Comparing distribution of harbour porpoise using generalized additive models and hierarchical Bayesian models with integrated nested laplace approximation

Laura D. Williamson [a,b,c,*], Beth E. Scott [b], Megan Laxton [d], Janine B. Illian [d], Victoria L.G. Todd [a,c], Peter I. Miller [e], Kate L. Brookes [f]

[a] *Ocean Science Consulting Limited, Dunbar, Scotland, UK*
[b] *School of Biological Sciences, University of Aberdeen, Aberdeen, Scotland, UK*
[c] *Environmental Research Institute, University of the Highlands and Islands, Thurso, UK*
[d] *School of Mathematics and Statistics, University of Glasgow, Glasgow, UK*
[e] *Plymouth Marine Laboratory, Plymouth, England, UK*
[f] *Marine Scotland Science, Marine Laboratory, Aberdeen, Scotland, UK*

ARTICLE INFO

ABSTRACT

Species Distribution Models (SDMs) are used regularly to develop management strategies, but many modelling methods ignore the spatial nature of data. To address this, we compared fine-scale spatial distribution predictions of harbour porpoise (*Phocoena phocoena*) using empirical aerial-video-survey data collected along the east coast of Scotland in August and September 2010 and 2014. Incorporating environmental covariates that cover habitat preferences and prey proxies, we used a traditional (and commonly implemented) Generalized Additive Model (GAM), and two Hierarchical Bayesian Modelling (HBM) approaches using Integrated Nested Laplace Approximation (INLA) model-fitting methodology. One HBM-INLA modelled gridded space (similar to the GAM), and the other dealt more explicitly in continuous space using a Log-Gaussian Cox Process (LGCP).

Overall, predicted distributions in the three models were similar; however, HBMs had twice the level of certainty, showed much finer-scale patterns in porpoise distribution, and identified some areas of high relative density that were not apparent in the GAM. Spatial differences were due to how the two methods accounted for autocorrelation, spatial clustering of animals, and differences between modelling in discrete vs. continuous space; consequently, methods for spatial analyses likely depend on scale at which results, and certainty, are needed.

For large-scale analysis (>5–10 km resolution, e.g. initial impact assessment), there was little difference between results; however, insights into fine-scale (<1 km) distribution of porpoise from the HBM model using LGCP, while more computationally costly, offered potential benefits for refining conservation management or mitigation measures within offshore developments or protected areas.

1. Introduction

Understanding environmental and anthropogenic drivers of species distributions is critical for identification of potential areas for protection, development of appropriate conservation management strategies, and mitigation of negative anthropogenic impacts. Collection of empirical environmental data, however, can be risky, costly, and time consuming; consequently, desk-based species distribution modelling provides a cost-effective, yet successful alternative method to address these factors (e.g. Bailey and Thompson, 2010; Norberg et al., 2019).

Many statistical techniques are used to model species distributions, but due to inherent complexity of analysing space, most do not fully address the spatial nature of data. Generalised Additive Models (GAMs; Hastie and Tibshirani, 1990) are one of the most commonly used and

* Corresponding author.

E-mail address: lw@osc.co.uk (L.D. Williamson).

well-established methods of performing species distribution modelling (Becker et al., 2020; Redfern et al., 2006; Wood, 2006). Over the past decade, there has been rapid development of specific statistical methodologies aiming to address the spatial nature of empirical data (e.g. Cameletti et al., 2012; Diggle et al., 2013; Illian et al., 2008; Law et al., 2009; Sadykova et al., 2017); however, such methods are complex and require a high level of statistical expertise and computational power. Practitioners are, therefore, often faced with a choice of which model to apply to their data, which can have considerable impacts on outputs (Norberg et al., 2019).

The development of point process models in ecological contexts (Soriano-Redondo et al., 2019; Yuan et al., 2017a), by using exact locations of points in space within a GAM function are being explored by some studies (Fithian and Hastie, 2013; Miller et al., 2019; Renner et al., 2015). These use an inhomogeneous Poisson process, maximum entropy, or infinitely weighted logistic regression to model data in a point process framework thereby taking advantage of the spatial information; however, at the time of the production of this study, these techniques were still under development for generalist use, and were therefore not explored further here.

