGE-SCALE SURVEY OF BLUE WHALES[1]

BY YUAN YUAN*, FABIAN E. BACHL†, FINN LINDGREN†, DAVID L. BORCHERS*, JANINE B. ILLIAN*,‡, STEPHEN T. BUCKLAND*, HÅVARD RUE§ AND TIM GERRODETTE¶

University of St Andrews, University of Edinburgh†, Norwegian University of Science and Technology‡, King Abdullah University of Science and Technology§ and NOAA National Marine Fisheries Service¶*

Distance sampling is a widely used method for estimating wildlife population abundance. The fact that conventional distance sampling methods are partly design-based constrains the spatial resolution at which animal density can be estimated using these methods. Estimates are usually obtained at survey stratum level. For an endangered species such as the blue whale, it is desirable to estimate density and abundance at a finer spatial scale than stratum. Temporal variation in the spatial structure is also important. We formulate the process generating distance sampling data as a thinned spatial point process and propose model-based inference using a spatial log-Gaussian Cox process. The method adopts a flexible stochastic partial differential equation (SPDE) approach to model spatial structure in density that is not accounted for by explanatory variables, and integrated nested Laplace approximation (INLA) for Bayesian inference. It allows simultaneous fitting of detection and density models and permits prediction of density at an arbitrarily fine scale. We estimate blue whale density in the Eastern Tropical Pacific Ocean from thirteen shipboard surveys conducted over 22 years. We find that higher blue whale density is associated with colder sea surface temperatures in space, and although there is some positive association between density and mean annual temperature, our estimates are consistent with no trend in density across years. Our analysis also indicates that there is substantial spatially structured variation in density that is not explained by available covariates.

1. Introduction. Distance sampling is a widely used set of survey methods for estimating animal density or abundance [Buckland et al. (2001, 2015)]. Conventional distance sampling methods (of which line transect and point transect methods are the most common) use a combination of model-based inference for estimating detection probability and design-based inference with Horvitz–Thompson-like estimators [Borchers et al. (1998)] for estimating density and

Received September 2016; revised June 2017.

[1]Supported by the Engineering and Physical Sciences Research Council (EPSRC)—EP/K041061/1 and EP/K041053/1.

Key words and phrases. Distance sampling, spatio-temporal modeling, stochastic partial differential equations, INLA, spatial point process.

abundance conditional on the detection probability estimates. While the design-based nature of the second stage in this two-stage estimation process [Buckland, Oedekoven and Borchers (2016)] confers robustness on density and abundance estimates when suitable designs are used, it severely restricts the spatial resolution at which such estimates can be obtained. This is because design-based inference requires adequate sampling units (strips for line transect surveys and circular plots for point transect surveys) in each area for which animal density or abundance is to be estimated. The low spatial resolution of estimates from this two-stage approach limits the utility of estimates obtained from conventional distance sampling methods as there is often interest in the distribution at high spatial resolution. As a result, there has been increasing interest in distance sampling methods that generate continuous spatial density surface estimates, and hence allow inference at an arbitrarily fine spatial scale.

In this paper, we consider a series of line transect surveys of blue whales (*Balaenoptera musculus*) in the Eastern Tropical Pacific Ocean [ETP, Gerrodette and Forcada (2005)], in which the focus of inference is on how density changes continuously in space, with respect to available explanatory variables, and across years. The surveys were designed for dolphins, not blue whales, so there are relatively few blue whale sightings. A continuous spatial model has the potential to borrow strength from data outside the lightly-sampled strata to improve overall inference.

One can obtain a continuous density model by using a spatial model of density in the second stage, rather than basing inference on the design in this stage. This is usually done by transforming the data to counts: discretizing the sampled strips into smaller spatial units in the case of line transects and specifying a model for the counts within each unit, using estimated detection probability as an offset to correct the counts for detectability. Hedley, Buckland and Borchers (1993) and Hedley and Buckland (2004) pioneered this approach and Niemi and Fernández (2010) developed a similar approach (but ignoring detection uncertainty). The R-package `dsm` [Miller et al. (2014)] implements the approach of Hedley, Buckland and Borchers (1993) and Hedley and Buckland (2004) using generalized additive models [Wood (2006)] to estimate a density surface from the count data. Either frequentist or Bayesian approaches can be used for the second stage [Oedekoven, Laake and Skaug (2015), Oedekoven et al. (2013)], and bootstrapping is often used to propagate the uncertainty of detectability estimated from the first stage. Williams et al. (2011) use a more direct approach to incorporate uncertainty of detectability: a random effect term is added in the second stage to characterize the uncertainty in the estimation of the detection function from the first stage.

One can also estimate the parameters of the detection function and the count model simultaneously [Conn, Laake and Johnson (2012), Johnson, Laake and Ver Hoef (2010), Moore and Barlow (2011), Oedekoven et al. (2014), Pardo et al. (2015), Royle, Dawson and Bates (2004), Royle and Dorazio (2008), Schmidt et al. (2012)]. This is known as a full-likelihood approach, as it involves specifying a

likelihood that incorporates both a detection function model and a spatial density model, allowing simultaneous estimation of both models.

Whether inference is in two stages or one, models that discretize searched strips or lines involve an element of subjectivity in choosing the size of the discrete units and a loss of spatial information because each discrete unit can have only one value of any spatial covariate attached to it, even though it might span an area incorporating a range of covariate values. In this paper, we develop a method that does not suffer from these problems, using a point process model.

Point process theory provides a flexible modeling framework for incorporating the underlying spatial or spatio-temporal stochastic processes and does not require discretization of spatial sampling units. Point process models have been used with ecological data to estimate smooth spatial density surfaces and are an obvious choice for the spatial model component of a full likelihood line transect model, although to date they have mainly been used in ecological applications with fully mapped point patterns [Wiegand and Moloney (2014)]: Stoyan (1982) formulated line transect data as observations of stationary point processes; Hedley, Buckland and Borchers (1993) considered point process models for point transect surveys, and Högmander (1991, 1995) constructed a marked point process model for line transect data with detection probability of an animal treated as a mark, but they used a detection model (in which each animal has a detection circle with variable radius) that was shown by Hayes and Buckland (1983) to be unrealistic and often resulting in biased inference.

Here, we develop a full likelihood point process model for line transect data, in which the detection process thins the underlying point process, and in which the detection model and the point process model are estimated simultaneously. In the context of the blue whale survey, an unknown point process governs the number and locations of the whales in space, and points are thinned (whales missed) with a probability that depends in an unknown way on distance from the known locations of lines. Such an approach is not new for modeling distance sampling data. The R-package `DSpat` [Johnson, Laake and Ver Hoef (2014, 2010)] uses a thinned point process model for line transect survey data. However, their method assumes the absence of residual spatial structure on the intensity level (whale density in our case), which is usually not the case in practice, and may result in biased estimates. We relax the independence assumption by using the stochastic partial differential equation approach [SPDE, Lindgren, Rue and Lindström (2011)] to incorporate a spatial or spatio-temporal random field for the underlying stochastic process of autocorrelated spatial or spatio-temporal random effects. For point process data in general, the SPDE approach avoids the need to aggregate observations [Simpson et al. (2016)], and it provides a flexible modeling framework for spatio-temporal random fields. We build our models in a Bayesian framework, which gives us a tool for fitting complicated models, and the advantage of being able to use integrated nested Laplace approximation [INLA, Rue, Martino and Chopin (2009)] for inference. INLA is a computationally efficient method for Bayesian inference using

numerical approximations instead of a sampling-based method such as Markov chain Monte Carlo algorithms. In addition, our modeling framework accommodates the models of the sort used by Johnson, Laake and Ver Hoef (2010) as a special case, as well as the second stage of the two-stage approach of Miller et al. (2013).

After describing the blue whale survey data in Section 2, we describe our model and computational methods in Sections 3 and Section 4. We then analyze the survey data in Section 5, investigating the underlying spatial stochastic process of blue whale density in this area, and how the blue whales respond to sea surface temperature in space and time. Finally, in Section 6, we discuss the results of the analysis, the utility of our modeling approach and extensions for more complicated scenarios.

2. The blue whale survey data. Line-transect cetacean surveys were carried out in the Eastern Tropical Pacific Ocean (ETP) between 1986 and 2007. Figure 1 shows the survey region and transect lines over this whole period, together with blue whale sightings. The survey area is 21.353 million square kilometres and is large enough that the curvature of the earth needs to be taken into account in the analysis. A total of 182 blue whale groups were sighted over all years, with a mean group size of 1.8 (standard deviation 2.1). In 1986–1990, 1998–2000, 2003 and 2006, the entire ETP area of was sampled. These complete surveys required two oceanographic research vessels (3 in 1998) for 120 sea days each. Transect search effort was stratified by area [Gerrodette and Forcada (2005)], and in 1992, 1993 and 2007, only part of the ETP area was sampled. These spatial differences in intensity of sampling need to be accounted for in modelling (see Section 4.2 for more detail). Data collection followed standardized line-transect protocols [Kinzey, Olson and Gerrodette (2000)]. Briefly, in workable conditions, a visual search for

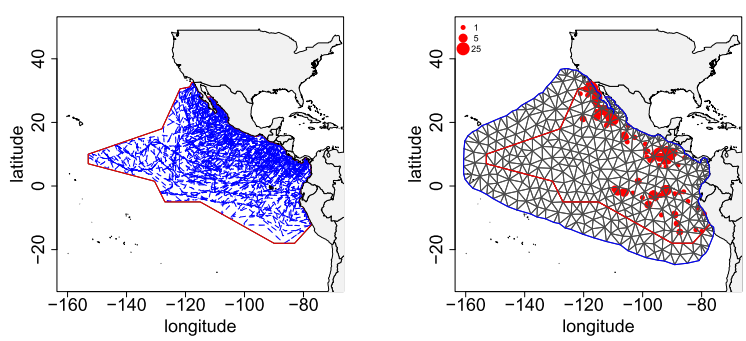

FIG. 1. *Plot of the ETP data: the left panel shows the transect lines surveyed from* 1986 *to* 2007, *and the right panel displays the sightings of blue whale groups* (*red dots*) *on top of the mesh used in our analysis: the radius of each dot is proportional to the logarithm of the group size plus* 1. *The red line is the ETP survey region boundary.*

cetaceans was conducted by a team of three observers on the flying bridge of each vessel during all daylight hours as the ship moved along the transect at a speed of 10 knots. Pedestal-mounted 25x binoculars were fitted with azimuth rings and reticles for angle and distance measurements. If a blue whale sighting was less than 10 km from the transect, the team went off-effort and directed the ship to leave the transect to approach the sighted animal(s). The observers identified the sighting to species or subspecies (if possible) and made group-size estimates.

The inference problem we address is how to model the density of blue whale groups across this survey region in a way that takes account of (i) the variable survey effort (transect lines) in space, (ii) the unknown probability of detecting a group from a line, with detection probability decreasing with distance from line, (iii) the dependence of density on explanatory variables (sea surface temperature in particular), (iv) how density changes over years, and (v) spatial fluctuation in blue whale density that cannot be explained by any available explanatory variables.