The aim of this study was to assess differences between use of traditional frequentist GAMs and explicit spatial-modelling approaches using Hierarchical Bayesian Models (HBM) fitted with Integrated Nested Laplace Approximation (INLA, Rue et al., 2009). To achieve this, we introduce the two methods (GAM and HBM-INLA) and identify three areas in which Bayesian developments offer potential benefits over traditional GAMs: (1) they account for spatial autocorrelation, (2) clustering of animals, and (3) model in continuous space. We then implement both methods using an example of real-world aerial-video survey data for harbour porpoise (*Phocoena phocoena*). Finally, we explore differences in resulting spatial predictions of porpoise distribution and conservation management implications.

1.1. Two approaches to species distribution modelling

1.1.1. Generalized additive models

Generalized Additive Models are popular for species distribution modelling because of their ability to encapsulate non-linear interactions between sightings and environmental covariates through use of smoothing functions (e.g. Booth et al., 2013; Embling et al., 2010; Williamson et al., 2016); however, their ability to capture spatial trends is limited without use of alternative, more complex additions (Fithian and Hastie, 2013; Miller et al., 2019; Renner et al., 2015; Scott-Hayward et al., 2014). Nonetheless, GAMs are relatively easy to implement by the user through various R libraries, such as `mgcv` (Wood, 2011), or similar. Counts y in a grid cell S_i, i=1,…N are modelled (response variable), and the model usually assumes that counts in each grid cell follow a Poisson distribution. The expected value in a grid cell s_i is modelled as a linear model of covariates x, i.e.,

$$g(y(s_i)) = x(s_i)^T \beta + \beta_0. \tag{1}$$

where, g is a log link function, $x(s_i)^T$ is a matrix of N observations on each of the k explanatory variables, and β is a vector of regression coefficients.

The GAM formulation in Eq. (1) does not address the problem of spatial autocorrelation, instead accounting for trends in data across larger geographical distances (Dormann et al., 2007). One common method for circumventing this problem is to include a bivariate smoothing term of latitude and longitude or a soap-film smoother (Wood et al., 2008) in the GAM itself. Soap-film *x,y* smoothers also include boundary information to prevent the model from predicting animals' presence in areas they are not capable of occupying (Wood et al., 2008), e.g. porpoises present on land. An *x,y* smoother accounts for similarity of the response variable in adjacent cells, reducing the effect (but not the amount) of spatial autocorrelation. The spatial

autocorrelation problem is solved by using more complex methods, such as including autocorrelation structures in a generalized additive mixed effects model (Carvalho et al., 2020).

1.1.2. Hierarchical Bayesian modelling using INLA

Recent developments in spatial statistical methodology have focused on producing a flexible, computationally efficient approach that is increasingly accessible to end users. Here, we used a continuously indexed Gauss-Markov random field based on the Stochastic Partial Differential Equation, SPDE (Lindgren et al., 2011). This provides an approximation to a continuous spatial field that allows users to work efficiently with complex spatially explicit models (Blangiardo et al., 2013). In addition, models were fitted based on INLA (Rue et al., 2009), a recently developed computationally efficient method for hierarchical Bayesian inference that is particularly suitable for complex spatial models, such as those based on the SPDE approach. Combining the two methods (SPDE and INLA) provides a suite of flexible models that allow for realistic spatial modelling within feasible computation timescales.

As in the GAM model, counts y in a grid cell s_i, i=1,…,N are modelled, but a Bayesian approach is used and spatial autocorrelation is modelled by a random field ξ, i.e.,

$$g(y(s_i)) = x(s_i)^T \beta + \beta_0 + \xi(s_i) \tag{2}$$

where, g is a log link function, and β is a vector of regression coefficients. In this HBM formula, the residual spatial autocorrelation is modelled explicitly by including the ξ term, which specifies a spatial or spatio-temporal random field (Rue and Held, 2005). This framework allows for smooth and nonlinear effects of covariates, time trends, seasonal effects, random intercepts and slopes, and spatio-temporal random effects; therefore, this class of model is highly flexible and can accommodate a wide range of paradigms.

1.2. Discrete vs. continuous space

Spatial data are usually gridded or segmented when using GAMs, which causes spatial information to be lost; however, spatially explicit approaches such as SPDE/INLA can model in continuous space. Since this implies much less information loss, SPDE/INLA approaches were expected to be more appropriate and capable of reflecting small-scale spatial behaviour that cannot be captured by gridding or segmenting sightings. Consequently, in this paper we made two comparisons. We initially used both approaches, GAM and HBM, to model data on the same spatial grid to establish how the approaches differed when the actual models were similar, with the expectation of similar results. We then moved towards modelling in continuous space using the SPDE/INLA approach that also allowed us to incorporate information on the exact location of transects, as well as the spatial structure formed by the locations of individual animals, i.e. their clustering (Yuan et al., 2017b).