We describe the statistical models and tools that we use to address this inference problem next, and then use these to address the blue whale inference problem.

3. The models. Spatial point processes model the locations of objects in space [Stoyan and Grabarnik (1991), van Lieshout (2000), Diggle (2003), Møller and Waagepetersen (2004), Illian et al. (2008)]. Before incorporating distance sampling, we consider spatial point patterns formed by objects, represented as collections of locations, $\boldsymbol{Y} \equiv \{\boldsymbol{s}_i, i = 1, \dots, n\}$. The point set $\boldsymbol{Y}$ is considered as a realisation from a random point process on a bounded domain Ω, where usually $\Omega \subset \mathbb{R}^2$. Since the ETP survey domain is large enough for the curvature of the Earth to matter (see Figure 1), we treat Ω as a subdomain of a sphere, $\Omega \subset \mathbb{S}^2$.

3.1. *Spatial hierarchical Poisson point process models.* For any subset $A \subseteq \Omega$, the number of objects in A is denoted $N_{\boldsymbol{Y}}(A)$. For an inhomogeneous point process, we define an intensity function $\Lambda(\boldsymbol{s})$ as

$$\Lambda(\boldsymbol{s}) = \lim_{\varepsilon \to 0} \frac{\mathrm{E}\{N_{\boldsymbol{Y}}[\mathcal{B}_\varepsilon(\boldsymbol{s}, t)]\}}{|\mathcal{B}_\varepsilon(\boldsymbol{s}, t)|}, \qquad \boldsymbol{s} \in \Omega,$$

where $\mathcal{B}_\varepsilon(\boldsymbol{s})$ is a ball of radius ε centered at $\boldsymbol{s}$. For all nonoverlapping subsets $A_1, \dots, A_m \subset \Omega$, an inhomogeneous Poisson point process has the following two conditions:

$$N_{\boldsymbol{Y}}(A_k) \sim \mathrm{Po}\left[\int_{A_k} \Lambda(\boldsymbol{s})\, \mathrm{d}\boldsymbol{s}\right], \qquad k = 1, \dots, m, \text{ and}$$

$$N_{\boldsymbol{Y}}(A_1), \dots, N_{\boldsymbol{Y}}(A_m) \text{ are mutually independent.}$$

Finally, we let $\Lambda(\boldsymbol{s})$ be a random process, and define the point pattern model conditionally on $\Lambda(\boldsymbol{s})$. The conditional likelihood for the entire point pattern $\boldsymbol{Y}$, relative

to a homogeneous Poisson process with intensity 1, is given by

$$\pi(\mathbf{Y} \mid \Lambda) = \exp\left(|\Omega| - \int_{\Omega} \Lambda(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}\right) \prod_{i=1}^{N_{\boldsymbol{Y}}(\Omega)} \Lambda(\boldsymbol{s}_i), \tag{3.1}$$

where $\boldsymbol{s}_i$ is the location of the ith observation. If $\log \Lambda(\boldsymbol{s})$ is modelled by a latent Gaussian linear model, the resulting hierarchical model is a doubly-stochastic log-Gaussian Cox process [Møller and Waagepetersen (2004)].

3.2. *Point process models in the context of distance sampling*. For wildlife surveys, only a proportion of the population in the domain of interest is observed, due to partial sampling of the domain, and failure to detect all animals in the sampled regions. Distance sampling provides a method to account for imperfect detection. In line transect surveys, an observer traces a path through space, searching a strip centered on the path. The probability of detecting an object typically decreases with distance from the observer. From a modelling perspective, this results in a *thinned* spatial point process with the intensity function scaled by the detection probability [Dorazio (2012), Hefley and Hooten (2016), Johnson, Hooten and Kuhn (2013)].

When deriving the appropriate likelihood model for an observed point pattern, the problem-specific underlying generative structure influences the potential dependence between point locations both over space and in time. It is therefore important to note that the thinning in transect surveys is neither a thinning of a fixed spatial point pattern, nor a thinning of a regular spatio-temporal point process. Instead, each object is characterised by a temporally evolving curve in space, describing its movement, and the observations are thinned snapshots of time-slices of the resulting point process of curves. In addition, the intensity may vary over time, and we write $\lambda(\boldsymbol{s}; t)$ for the spatial point intensity for the full time-slice point pattern at time t, and $\Lambda(\boldsymbol{s}; t)$ for the intensity of the observationally thinned version. The assumptions about the movements of the observer and the objects affect what approximations are allowed in practical calculations.

Traditionally, the detection probability for an object located at a given perpendicular distance z from the path of the observer is modelled by a *detection function* $g(z)$. Assuming that the observer is moving with constant speed v along a straight line, standard Poisson process theory yields the probability of detecting an object located at $\boldsymbol{s}_0$ as a function of the perpendicular distance $z(\boldsymbol{s}_0)$,

$$\mathrm{P}(\text{object at } \boldsymbol{s}_0 \text{ detected} \mid \boldsymbol{s}_0 \in \boldsymbol{Y}) = 1 - \exp\{-h[z(\boldsymbol{s}_0)]/v\} = g(z(\boldsymbol{s}_0), v),$$

where $h(\cdot)$ is an aggregated detection *hazard* along the path, and $g(\cdot, \cdot)$ is the aggregated detection function, with explicit dependence on v. The standard approach is to model either the aggregated detection probability $g(z, v)$, or the aggregated hazard $h(z)$. Under simple assumptions about the observers, Hayes and Buckland (1983) derived the commonly-used *hazard-rate model*, given by $h(z) =$

$-(z/\sigma)^{-b}$, $b, \sigma > 0$. The half-normal detection function $g(z) = \exp[-z^2/(2\sigma^2)]$, $\sigma > 0$ is another widely-used model. While the hazard-rate model is more flexible than the half-normal detection model, only the latter results in a log-linear probability model. For this reason, the hazard-rate model does not fit directly into the existing INLA estimation software [Rue, Martino and Chopin (2009)], and instead we use a semiparametric detection model, which we introduce in Section 3.6, to give us a more flexible model than the half-normal.

3.3. *Line transect point process likelihood.* For line transects, assuming that environmental and other observational conditions that might affect detectability remain constant along suitably short and straight *transect segments*, we can formulate a tractably simple version of the likelihood. The region of space swept by the transect path is assumed to consist of a sequence of rectangular transect segment strips $\{\mathcal{C}_1, \dots, \mathcal{C}_K\}$, so that $\mathcal{C}_{k(t)}$ is the transect strip at time t. Writing $\lambda(\boldsymbol{s}; t)$ for the intensity of potentially observable objects, and introducing transect-dependent detection functions $g_{k(t)}(\boldsymbol{s})$, the intensity for the thinned observational point process is $\Lambda(\boldsymbol{s}; t) = \lambda(\boldsymbol{s}; t) g_{k(t)}(\boldsymbol{s})$. Under some loose assumptions (see Supplement A.1 [Yuan et al. (2017)]), the joint conditional likelihood for the observed point pattern is the product of the conditional likelihoods for each individual transect segment,

$$\pi(\boldsymbol{Y} \mid \Lambda) = \exp\left(\sum_{k=1}^{K} |\mathcal{C}_k| - \sum_{k=1}^{K} \int_{\mathcal{C}_k} \Lambda(\boldsymbol{s}; t_{\mathcal{C}_k})\, d\boldsymbol{s}\right) \prod_{i=1}^{N_{\boldsymbol{Y}}} \Lambda(\boldsymbol{s}_i; t_i), \tag{3.2}$$

where $N_{\boldsymbol{Y}} = \sum_{k=1}^{K} N_{\boldsymbol{Y}}(\mathcal{C}_k)$ is the total number of observed objects, located at $(\boldsymbol{s}_i, t_i)$, $i = 1, \dots, N_{\boldsymbol{Y}}$. We do not specifically address the issue of *marks* here (features or quantities associated with detected groups or animals). Marks that do not affect the detection probability can be modelled alongside the object intensity $\lambda(\boldsymbol{s}; t)$, including possible common fixed effects and dependent random effects [Illian, Sørbye and Rue (2012)]. However, marks that do affect the detection probability, such as the sizes of groups of animals, require a joint likelihood expression for the extended dimension point process of object locations and their marks, which is a topic for further development.

3.4. *A Bayesian hierarchical spatio-temporal point process model.* Following the classical approach for log-Gaussian Cox processes, we let the logarithm of the intensity $\lambda(\boldsymbol{s}; t)$ be a Gaussian process, with linear covariates $\boldsymbol{x}(\boldsymbol{s}, t)$, and a zero mean additive Gaussian spatial or spatio-temporal random field $\xi(\boldsymbol{s}, t)$ [Møller, Syversveen and Waagepetersen (1998), Møller and Waagepetersen (2004, 2007)]. For computational efficiency, we use the INLA method for numerical Bayesian inference with Gaussian Markov random fields [Rue, Martino and Chopin (2009), Illian, Sørbye and Rue (2012), Simpson, Lindgren and Rue (2012)], but the general methodology is not tied to a specific inferential framework.

In the likelihood given by (3.2), the log of the thinned intensity is given by

$$\begin{aligned}\log[\Lambda(\boldsymbol{s};t)] &= \log[\lambda(\boldsymbol{s};t)] + \log[g_{k(t)}(\boldsymbol{s})] \\ &= \boldsymbol{x}(\boldsymbol{s},t)^\top \boldsymbol{\beta} + \xi(\boldsymbol{s},t) + \log[g_{k(t)}(\boldsymbol{s})],\end{aligned} \tag{3.3}$$

where we assume Gaussian priors for $\boldsymbol{\beta}$, and a Gaussian random field ξ. If the logarithm of the detection probability model is linear in its parameters, this results in a joint linear model with latent Gaussian components.

In general, any link function and spatially coherent linear predictor could be used for Λ. The point process likelihood only requires Λ to be well-defined pointwise, and integrable. In practice, the numerical integration schemes used for practical likelihood evaluation (see Section 4) require piecewise continuity and differentiability. Covariates affecting $\lambda(\boldsymbol{s};t)$ need to be available throughout the transect region for parameter inference, and throughout the domain of interest for spatial prediction. For practical implementation reasons, spatial covariates are projected onto the same computational function space as the latent field ξ (see Section 4.1). Covariates affecting $g_k(\boldsymbol{s})$ need to be available for each transect segment. Within-segment variation in detectability would require a more expensive numerical integration scheme in Section 4.1, equivalent to splitting segments until they were sufficiently short for our assumption of constant detectability within each segment to be fulfilled. As noted at the end of Section 3.3, marks for individuals are currently only allowed if they do not affect the detection probability.

The full model is given by the following hierarchy:

$$\pi(\boldsymbol{Y},\xi,\boldsymbol{\beta},\boldsymbol{\beta}_g,\boldsymbol{\theta}) = \pi(\boldsymbol{Y} \mid \xi,\boldsymbol{\beta},\boldsymbol{\beta}_g,\boldsymbol{\theta})\pi(\xi \mid \boldsymbol{\theta})\pi(\boldsymbol{\beta} \mid \boldsymbol{\theta})\pi(\boldsymbol{\beta}_g \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}),$$

where $\boldsymbol{\beta}_g$ are parameters controlling the detection model, $\boldsymbol{\theta}$ are further model parameters, such as precision parameters for the latent Gaussian variables. Each of the prior densities $\pi(\xi \mid \boldsymbol{\theta})$, $\pi(\boldsymbol{\beta} \mid \boldsymbol{\theta})$, $\pi(\boldsymbol{\beta}_g \mid \boldsymbol{\theta})$, and $\pi(\boldsymbol{\theta})$ are controlled by hyperparameters. Note that in some software packages, including INLA, the parameters $\boldsymbol{\theta}$ themselves are referred to as hyperparameters.

For given prior distributions, the goal is to compute the posterior densities for the latent variables, optionally with $\boldsymbol{\theta}$ integrated out:

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{Y}) \propto \left.\frac{\pi(\boldsymbol{Y},\xi,\boldsymbol{\beta},\boldsymbol{\beta}_g,\boldsymbol{\theta})}{\pi(\xi,\boldsymbol{\beta},\boldsymbol{\beta}_g \mid \boldsymbol{Y},\boldsymbol{\theta})}\right|_{(\xi,\boldsymbol{\beta},\boldsymbol{\beta}_g)=(\xi^*,\boldsymbol{\beta}^*,\boldsymbol{\beta}_g^*)},$$

$$\pi(\xi,\boldsymbol{\beta},\boldsymbol{\beta}_g \mid \boldsymbol{Y}) = \int \pi(\xi,\boldsymbol{\beta},\boldsymbol{\beta}_g \mid \boldsymbol{Y},\boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid \boldsymbol{Y})\,\mathrm{d}\boldsymbol{\theta},$$

where $(\xi^*,\boldsymbol{\beta}^*,\boldsymbol{\beta}_g^*)$ is an arbitrary latent variable state vector. In the INLA method [Rue, Martino and Chopin (2009)], this is achieved approximately by replacing $\pi(\xi,\boldsymbol{\beta},\boldsymbol{\beta}_g \mid \boldsymbol{Y},\boldsymbol{\theta})$ with various Gaussian or near-Gaussian approximations, and integrating numerically over $\boldsymbol{\theta}$.

3.5. *Stochastic PDE models*. The general model construction requires no particular assumptions on how the spatial or spatio-temporal random field $\xi(\boldsymbol{s}, t)$ is modelled or treated computationally. The only requirement is that the model can be written as a latent Gaussian random field in such a way that the model likelihood can be evaluated numerically. In the context of INLA, that means that we need to construct a Gaussian Markov random field representation of the continuous space process. The traditional approach is to discretize space into a lattice and count the number of sighted points in each lattice cell, but here we take an alternative approach that allows us to use the true sighting locations, and to let $\lambda(\boldsymbol{s}; t)$ vary continuously through space. The results from Lindgren, Rue and Lindström (2011) show how to take advantage of the connection between Gaussian Markov random fields of graphs and stochastic partial differential equations in continuous space. Some details of such models are given in Supplement A.2 and the computational implications are discussed in Section 4.1.

3.6. *Log-linear detection function models*. As noted in Section 3.2, the hazard-rate model is not a log-linear model, which means that estimating the parameters does not directly fall under the latent Gaussian model framework of the INLA estimation software [Rue, Martino and Chopin (2009)]. In contrast, the half-normal detection model is $g_k(\boldsymbol{s}) = \exp[-z_k(\boldsymbol{s})^2/(2\sigma_{g,k}^2)]$, where $z_k(\boldsymbol{s})$ is the perpendicular distance from $\boldsymbol{s}$ to the kth transect line segment, and $\sigma_{g,k}$ are scale parameters. This can be written in log-linear form as $\log[g_k(\boldsymbol{s})] = \beta_{g,k} z_k^*(\boldsymbol{s})$, where $z_k^*(\boldsymbol{s}) = -z_k(\boldsymbol{s})^2/2$, and $\beta_{g,k} = 1/\sigma_{g,k}^2$. To allow more flexibility within the log-linear framework, we introduce a semiparametric piecewise quadratic model for the logarithm of the detection function, based on a one-dimensional version of the SPDE in the previous section.

For ease of presentation, first assume that the detection probability is the same for all transects k, so that we can write $G[z(\boldsymbol{s})] = -\log[g_k(\boldsymbol{s})]$. The prior distribution for $G(z)$ is then defined by the spline-like stochastic differential equation

$$\gamma \frac{\mathrm{d}^2 G(z)}{\mathrm{d}z^2} = \mathcal{W}(t), \qquad t \in [0, z_{\max}] \subset \mathbb{R}, \tag{3.4}$$

$$G(0) = 0, \qquad \left.\frac{\mathrm{d}G(z)}{\mathrm{d}z}\right|_{z=0} = 0, \tag{3.5}$$

where $z_{\max}$ is the maximal detection distance, $\gamma > 0$ is a smoothness parameter, and $\mathcal{W}(t)$ is a white noise process. The boundary constraints ensure that the detection probability at distance $z = 0$ is 1, and that the probability is flat near $z = 0$.

Let $(0, z_1, z_2, \dots, z_p)$ be breakpoints for piecewise quadratic B-spline basis functions $B_i(z)$ [Farin (2002)], such that $z_p = z_{\max}$, with the simplest choice of breakpoints being $z_i = i z_{\max}/p$, $i = 0, 1, \dots, p$. The nonparametric model for

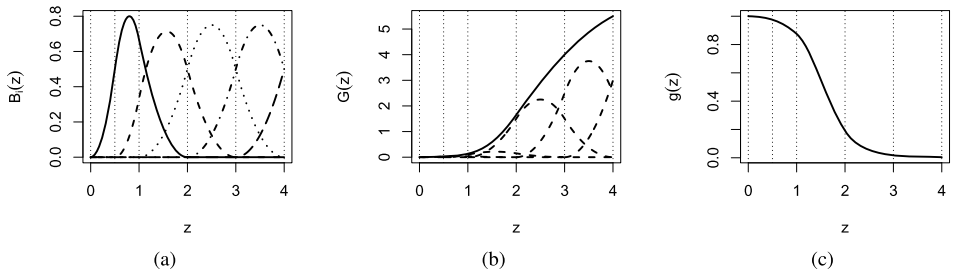

FIG. 2. *Illustration of the semiparametric detection function model*: (a) *B-spline basis functions* $B_i(z)$, $i = 1, \dots, 5$, *of order* 2 *that fulfill the boundary conditions* (3.5), (b) *the weighted functions* $\beta_i B_i(z)$ (*dashed*), *their sum* $G(z)$ (*solid*), *and* (c) *the detection probability* $g(z) = \exp[-G(z)]$. *The breakpoints are* $(0, z_1, \dots, z_5) = (0, 0.5, 1, 2, 3, 4)$ *and the weights are* $(\beta_1, \dots, \beta_5) = (0.05, 0.3, 3, 5, 6)$. *Note that the detection function does not automatically have more irregular behavior where the breakpoints are close together, and that this is probabilistically regulated by the stochastic differential equation* (3.4).

$G(z)$ can then be used to construct a finite dimensional model

$$G(z) = \sum_{i=1}^{p} \beta_i B_i(z), \tag{3.6}$$

where the p basis functions only include $B_i(\cdot)$ that fulfill the boundary conditions (3.5), as shown in Figure 2. The joint multivariate Gaussian prior distribution for $(\beta_1, \dots, \beta_p) \sim \mathrm{N}(\mathbf{0}, \boldsymbol{Q}_\beta^{-1}(\gamma))$ is constructed with the same finite element technique that will be used for the spatial SPDE in Section 4.1. For uniform breakpoint spacing, increasing p will make this discrete model converge to the continuous domain model, but for finite p the model is effectively a piecewise quadratic semiparametric model.

Imposing a monotonicity constraint on $g(z)$ is possible by replacing the basis functions for $G(z)$ with increasing basis functions, and mandating positivity of the β_i coefficients [Ramsay (1988)]. However, because the latter is currently only implemented in INLA for independent β_i, this is restricted to small p, such as 2 or 3, since an independence prior would result in a nonsmooth function for larger p. An alternative that is feasible when sampling from the posterior distribution is to simply reject all nonincreasing samples of $G(z)$. If the data are informative, the smoothing from the prior can in practice be enough to yield monotonic estimates without including an explicit constraint.

Relaxing the assumption that the detection function is the same for all transects is most easily done by adding further linear terms to $\log g_k(\boldsymbol{s})$ based on observed or constructed covariates that depend on k.

4. Computational methods. There are several practical considerations for evaluating the likelihoods and representing the random fields in such a way that

large dense matrices can be avoided. In Section 4.1, we give a brief overview of the essentials for translating stochastic PDEs into manageable Gaussian Markov random fields (GMRF), and Section 4.2 presents a numerical integration scheme for the point process likelihood.

4.1. *The stochastic partial differential equation approach.* The SPDE/GMRF approach works by replacing the continuous domain stochastic PDE model with a finite dimensional Gaussian Markov random field for basis function weights defined on a triangulation of the domain of interest, such that the sparse precision matrix leads to a good approximation of the continuous space SPDE solutions. Given a triangulation mesh (see the right panel of Figure 1 for the triangulation used for the ETP survey area), Lindgren, Rue and Lindström (2011) define a finite element representation [Brenner and Scott (2008)] of ξ from (A.1),

$$\xi(\boldsymbol{s}) = \sum_{j=1}^{m} w_j \phi_j(\boldsymbol{s}), \tag{4.1}$$

where $w_1, \dots, w_m$ are stochastic weights, and ϕ_j, $j = 1, \dots, m$, are deterministic piecewise linear basis functions defined for each node on the mesh: ϕ_j equals 1 at mesh node j and 0 in all the other mesh nodes. The weight vector $\boldsymbol{w} \equiv (w_1, \dots, w_m)^\top$ is a GMRF with its Markovian properties defined by the mesh structure. It follows that $\boldsymbol{w}$ determines the stochastic properties of (4.1) and $\boldsymbol{w}$ is chosen in a way that the distribution of (4.1) approximates the distribution of the solution to the SPDE (A.1). As shown by Lindgren, Rue and Lindström (2011), for the SPDE in (A.1), the resulting weight distribution is $\boldsymbol{w} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{Q}(\tau, \kappa)^{-1})$, where the sparse precision matrix $\boldsymbol{Q}(\tau, \kappa)$ is a polynomial in the parameters τ and κ, and is obtained through finite element calculations.