1.3. Spatial clustering of animals

Animals do not always move independently. At times, some species can form large groups in small areas for several reasons, such as social cohesion or specific attraction to the location as a result of physical and/or environmental covariates. Plants and animals are often clustered spatially according to a Poisson process, which inappropriately assumes complete spatial randomness between clusters. To account for this, a Cox process (Møller and Waagepetersen, 2007; Waagepetersen and Schweder, 2006) is a flexible type of spatial point procedure which improves accuracy by modelling spatial patterns relative to observed or unobserved spatial trends (Illian et al., 2012). In this way, the spatial pattern can be modelled along with traits of individual observations such as markings, scars, size, or colour and environmental covariates (Illian et al., 2012). Since clustering violates Poisson process assumptions, this obliges the SPDE to capture low-range peaks in detection,

rather than a large-scale distribution trend. An alternative (and often implemented) approach is to use the negative binomial distribution to account for overdispersion which means that the SPDE does not have to accommodate spikes to capture the large-scale trend; however, while often used to avoid violation of model assumptions, it does not explicitly model overdispersion. Moreover, there is no obvious way of extending this to continuous space, so analysis is restricted to using gridded bins.

When using hierarchical Bayesian models, localised clustering can be differentiated from large-scale patterns using the Log-Gaussian Cox Process (LGCP) method which models exact locations of points (Diggle et al., 2013). The model assumes point locations are scattered independently in space as a Poisson process (top level of the hierarchy), given a model of the point density (formally the "intensity" of the process – the middle level of the hierarchy) and a set of priors for all model parameters (bottom level of the hierarchy). The log of intensity Λ in a location $s \in R$ is modelled as a linear model of covariates x and a random field ξ that is again approximated by an SPDE:

$$log(\Lambda(s;t)) = x(s)^T\beta + \beta_0 + \xi(s) \qquad (3)$$

1.4. Case study

Harbour porpoise is the most numerous cetacean species in European Atlantic Shelf waters (Hammond et al., 2021) which is also where most anthropogenic activities occur. Consequently, they are threatened by, inter alia, unsustainable incidental catches in fishing gear and prey removal, various forms of chemical and underwater noise pollution, and other anthropogenic pressures (e.g. Culik, 2004; Hammond et al., 2021; MacLeod et al., 2007). Knowledge of porpoise distribution in relation to environmental drivers is therefore critical for effective implementation of management and mitigation measures.

Depth, slope, bottom temperature, and sediment have been significant in previous models of porpoise distribution (e.g. Booth et al., 2013; Brookes et al., 2013; Sadykova et al., 2017; Williamson et al., 2016). Other variables included chlorophyll-*a*, net primary productivity, and fronts and tidal mixing (e.g. current speed, vertical shear, and potential energy anomaly to quantify stratification strength), which cause prey aggregation, and have been correlated previously with the distribution of fish and marine predators (e.g. Embling et al., 2012; Hofmann and Powell, 1998; Olson and Backus, 1985; Scott et al., 2010).

Previous distribution modelling in the Moray Firth, north-east Scotland (part of the current study area; Fig. 1) found a relatively high density of animals in the Smith Bank, a shallow (30–40 m) sand bank in the middle of the Firth (Brookes et al., 2013; Williamson et al., 2016). This is likely because the Smith Bank provides suitable habitat for sandeels (Ammodytidae) and whiting, *Merlangius merlangus* (Hopkins, 2011), two of the most common prey species for porpoise in this area (Santos and Pierce, 2003). Knowledge of porpoise distribution along the East coast of Scotland, however, is lacking. Consequently, the present analysis combines previously collected aerial-video-survey data with a newer dataset, which allowed us to investigate distribution along the entire east coast of Scotland.

Aerial video surveys use video cameras to record strip transects below the survey aircraft and are thus different to visual surveys by human observers (Buckland et al., 2012). Moreover, video surveys provide a permanent record of sightings, which can be beneficial for clarifying mis- or ambiguous identifications, anomalous results, improving data transparency, and audit; indeed, some regulatory authorities now require such a permanent record (BSH, 2013). Estimates of relative density of porpoise modelled from video-survey data have been found previously to correlate strongly with relative density estimated from visual surveys (Williamson et al., 2016).