The practical implication of this construction is that instead of directly using the covariances from (A.2) in Supplement A, which results in dense covariance matrices and high computational cost, $\mathcal{O}(m^3)$, the SPDE/GMRF approach links the continuous and discrete domains in such a way that the computational cost is reduced to $\mathcal{O}(m^{1.5})$. The computational advantages of GMRFs [Rue and Held (2005)] is strengthened by using INLA for Bayesian inference [Rue, Martino and Chopin (2009)]. For the case of fully observed log-Gaussian Cox point processes, the in-depth analysis by Simpson et al. (2016) of the combined approximation errors induced by the basis function expansion in combination with the likelihood approximation error shows that the resulting approximate posterior distribution is close to the true posterior distribution. Since the integration scheme in the following section is constructed in the same way, we do not include a detailed approximation error analysis here, but note that the SPDE/GMRF approximation is likely to be the largest source of approximation error. Point patterns are relatively uninformative about the latent intensity, which has the practical effect that the realizations of the fields in the posterior distribution are typically smoother than in

directly observed process problems. Hence, the approximation error is very small as long as the triangle mesh edges are short compared with the spatial scales of the covariates and of the point pattern intensity variability.

4.2. *Numerical point process likelihood evaluation.* Combining the general distance sampling point pattern likelihood (3.2) with the log-linear model structure for $\Lambda(\boldsymbol{s}; t)$ from (3.3) results in a log-likelihood for the observed point pattern,

$$\begin{aligned}
&\log \pi(\boldsymbol{Y} \mid \lambda, g) \\
&\quad = \sum_{i=1}^{N_{\boldsymbol{Y}}} \{\boldsymbol{x}(\boldsymbol{s}_i, t_i)^\top \boldsymbol{\beta} + \xi(\boldsymbol{s}_i, t_i) - \log[g_{k(t_i)}(\boldsymbol{s}_i)]\} \\
&\qquad - \sum_{k=1}^{K} \int_{\mathcal{C}_k} \exp\{\boldsymbol{x}(\boldsymbol{u}, t_k)^\top \boldsymbol{\beta} + \xi(\boldsymbol{u}, t_k) - \log[g_k(\boldsymbol{u})]\} \, \mathrm{d}\boldsymbol{u} + \sum_{k=1}^{K} |\mathcal{C}_k|,
\end{aligned} \tag{4.2}$$

where the first term evaluates the log-intensity at the observed locations, and the second term integrates the intensity over the sampled transect segments. The log-likelihood (4.2) is in general analytically intractable as it requires integrals of the exponential of a random field. Therefore, we use numerical integration to evaluate (4.2), and the remainder of this section describes an integration scheme to approximate the integrals efficiently. As noted in Section 3.4, we assume that the covariates $x(\boldsymbol{s}, t)$ are expressed using the same piecewise linear basis functions as ξ. For cases where a covariate has a much finer resolution than the one needed for ξ, the efficient integration scheme developed here is not appropriate, and further research is needed to develop an integration method that can deal with that without incurring a high computational cost.

For distance sampling surveys, transect areas describe subsets of the earth's surface. The most natural representation of transect areas would therefore be subsets $\mathcal{C}_k \subseteq \mathbb{S}^2$ of the sphere, leading to surface integration in the Poisson process likelihood. However, the small scale at which earthbound observers are capable of probing their environment lends itself to easily justifiable simplifications of the numerical integration. Apart from environmental conditions such as the weather, the curvature of the earth puts an upper bound on the distance at which an observer with a given elevation can actually detect an animal. We therefore approximate the surface integrals over $\mathcal{C}_k \subset \mathbb{S}^2$ by integrals over $\widetilde{\mathcal{C}}_k \subset \mathbb{R}^2$,

$$\begin{aligned}
I_{\mathcal{C}_k} &= \int_{\mathcal{C}_k} \lambda(\boldsymbol{u}; t_k) g_k(\boldsymbol{u}) \mathrm{d}\mathcal{C}_k(\boldsymbol{u}) \\
&= \int_{\widetilde{\mathcal{C}}_k} \lambda\big(\boldsymbol{u}_k(l, z); t_k\big) g_k\big(\boldsymbol{u}_k(l, z)\big) \left\| \frac{\partial \boldsymbol{u}_k}{\partial l} \times \frac{\partial \boldsymbol{u}_k}{\partial z} \right\| \mathrm{d}l \, \mathrm{d}z \\
&\approx \int_{\widetilde{\mathcal{C}}_k} \lambda\big(\boldsymbol{u}_k(l, z); t_k\big) g_k\big(\boldsymbol{u}_k(l, z)\big) \, \mathrm{d}l \, \mathrm{d}z,
\end{aligned}$$

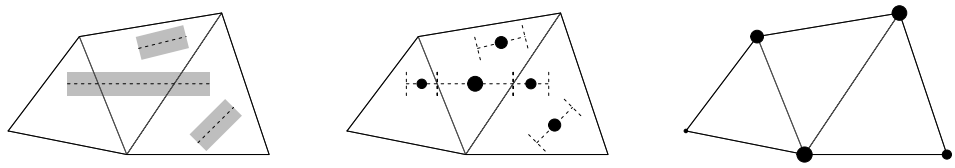

FIG. 3. *Integration scheme for transect lines. Panel* (a) *shows three triangles of a mesh, and three transects. When the width of the transects is small compared to the size of the triangles, and thus the slope of the intensity has small variability perpendicular to each segment, the areal integration can be reduced to an integration along the observer's trajectory weighted by the width of the transect. Trajectories that reside in multiple triangles can be split at the triangle edges such that their midpoints serve as integration points, assuming linear intensity within the triangles* (b). *Under the same assumption, integration points from each triangle are accumulated into re-weighted points at the mesh vertices* (c). *The areas of the filled circles are proportional to the integration weights.*

where we use a transect-specific parameterization $\boldsymbol{u}_k$ at coordinate l along and distance z to the transect line, respectively. If R is the radius of the earth, then the Jacobian is $\cos(z/R)$, which gives an approximation error of a factor less than $5 \cdot 10^{-6}$ even in the extreme and unrealistic case of an observer at 31 metres above a calm sea looking at the horizon 20 kilometres away.

Another fact that we can utilize is that the detection function g does not depend on the position of the observer along the line but only on the distance of an observation from the line. Similarly, if the transect line is narrow compared to the spatial rate of change in the intensity function, we can substitute the evaluation of λ by an evaluation at the center of the transect line, $\widehat{z} = 0$ (see Figure 3). That is,

$$I_{\mathcal{C}_k} \approx \int_{\widetilde{\mathcal{C}}_k} \lambda\big(\boldsymbol{u}_k(l, \widehat{z}); t_k\big) g_k\big(\boldsymbol{u}_k(\widehat{l}, z)\big) \mathrm{d}l \, \mathrm{d}z,$$

together with an arbitrary coordinate $\widehat{l}$ along the transect line.

REMARK. In a stand-alone implementation, the integral could be written as a product of two one-dimensional integrals, and even evaluated exactly due to the log-linearity of the model. Unfortunately, the resulting structure cannot be expressed using only evaluations of products of λ and g, which is a requirement imposed by the internal structure of the INLA implementation, so we do not use that approach here.

We can also make use of the fact that λ lives on a mesh. If the mesh triangles are small enough, the log-linear function $\lambda(\cdot)$ is approximately linear within each triangle. By splitting a transect line $\mathcal{C}_k$ into segments $\mathcal{C}_{k,j}$, $j \in 1, \ldots, J$, each of

which resides in a single triangle (see Figure 3), we obtain a Gaussian quadrature method of order one,

$$I_{\mathcal{C}_k} \approx \sum_{j=1}^{J} \int_{\widetilde{\mathcal{C}}_{k,j}} \lambda\big(\boldsymbol{u}_{kj}(l, \widehat{z}); t_k\big) g_k\big(\boldsymbol{u}_{kj}(\widehat{l}, z)\big)\, \mathrm{d}l\, \mathrm{d}z$$
$$\approx \sum_{j=1}^{J} w_{k,j} \int_{z} \lambda\big(\boldsymbol{u}_{kj}(l_{k,j}, \widehat{z}); t_k\big) g_k\big(\boldsymbol{u}_{kj}(\widehat{l}, z)\big)\, \mathrm{d}z.$$

Here, $l_{k,j}$ is half of segment j's length $w_{k,j}$. The integration over the distance parameter can now be approximated by a quadrature rule with an equidistant scheme, so that

$$I_{\mathcal{C}_k} \approx \sum_{j=1}^{J} \sum_{r=1}^{R} \widetilde{w}_{k,j} \lambda\big(\boldsymbol{u}_{kj}(l_{k,j}, \widehat{z}); t_k\big) g_k\big(\boldsymbol{u}_{kj}(\widehat{l}, z_r)\big),$$

where $\widetilde{w}_{k,j} = \frac{2z_{\max}}{R} w_{k,j}$ with maximal detection distance $z_{\max}$, and we substitute $\widehat{l} = l_{k,j}$. We can then write

$$I_{\mathcal{C}_k} \approx \sum_{j=1}^{J} \sum_{r=1}^{R} \widetilde{w}_{k,j} \lambda(\widetilde{\boldsymbol{u}}_{k,j,r}; t_k) g_k(\boldsymbol{u}_{k,j,r}),$$

where $\boldsymbol{u}_{k,j,r}$ are points on the perpendicular line through the midpoint of transect k's segment j, and $\widetilde{\boldsymbol{u}}_{k,j,r}$ is the midpoint of each subsegment line.

As a last step, we can again make use of the assumption that the function we are integrating over is approximately linear within a given triangle. It is straightforward to show (see Supplement A.3) that this means that each integration point can be expressed by an evaluation of the function at the triangle vertices weighted by the within-triangle Barycentric coordinates of the original point [Farin (2002)]. We can therefore summarize integration points that reside in the same triangle and share a common time coordinate t_k to such evaluations at the mesh vertices, illustrated in Figure 3. This can, depending on the problem structure, lead to a significant reduction in the respective computational workload.

The approximation error from treating the log-linear function within each triangle as linear can be reduced by subdividing each triangle into four. However, evaluating the function at the mid-points of the original triangle edges as well as the original vertices leads to an increase in the computational cost of at least a factor of four, since the number of edges is approximately three times the number of vertices in the mesh.

5. Estimating blue whale density from the Eastern Tropical Pacific surveys.

5.1. *The ETP surveys*. We use the above methods to predict the blue whale group density over the ETP survey area for the each of the survey years, and to study the effect of sea surface temperature (SST) on the blue whale group density. The function of interest for density estimation is the intensity of the point process before thinning, denoted $\lambda(\boldsymbol{s}, t)$ in (3.3), which we refer to as the group density.

These data have been analyzed before: Forney et al. (2012) used GAMs to estimate encounter rate, with a two-stage estimation approach and gridded data (counts of detections within small segments of transects). Pardo et al. (2015) also used gridded data, modelling log density in each grid cell as a polynomial function of the absolute dynamic topography, a spatially referenced variable that indicates vertical transport of nutrients, and thus productivity. While they included a random component in their density model, it had no spatial structure, assuming independent residuals among grid cells. They estimated all model parameters simultaneously in a hierarchical Bayesian framework. Two key differences between their and our model structures are that we use the ungridded data (i.e., the point locations of each detection rather than counts in user-defined grid cells) in our analysis, and we use a *spatially structured* Gaussian random field to capture spatial variation in density that is not explained by the observed explanatory variable.

Because the blue whale group size is small (mean of 1.8 and standard deviation 2.1) and the size is easily established, it is realistic to treat the group size of blue whales as known without error [Gerrodette, Perryman and Barlow (2002)]. We assume that the group size does not affect the detection probability. The detection of cetaceans on ship surveys also depends on wind conditions, but this is less important for blue whales because of their large body size and conspicuous blows. Therefore, we assume the detection probability depends only on the perpendicular distance for the blue whales in the ETP survey. In our analysis, we truncate the data at perpendicular distance $w = 6$ km. We also assume that distances were observed without error for each detected animal group and we fit a semiparametric model given by (3.6) to estimate the detection function.

To build a spatio-temporal model using the SPDE approach described in Section 4.1, we start by constructing a mesh for the ETP survey as shown in Figure 1. The ETP survey is bounded partially by the coastline and partially by the red line of Figure 1. We use a simplified representation of the actual coastline as the mesh boundary to incorporate the boundary effect because a physical boundary such as the coastline has a strong effect—in this case, there will be no blue whales on land. We use a simplified representation of the coastline, because the actual coastline is too 'angular', and hence problematic for the SPDE approach [Lindgren and Rue (2015)]. Meanwhile, we extend the boundary further in the northwest and south directions in the Eastern Tropical Pacific, to exclude the boundary effect for the part of the survey area that is bounded by the red line in Figure 1. Given the low sighting rate of the blue whales, there is little information contained in the data

to fit a complicated spatio-temporal stochastic process for the random field, such as the AR(1) temporal process used by Cameletti et al. (2013); even the simpler version of the spatio-temporal process with replicated spatial field over time is not feasible.

We consider three models, all with Gaussian random fields in space alone. Model 0 has latitude and longitude fixed effects but ignores SST. Model 1 has a temporal SST fixed effect together with spatial residual SST fixed effects for each year. Model 2 has a temporal SST fixed effect and a spatial SST fixed effect together with spatio-temporal residual SST fixed effects.

5.2. *Incorporating a spatio-temporal environmental covariate*: *Sea surface temperature* (*SST*). Based on the Simple Ocean Data Assimilation (SODA) model (http://apdrc.soest.hawaii.edu/datadoc/soda_2.2.4.php), the SST data are available on a fine grid over the ETP survey area on a monthly scale between 1986 and 2007. First, within each year, SST is averaged over the months July to December, during which the survey was conducted. Second, these temporally averaged SST values are spatially smooth, and can be projected onto the mesh of the survey area with only minor loss of fine-scale information. Piecewise linear interpolation is used to calculate the SST for any given location and year, denoted by $\mathrm{sst}(\boldsymbol{s}, t)$.

Figure 4 shows the centered SST averaged over time and the centered SST averaged over space. There is both spatial and inter-year variation in SST, and we use hierarchical centering to separate the annual and spatial effects of SST.

Hierarchical centering is a commonly used technique in multilevel modelling [Kreft, de Leeuw and Aiken (1995)], and we consider two different centering schemes for SST here. Model 0 does not incorporate SST, Model 1 incorporates

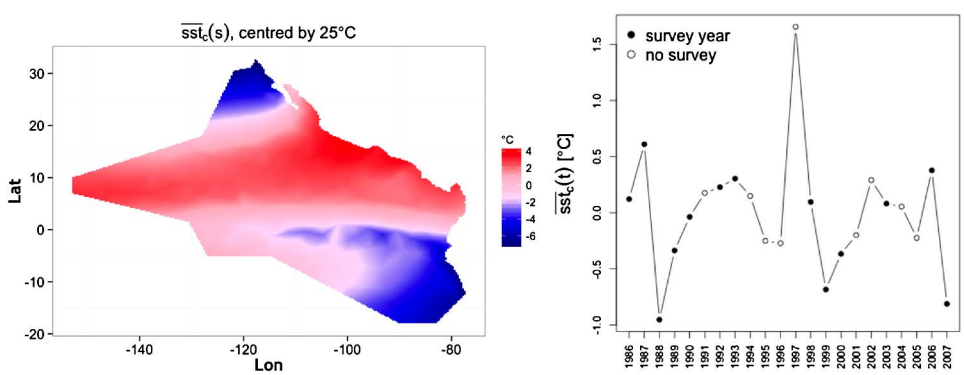

FIG. 4. *Sea surface temperature. The centered time-average temperature* $\overline{\mathrm{sst}}_c(\boldsymbol{s})$ *is shown in the left panel, while the centered space-averaged temperature* $\overline{\mathrm{sst}}_c(t)$ *is shown in the right panel. Solid circles represent survey years and the empty circles represent nonsurvey years. The SST in* 1997 *is extreme relative to all survey years.*

SST using within-year centering, and Model 2 uses space–time centering. They all have the same SPDE specification of the latent spatial Gaussian random field:

1. Model 1: within-year centering. This model has two SST components, the spatially averaged SST for each year, and the spatial SST patterns centered within each year. Let Ω denote the bounded ETP survey area in Figure 1. We use $\overline{\text{sst}}_c(t)$ to denote the SST averaged over the ETP survey area for year t after centering,

$$\overline{\text{sst}}_c(t) = \frac{1}{|\Omega|}\int_{\Omega} \text{sst}(\boldsymbol{s}, t)\,\mathrm{d}\boldsymbol{s} - \overline{\text{sst}}, \tag{5.1}$$

where $\overline{\text{sst}}$ denotes the overall average of SST, $\overline{\text{sst}} = \int_{\Omega\times\mathbb{T}} \text{sst}(\boldsymbol{s}, t)\,\mathrm{d}\boldsymbol{s}\,\mathrm{d}t/(|\Omega| \times |\mathbb{T}|)$, with $\mathbb{T}$ denoting the set of survey years. Then the SST centered within year t for location $\boldsymbol{s}$, $\text{sst}_{\text{cwy}}(\boldsymbol{s}, t)$, is defined as

$$\text{sst}_{\text{cwy}}(\boldsymbol{s}, t) = \text{sst}(\boldsymbol{s}, t) - \overline{\text{sst}}_c(t) - \overline{\text{sst}}. \tag{5.2}$$

2. Model 2: space–time centering. This model separates the spatial and temporal patters from a spatio-temporal interaction and has three SST components. These are the $\overline{\text{sst}}_c(t)$ given by (5.1), the SST averaged over years for each location, $\overline{\text{sst}}_c(\boldsymbol{s})$ given by (5.3), and the SST residuals, $\text{sst}_{\text{res}}(\boldsymbol{s}, t)$ given by (5.4),

$$\overline{\text{sst}}_c(\boldsymbol{s}) = \frac{1}{|\mathbb{T}|}\int_{\mathbb{T}} \text{sst}(\boldsymbol{s}, t)\mathrm{d}t - \overline{\text{sst}}, \tag{5.3}$$

$$\text{sst}_{\text{res}}(\boldsymbol{s}, t) = \text{sst}(\boldsymbol{s}, t) - \overline{\text{sst}}_c(t) - \overline{\text{sst}}_c(\boldsymbol{s}) - \overline{\text{sst}}. \tag{5.4}$$

$\overline{\text{sst}}_c(t)$ indicates whether a year is relatively warm or cold after averaging over the survey area, and similarly, $\overline{\text{sst}}_c(\boldsymbol{s})$ indicates whether a location is relatively warm or cold after averaging over the survey years. The SST residual, $\text{sst}_{\text{res}}(\boldsymbol{s}, t)$, contains information about the interaction between temporal pattern and the spatial average of SST.

Plots of raw and centered SST used in the analysis are given in Supplement B.1.

La Niña conditions are characterized by a band of cooler waters in 1988, 1999, and 2007, and El Niño conditions by a much wider band of warm ocean water in 1997 [1987 is a moderately strong El Niño year according to the scale by Wolter and Timlin (2011)]. Centering SST strongly captures the El Niño/La Niña oscillations that occur at irregular intervals in the ETP survey area (see Supplement B.1 for more detail). The temporal effect of SST after centering using (5.1) (see Figure 4) correctly reflects the La Niña conditions in 1988 and 2007, and strongly highlights the El Niño year 1997 as an outlier. Unfortunately, no survey was conducted in 1997. Given the time series of $\overline{\text{sst}}_c(t)$ for the survey years and nonsurvey years in the right panel of Figure 4, it is obviously problematic to predict for 1997 using a model fitted on the data from the survey years, which are represented by the filled circles in Figure 4. Therefore, we make predictions for all years except 1997.

TABLE 1
The posterior estimates for the fixed-effects coefficients for each model

Model	Parameter	Mean	Std.dev.	Quantile 2.5%	50%	97.5%
Model 0	β_0	−12.29	2.29	−18.04	−11.99	−8.56
	β_{lon}	0.10	0.07	−0.05	0.10	0.26
	β_{lat}	0.01	0.09	−0.22	0.02	0.16
Model 1	β_0	−4.58	3.04	−11.00	−4.44	1.06
	$\beta_{\overline{\text{sst}}_c(t)}$	0.79	0.21	0.38	0.78	1.20
	$\beta_{\text{sst}_{\text{cwy}}}$	−0.28	0.10	−0.48	−0.28	−0.07
Model 2	β_0	−11.85	2.24	−17.34	−11.60	−7.85
	$\beta_{\overline{\text{sst}}_c(t)}$	0.73	0.21	0.32	0.73	1.16
	$\beta_{\overline{\text{sst}}_c(\boldsymbol{s})}$	−0.60	0.14	−0.88	−0.60	−0.34
	$\beta_{\text{sst}_{\text{res}}}$	0.22	0.17	−0.10	0.22	0.55

5.3. *Results*. Table 1 summarizes the posterior density of the regression coefficients for each model.

5.3.1. *The effects of longitude and latitude*. Model 0 contains only longitude and latitude as covariates. The 95% posterior credible intervals for the regression coefficients from this model both include zero with medians very close to zero. This suggests that there is no large-scale log-linear spatial effect that can be explained by longitude and latitude. This interpretation is supported by the results from models that include SST. Specifically, when we add longitude and latitude to Models 1 and 2, the 95% posterior credible intervals of the longitude and latitude regression parameters still include zero. We therefore exclude longitude and latitude, and henceforth consider only Models 1 and 2.

5.3.2. *The effects of SST*. From (5.2), (5.3), and (5.4), we have $\text{sst}_{\text{cwy}}(\boldsymbol{s}, t) = \overline{\text{sst}}_c(\boldsymbol{s}) + \text{sst}_{\text{res}}(\boldsymbol{s}, t)$, so that $\beta_{\text{sst}_{\text{cwy}}}$ in Model 1 amounts to combining $\beta_{\overline{\text{sst}}_c(\boldsymbol{s})}$ and $\beta_{\text{sst}_{\text{res}}}$ in a single parameter. The negative posterior median and 95% credible interval (2.5% to 97.5% quantiles) of $\beta_{\overline{\text{sst}}_c(\boldsymbol{s})}$ and of $\beta_{\text{sst}_{\text{cwy}}}$ indicate that locations that are colder on average over the years are expected to have more blue whale groups than locations that are warmer on average, while the opposite sign of the spatio-temporal interaction $\beta_{\text{sst}_{\text{res}}}$ indicates that this effect is weaker at locations with higher temperature in a given year than the across-year average temperature at the location.

The posterior median estimates of $\beta_{\overline{\text{sst}}_c(t)}$ are similar for Models 1 and 2, indicating that the effect of warmer average temperature in a year, conditional on the spatial effect and random field, is to increase density. The ETP survey design is

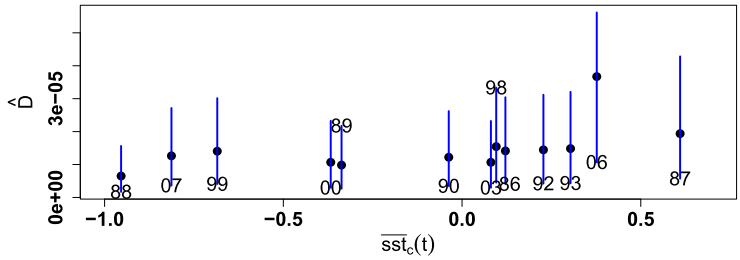

FIG. 5. *The predicted number of blue whale groups per unit area* ($\hat{D}$) *from Model* 2, *together with* 95% *credible intervals against centered mean annual temperature* $\overline{\text{sst}}_c(t)$. *Numbers indicate the year in question.*

not balanced in that it does not have survey effort in every year along each transect that was surveyed in any year and as a result we need to be a bit cautious about interpreting parameters. To investigate the effect of annual mean temperature, we therefore also considered the posterior distribution of the predicted number of blue whale groups per unit area. This is shown in Figure 5. While this plot is consistent with the estimates of $\beta_{\overline{\text{sst}}_c(t)}$ from Models 1 and 2, it is also consistent with an hypothesis of no change in average density across the years, as a horizontal line falls well within the 95% credible intervals of all estimates. The other notable feature of the plot is the unusually high estimated density for the second- warmest year, 2006. The reasons for this are unclear.

5.3.3. *Posterior median density and its relative uncertainty.* The posterior median of blue whale density, $\lambda(\boldsymbol{s}; t)$, for year 1986 is shown in the top panel of Figure 6 for Models 0, 1, and 2, respectively. The top three plots of Figure 6 are very similar, with areas of higher blue whale group density in the north off the coast of Baja California, in the area of the Costa Rica Dome off the coast of Central America, and in the south-east in the vicinity of the Galapagos Islands. This pattern of the blue whale group density is consistent across all the models implemented, and reflects what we observe in the sightings data in the right panel of Figure 1. This observed spatial pattern is also in general agreement with previous analysis of blue whale sighting data in the ETP [Forney et al. (2012), Pardo et al. (2015)]. Similar plots for 1986–2007 (omitting the very strong El Niño year 1997), are given in Supplement B.2.

We use the relative width of the 95% posterior credible interval (RWPCI) as a measure of the relative uncertainty for the predicted $\lambda(\boldsymbol{s}; t)$ of the ETP survey area. We define the RWPCI as the inter-quartile range divided by the median,

$$\text{RWPCI} = (Q_3 - Q_1)/Q_2. \tag{5.5}$$

When the posterior distribution is approximately Gaussian, the RWPCI is about 1.35 times the ratio of the posterior standard deviation to the posterior median. The

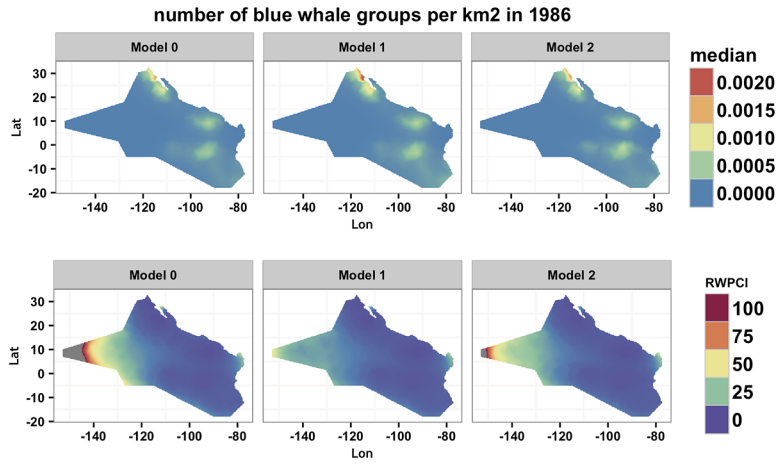

FIG. 6. *The posterior median (top) and RWPCI from* (5.5) *(bottom) of the ETP blue whale groups density in* 1986 *using Models* 0, 1, *and* 2 *in Table* 1. *The RWPCI colour palette is cut off at* 100 *to exclude the extreme values at the western corner of the ETP survey area.*

bottom panel of Figure 6 shows the spatial structure of the RWPCI in 1986 for each of the three models, and this pattern persists across years (see Supplement B.2). The far west of the survey region has very high relative uncertainty because it is close to the edge of the mesh boundary shown in Figure 1 and there are no sightings in that area. The spatial random field has high uncertainty in this area: regions of low $\lambda(\boldsymbol{s}; t)$ tend to have higher uncertainty associated with the latent field. The slowly varying standard deviation of the latent field in Figure 7 is likely due to a

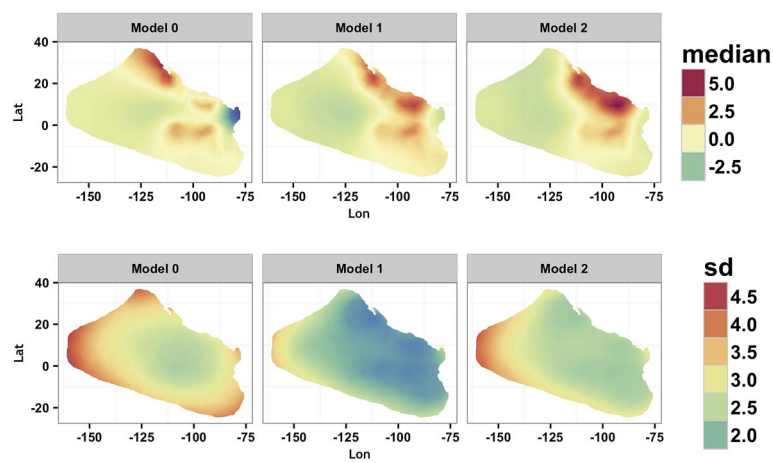

FIG. 7. *The posterior median and standard deviation of the latent field* (4.1) *for Models* 0, 1 *and* 2.

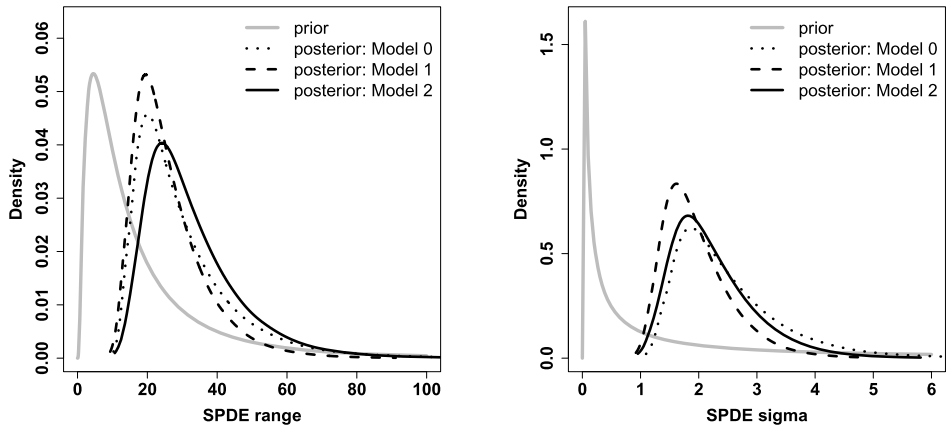

FIG. 8. *The posterior densities of the SPDE parameters using Models* 0, 1, *and* 2. *The left panel is for the range parameter* ρ (*see Section* 3.5 *for its definition*), *and the right panel for the marginal standard deviation* σ_ξ *in* (A.1).

combination of large spatial range (see Figure B.8 of Supplement B) and the fact that the observed point pattern is not very informative about the latent field.

5.3.4. *SPDE parameters and detection function*. Prior sensitivity tests of the SPDE parameters showed the posterior median of $\lambda(\boldsymbol{s}; t)$ to be less sensitive to prior specification than is its variance. Details of the SPDE prior specification are given in Supplement B.3. Figure 8 displays the posterior densities of the SPDE parameters for Models 0, 1, and 2 using the same prior.

The large range of the Matérn covariance function is consistent with the latent Gaussian random field [ξ of (3.3)] shown in Figure 7. There is little difference among the models for either the posterior detection function or the 95% credible band (see Figure B.9 of Supplement B).

5.4. *Exploratory model checking*. Let $\eta(\boldsymbol{s}, t)$ denote the log-intensity defined by the fixed effects and random field components of (3.3),

$$\eta(\boldsymbol{s}, t) = \log[\lambda(\boldsymbol{s}; t)] = \boldsymbol{x}(\boldsymbol{s}, t)^\top \boldsymbol{\beta} + \xi(\boldsymbol{s}, t).$$

To investigate the role of the components and the possibility of confounding, we consider the variability around the posterior mean of the overall averages of $\eta(\boldsymbol{s}, t)$;

$$M_\eta = \frac{1}{|\Omega| \times |\mathbb{T}|} \int_{\boldsymbol{s} \in \Omega} \int_{t \in \mathbb{T}} E[\eta(\boldsymbol{s}, t) \mid \boldsymbol{Y}] \, \mathrm{d}t \, \mathrm{d}\boldsymbol{s},$$

and similarly for the components, so that $M_\eta = M_\beta + M_\xi$. The posterior expected squared deviation of $\eta(\boldsymbol{s}, t)$ from M_η can be split into contributions from the fixed

TABLE 2
The posterior space–time averages $V_\eta(\Omega, \mathbb{T})$, $V_\beta(\Omega, \mathbb{T})$, $V_\xi(\Omega, \mathbb{T})$, covariance $C_{\beta,\xi}(\Omega, \mathbb{T})$ and correlation $\rho_{\beta,\xi}(\Omega, \mathbb{T})$ for Models 0, 1, *and* 2

	$V_\eta(\Omega, \mathbb{T})$	$V_\beta(\Omega, \mathbb{T})$	$V_\xi(\Omega, \mathbb{T})$	$C_{\beta,\xi}(\Omega, \mathbb{T})$	$\rho_{\beta,\xi}(\Omega, \mathbb{T})$
Model 0	7.32	10.05	9.86	−6.29	−0.63
Model 1	5.04	3.19	5.97	−2.06	−0.47
Model 2	6.82	7.51	9.40	−5.05	−0.60

effects $\boldsymbol{x}(\boldsymbol{s}, t)^\top \boldsymbol{\beta}$ and the random field $\xi(\boldsymbol{s}, t)$;

$$\begin{aligned} V_\eta(\boldsymbol{s}, t) &= E\{[\eta(\boldsymbol{s}, t) - M_\eta]^2 \mid \boldsymbol{Y}\} \\ &= E\{[\boldsymbol{x}(\boldsymbol{s}, t)^\top \boldsymbol{\beta} - M_\beta]^2 \mid \boldsymbol{Y}\} + E\{[\xi(\boldsymbol{s}, t) - M_\xi]^2 \mid \boldsymbol{Y}\} \\ &\quad + 2E\{[\boldsymbol{x}(\boldsymbol{s}, t)^\top \boldsymbol{\beta} - M_\beta][\xi(\boldsymbol{s}, t) - M_\xi] \mid \boldsymbol{Y}\} \\ &= V_\beta(\boldsymbol{s}, t) + V_\xi(\boldsymbol{s}, t) + 2C_{\beta,\xi}(\boldsymbol{s}, t). \end{aligned}$$

For all models, $\xi(\boldsymbol{s}, t)$ is constant over time, and we define the averages across time, $V_\eta(\boldsymbol{s}, \mathbb{T})$, $V_\beta(\boldsymbol{s}, \mathbb{T})$, and $V_\xi(\boldsymbol{s}, \mathbb{T})$, shown in Figure 9 for Model 1 and in Supplement B.5 for all models. It is clear that the random field component $\xi(\boldsymbol{s})$ captures information not available in the SST components.

The full space–time averages $V_\eta(\Omega, \mathbb{T})$, $V_\beta(\Omega, \mathbb{T})$, and $V_\xi(\Omega, \mathbb{T})$ are the variances when probing the posterior distributions at a uniformly chosen random locations on $\Omega \times \mathbb{T}$. The remainder term $C_{\beta,\xi}(\Omega, \mathbb{T})$ is the posterior covariance between the fixed effect and random field contributions to the variability. We also define the correlation $\rho_{\beta,\xi}(\Omega, \mathbb{T}) = C_{\beta,\xi}(\Omega, \mathbb{T})/\sqrt{V_\beta(\Omega, \mathbb{T})V_\xi(\Omega, \mathbb{T})}$. A large negative value for the covariance or correlation indicates confounding. Table 2 shows the space–time averages, covariance, and correlation for the three models. The correlations are not very small, suggesting that there is some confounding, although it is not severe. While these diagnostics do not give direct guidance for model selection, they highlight the clear contribution of the random field component of each model.

6. Discussion. Unlike previous methods used to analyse these and similar survey data, our spatio-temporal point process model preserves the sighting locations, models the effect of explanatory variables continuously in space, and models spatial correlation that cannot be explained by such variables. It generalizes the approach of Johnson, Laake and Ver Hoef (2010), which models density as a nonhomogeneous Poisson process, using actual sighting locations, but neglecting residual spatial correlation. Unlike Johnson, Laake and Ver Hoef (2010), we model residual spatial intensity. It also generalises the approach of Pardo et al. (2015), who included a model for residual spatial intensity in their analysis of ETP blue

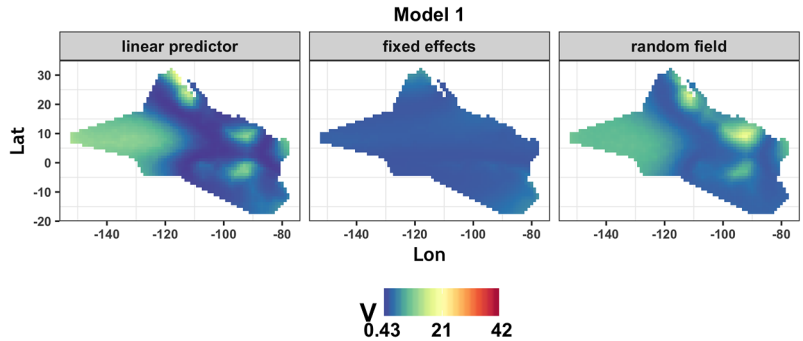

FIG. 9. *Variability measures V_η, V_β, and V_ξ for Model* 1.

whale data, but with no spatial structure on their residual model. We found substantial evidence for residual spatial structure in our analysis

It is rarely the case that spatial data are independent, and assuming independence when data are dependent can lead to biased variance estimation, spurious significance of covariates, and overfitting [Cressie (1993), Hanks et al. (2015)]. Use of a GMRF allows us to model spatially autocorrelated random effects, and model patterns in residuals that cannot be explained by available covariates. As shown in Section 5.3, the spatial pattern captured by the GMRF in Figure 7 plays an important role in estimating the spatial distribution of ETP blue whale groups, shown in Figure 6. Because the underlying mechanisms that dictate the distribution of blue whales in space and time are probably quite complex, it is unlikely that SST alone could adequately explain the distribution, so that drawing inferences about the effect of SST based on a model without modelling spatial correlation may result in misleading biological interpretations.

The analysis of Pardo et al. (2015) modelled blue whale density spatially as a function of absolute dynamic topography (ADT), which, like SST, predicted fewer blue whales in warmer regions. Because the model did not separate the temporal and spatial effects of ADT, large changes in ETP blue whale abundance were predicted from year to year, with few whales in warm (El Niño) years and many whales in cool (La Niña) years. Because blue whales have long life spans and reproduce slowly, and because tagging has shown that blue whales migrate to tropical waters every year, regardless of El Niño variations [see tracks in Bailey et al. (2009), for example], high interannual variation in true abundance seems unlikely. The hierarchical centering scheme in Section 5.2 separates the temporal and spatial effects of SST and accommodates situations in which whales make choices about habitat use *relative* to the other choices available to them, and this leads to what is arguably a more biologically plausible model with less interannual variation. While our estimates are consistent with true blue whale group density being unchanged over the surveys, point estimates of this density do tend to be higher in warmer years. This is unexpected and warrants further investigation.

Pardo et al. (2015) argue that ADT is a better predictor of blue whale density than SST because ADT contains information about subsurface as well as surface conditions. Notwithstanding this, our model is able to pick up structure in the data beyond that which can be attributed to SST. For example, Pardo et al. (2015) predict high densities on the Costa Rica Dome (approximately 10°N, 90°E) on the basis of ADT; we do the same by means of the GMRF (see Figure 7) even though SST does not suggest high densities here. By also modelling spatial autocorrelation, our model does not run the risk of drawing biased inference about the effects of explanatory variables (SST here) due to unmodelled correlation. We found that the estimated Gaussian random field is somewhat correlated with the fixed effects associated with SST. As a result, the interpretation of the fixed effects is not as clear cut as it would be were the Gaussian random field and fixed effects independent.

Considering our models in the more general context of point process modelling, the data structure we consider here differs from the point patterns typically analyzed in the point process literature [but see Waagepetersen and Schweder (2006)]. These usually comprise a point pattern that has been observed completely in a finite observation window that is a subset of $\mathbb{R}^2$, say. Unless finite point processes are explicitly considered, the standard assumption is that the point process continues in the same way outside the observation window. For interpretation, this implies that the analysis is only informative if the processes of interest are operating at a spatial scale that is captured within the (frequently single) subsample that is available. Further, there is an additional assumption that every point in the observation window has been observed, so that the detection probability is one within the observation window and zero elsewhere.

Our method extends such methods to deal with situations in which the processes of interest reflected in a spatial pattern, such as habitat preference, operates at a larger spatial scale than the sampled regions, when it may be impossible to fully sample an area that captures that scale. It also accommodates situations in which detection probability is unknown and not one, even within the sampled region. In wildlife sampling literature, this has often been dealt with in two stages, first estimating detection probability and then estimating spatial distribution conditional on the estimated detection probability. Our approach integrates the two, estimating detection probability simultaneously with the point process parameters.

We expect to see advances in spatio-temporal inference when there are covariates that affect both the thinning process and the density surface [Dorazio (2012)]. We also expect further development of methods to assess goodness of fit, as such methods are somewhat lacking for spatial and spatio-temporal inference.

The point process model in Section 3 can also be extended to a marked point process model to incorporate group size in the model and allow detection probability to depend on group size. We also anticipate that our approach will be extended to deal with more complex observation processes and for other survey types—for spatial capture-recapture sampling [Borchers and Efford (2008), Royle and Young (2008)], for example, for situations in which detection probabilities change over time, or when there is unknown spatially varying sampling effort.

Acknowledgements. We thank the captains, crews and observers on the NOAA research vessels, and the support staff at the Southwest Fisheries Science Center, for the collection of line-transect data in the ETP over many years.

SUPPLEMENTARY MATERIAL

Theory and analysis details (DOI: 10.1214/17-AOAS1078SUPP; .pdf). Supplement A gives theory for Sections 3.3 and 4.2. Supplement B gives details for the analysis in Section 5.

REFERENCES

BAILEY, H. B., MATE, B. R., PALACIOS, D. M., IRVINE, L., BOGRAD, S. J. and COSTA, D. P. (2009). Behavioural estimation of blue whale movements in the Northeast Pacific from state–space model analysis of satellite tracks. *Endanger. Species Res*. **10** 93–106.

BORCHERS, D. L. and EFFORD, M. G. (2008). Spatially explicit maximum likelihood methods for capture–recapture studies. *Biometrics* **64** 377–385. MR2432407

BORCHERS, D. L., BUCKLAND, S. T., GOEDHART, P. W., CLARKE, E. D. and HEDLEY, S. L. (1998). Horvitz–Thompson estimators for double-platform line transect surveys. *Biometrics* **54** 1221–1237.

BRENNER, S. C. and SCOTT, L. R. (2008). *The Mathematical Theory of Finite Element Methods*, 3rd ed. *Texts in Applied Mathematics* **15**. Springer, New York. MR2373954

BUCKLAND, S. T., OEDEKOVEN, C. S. and BORCHERS, D. L. (2016). Model-based distance sampling. *J. Agric. Biol. Environ. Stat*. **21** 58–75. MR3459294

BUCKLAND, S. T., ANDERSON, D. R., BURNHAM, K. P., LAAKE, J. L., BORCHERS, D. L. and THOMAS, L. (2001). *Introduction to Distance Sampling*: *Estimating Abundance of Biological Populations*, 1st ed. Oxford Univ. Press, Oxford.

BUCKLAND, S. T., REXSTAD, E. A., MARQUES, C. S. and OEDEKOVEN, C. S. (2015). *Distance Sampling*: *Methods and Applications*. Springer, Berlin.

CAMELETTI, M., LINDGREN, F., SIMPSON, D. and RUE, H. (2013). Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Adv. Stat. Anal*. **97** 109–131. MR3045763

CONN, P. B., LAAKE, J. L. and JOHNSON, D. S. (2012). A hierarchical modeling framework for multiple observer transect surveys. *PLoS ONE* **7** e42294.

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*. Wiley, New York. MR1239641

DIGGLE, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*, 2nd ed. Hodder Arnold, London.

DORAZIO, R. M. (2012). Predicting the geographic distribution of a species from presence-only data subject to detection errors. *Biometrics* **68** 1303–1312. MR3040037

FARIN, G. E. (2002). *Curves and Surfaces for CAGD*: *A Practical Guide*, 5th ed. Academic Press, New York.

FORNEY, K. A., FERGUSON, M. C., BECKER, E. A., FIEDLER, P. C., REDFERN, J. V., BARLOW, J., VILCHIS, I. L. and BALLANCE, L. T. (2012). Habitat-based spatial models of cetacean density in the eastern Pacific Ocean. *Endanger. Species Res*. **16** 113–133.

GERRODETTE, T. and FORCADA, J. (2005). Non-recovery of two spotted and spinner dolphin populations in the eastern tropical Pacific Ocean. *Mar. Ecol. Prog. Ser.* **291** 1–21.

GERRODETTE, T., PERRYMAN, W. and BARLOW, J. (2002). Calibrating group size estimates of dolphins in the Eastern Tropical Pacific Ocean. Administrative Report LJ-02-08. 20 p.

HANKS, E. M., SCHLIEP, E. M., HOOTEN, M. B. and HOETING, J. A. (2015). Restricted spatial regression in practice: Geostatistical models, confounding, and robustness under model misspecification. *Environmetrics* **26** 243–254. MR3340961

HAYES, R. J. and BUCKLAND, S. T. (1983). Radial-distance models for the line-transect method. *Biometrics* **39(1)** 29–42.

HEDLEY, S. L. and BUCKLAND, S. T. (2004). Spatial models for line transect sampling. *J. Agric. Biol. Environ. Stat.* **9** 181–199.

HEDLEY, S. L., BUCKLAND, S. T. and BORCHERS, D. L. (2004). Spatial distance sampling models. In *Advanced Distance Sampling* (S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers and L. Thomas, eds.) 48–70. Oxford Univ. Press, Oxford.

HEFLEY, T. J. and HOOTEN, M. B. (2016). Hierarchical species distribution models. *Curr. Landsc. Ecol. Rep.* **1** 87–97.

HÖGMANDER, H. (1991). A random field approach to transect counts of wildlife populations. *Biom. J.* **33** 1013–1023. MR1201875

HÖGMANDER, H. (1995). *Methods of Spatial Statistics in Monitoring Wildlife Populations*. Univ. Jyväskylä, Jyväskylä.

ILLIAN, J. B., SØRBYE, S. H. and RUE, H. (2012). A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *Ann. Appl. Stat.* **6** 1499–1530. MR3058673

ILLIAN, J., PENTTINEN, A., STOYAN, H. and STOYAN, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. Wiley, Chichester. MR2384630

JOHNSON, D. S., HOOTEN, M. B. and KUHN, C. E. (2013). Estimating animal resource selection from telemetry data using point process models. *J. Anim. Ecol.* **83** 1155–1164.

JOHNSON, D. S., LAAKE, J. L. and VER HOEF, J. M. (2010). A model-based approach for making ecological inference from distance sampling data. *Biometrics* **66** 310–318. MR2756719

JOHNSON, D. S., LAAKE, J. L. and VER HOEF, J. M. (2014). DSpat: Spatial modelling for distance sampling data. R package version 0.1.6.

KINZEY, D., OLSON, P. and GERRODETTE, T. (2000). Marine Mammal data collection procedures on research ship line-transect surveys by the Southwest Fisheries Science Center. Technical Report National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Southwest Fisheries Science Center.

KREFT, I. G. G., DE LEEUW, J. and AIKEN, L. S. (1995). The effect of different forms of centering in hierarchicallinear models. *Multivar. Behav. Res.* **30** 1–21.

LINDGREN, F. and RUE, H. (2015). Bayesian spatial modelling with R-INLA. *J. Stat. Softw.* **63(19)**.

LINDGREN, F., RUE, H. and LINDSTRÖM, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 423–498. MR2853727

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 Ecol. Evol.* **4** 1001–1010.

MILLER, D. L., REXSTAD, E. A., BURT, M. L., BRAVINGTON, M. V. and HEDLEY, S. L. (2014). dsm: Density surface modelling of distance sampling data. R package version 2.2.5.

MØLLER, J., SYVERSVEEN, A. R. and WAAGEPETERSEN, R. P. (1998). Log Gaussian Cox processes. *Scand. J. Stat.* **25** 451–482. MR1650019

MØLLER, J. and WAAGEPETERSEN, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. *Monographs on Statistics and Applied Probability* **100**. Chapman & Hall/CRC, Boca Raton, FL. MR2004226

MØLLER, J. and WAAGEPETERSEN, R. P. (2007). Modern statistics for spatial point processes. *Scand. J. Stat.* **34** 643–684. MR2392447

MOORE, J. E. and BARLOW, J. (2011). Bayesian state–space model of fin whale abundance trends from a 1991–2008 time series of line-transect surveys in the California Current. *J. Appl. Ecol.* **48** 1195–1205.

NIEMI, A. and FERNÁNDEZ, C. (2010). Bayesian spatial point process modeling of line transect data. *J. Agric. Biol. Environ. Stat.* **15** 327–345. MR2787262

OEDEKOVEN, C. S., LAAKE, J. L. and SKAUG, H. J. (2015). Distance sampling with a random scale detection function. *Environ. Ecol. Stat.* **22** 725–737. MR3424098

OEDEKOVEN, C. S., BUCKLAND, S. T., MACKENZIE, M. L., EVANS, K. O. and BURGER, L. W. (2013). Improving distance sampling: Accounting for covariates and non-independency between sampled sites. *J. Appl. Ecol.* **50** 786–793.

OEDEKOVEN, C. S., BUCKLAND, S. T., MACKENZIE, M. L., KING, R., EVANS, K. O. and BURGER, L. W. JR. (2014). Bayesian methods for hierarchical distance sampling models. *J. Agric. Biol. Environ. Stat.* **19** 219–239. MR3257912

PARDO, M. A., GERRODETTE, T., BEIER, E., GENDRON, D., FORNEY, K. A., CHIVERS, S. J., BARLOW, J. and PALACIOS, D. M. (2015). Inferring cetacean population densities from the absolute dynamic topography of the ocean in a hierarchical Bayesian framework. *PLoS ONE* **10** e0120727. DOI:doi:10.1371/journal.pone.0120727.

RAMSAY, J. O. (1988). Monotone regression splines in action. *Statist. Sci.* **3** 425–441.

ROYLE, J. A., DAWSON, D. K. and BATES, S. (2004). Modeling abundance effects in distance sampling. *Ecology* **85** 1591–1597.

ROYLE, J. A. and DORAZIO, R. M. (2008). *Hierarchical Modelling and Inference in Ecology*. Academic Press, London, UK.

ROYLE, J. A. and YOUNG, K. V. (2008). A hierarchical model for spatial capture–recapture data. *Ecology* **89** 2281–2289.

RUE, H. and HELD, L. (2005). *Gaussian Markov Random Fields*: *Theory and Applications*. *Monographs on Statistics and Applied Probability* **104**. Chapman & Hall/CRC, Boca Raton, FL. MR2130347

RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392. MR2649602

SCHMIDT, J. H., RATTENBURY, K. L., LAWLER, J. P. and MACCLUSKIE, M. C. (2012). Using distance sampling and hierarchical models to improve estimates of Dall's sheep abundance. *J. Wildl. Manag.* **76(2)** 317–327.

SIMPSON, D., LINDGREN, F. and RUE, H. (2012). Think continuous: Markovian Gaussian models in spatial statistics. *Spat. Stat.* **1** 16–29.

SIMPSON, D., ILLIAN, J. B., LINDGREN, F., SØRBYE, S. H. and RUE, H. (2016). Going off grid: Computationally efficient inference for log-Gaussian Cox processes. *Biometrika* **103** 49–70. MR3465821

STOYAN, D. (1982). A remark on the line transect method. *Biom. J.* **24** 191–195. MR0662894

STOYAN, D. and GRABARNIK, P. (1991). Second-order characteristics for stochastic structures connected with Gibbs point processes. *Math. Nachr.* **151** 95–100. MR1121200

VAN LIESHOUT, M. N. M. (2000). *Markov Point Processes and Their Applications*. Imperial College Press, London. MR1789230

WAAGEPETERSEN, R. and SCHWEDER, T. (2006). Likelihood-based inference for clustered line transect data. *J. Agric. Biol. Environ. Stat.* **11(3)** 264–279.

WIEGAND, T. and MOLONEY, K. A. (2014). *Handbook of Spatial Point-Pattern Analysis in Ecology*. CRC Press, Boca Raton, FL. MR3155590

WILLIAMS, R., HEDLEY, S. L., BRANCH, T. A., BRAVINGTON, M. V., ZERBINI, A. N. and FINDLAY, K. P. (2011). Chilean blue whales as a case study to illustrate methods to estimate abundance and evaluate conservation status of rare species. *Conserv. Biol.* **25** 526–535.

WOLTER, K. and TIMLIN, M. S. (2011). El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext). *Int. J. Climatol.* **31** 1074–1087.

WOOD, S. N. (2006). *Generalized Additive Models*: *An Introduction with R*. Chapman & Hall/CRC, Boca Raton, FL. MR2206355

YUAN, Y., BACHL, F. E., LINDGREN, F., BORCHERS, D. L., ILLIAN, J. B., BUCKLAND, S. T., RUE, H. and GERRODETTE, T. (2017). Supplement to "Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales." DOI:10.1214/17-AOAS1078SUPP.

Y. YUAN
D. L. BORCHERS
J. B. ILLIAN
S. T. BUCKLAND
SCHOOL OF MATHEMATICS AND STATISTICS
UNIVERSITY OF ST ANDREWS
THE OBSERVATORY, BUCHANAN GARDENS
ST ANDREWS, KY16 9LZ
UNITED KINGDOM
E-MAIL: yy84@st-andrews.ac.uk
dlb@st-andrews.ac.uk
jbi@st-andrews.ac.uk
steve@st-andrews.ac.uk

F. E. BACHL
F. LINDGREN
SCHOOL OF MATHEMATICS
UNIVERSITY OF EDINBURGH
JAMES CLERK MAXWELL BUILDING
THE KING'S BUILDINGS
PETER GUTHRIE TAIT ROAD
EDINBURGH, EH9 3FD
UNITED KINGDOM
E-MAIL: fabian.bachl@ed.ac.uk
finn.lindgren@ed.ac.uk

H. RUE
CEMSE DIVISION
KING ABDULLAH UNIVERSITY OF SCIENCE AND TECHNOLOGY
THUVAL 23955-6900
SAUDI ARABIA
E-MAIL: haavard.rue@kaust.edu.sa

T. GERRODETTE
SOUTHWEST FISHERIES SCIENCE CENTER
NOAA NATIONAL MARINE FISHERIES SERVICE
8901 LA JOLLA SHORES DRIVE
LA JOLLA, CALIFORNIA 92037
USA
E-MAIL: tim.gerrodette@noaa.gov