1.5. Aims

The purpose of this study was to compare two modelling frameworks
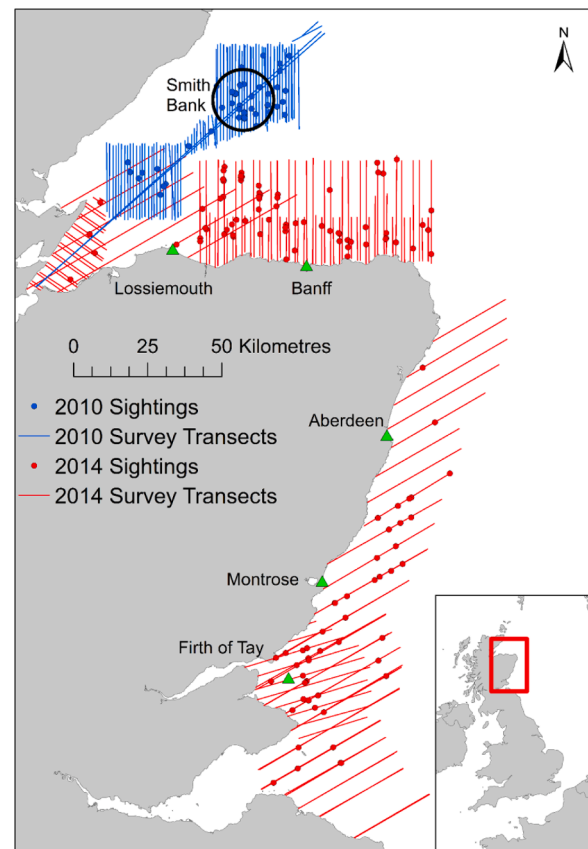

Fig. 1. Map of the study area showing locations of survey transects (lines) and sightings (dots) of harbour porpoise along the east coast of Scotland in 2010 and 2014. Inset shows the study area in relation to the British Isles.

to investigate differences between results obtained from a spatial-modelling technique (HBM using INLA) and a GAM which is used commonly by ecologists. We illustrated differences between these methods in how they model spatial autocorrelation, groups of individuals, and discrete vs. continuous space. We performed this using a case study of aerial-video transect-survey data to analyse harbour porpoise distribution along the East Coast of Scotland and discussed the potential management implications of results.

2. Methods

2.1. Case study data collection

Aerial digital video surveys covering a total of 5762 km were performed by HiDef Aerial Surveying Ltd. in August and September 2010 and 2014 along the east coast of Scotland (Fig. 1 and Table 1). Video data were processed by qualified, trained, and experienced observers, who extracted all non-avian objects for identification by specialists at WWT Consulting Ltd., resulting in a total of 303 individual porpoises observed.

Table 1
Details of digital video surveys performed in August and September 2010 and 2014.

	2010	2014	All
Survey effort (km)	2155	3607	5762
Number of porpoise	97	206	303
Survey height (m)	244–457	560	–
Strip width (m)	80–150	500	–

2.2. Data preparation for gridded analysis

Once porpoise observations were extracted, a 5×5 km grid, matching previous marine mammal work (Jones et al., 2015), was created which covered the entire survey area. Transect lines were intersected to the grid and the total length of transects and number of porpoises sighted within each grid cell was calculated.

Depth on a raster grid of approximately 180×180 m and polygons of sediment type at a 1:250,000 scale were provided by SeaZone Solutions Ltd. (2005a; 2005b). Sediment types were classified using a Folk triangle (Folk, 1954), and expressed as the proportion of sediment that was sand or gravelly sand within each grid cell, based on previous studies of porpoise habitat association in this area (Brookes et al., 2013; Williamson et al., 2016). Seabed slope was calculated in degrees using the Slope tool in ESRI ArcGIS 10.2.1 (ESRI, Redlands, California, USA).

Modelled environmental data were provided from runs of the NEMO-ERSEM 3D coupled hydrodynamic-ecosystem model (Nucleus for European Modelling of the Ocean; Edwards et al., 2012; Madec, 2008; O'Dea et al., 2012). Parameters available included: bottom temperature (ºC), potential energy anomaly (J/m^3; which is the energy required to mix the water column completely), depth-averaged current speed (m/s), vertical velocity shear (m/s), depth-averaged vertical velocity (m/day), depth-integrated net primary production (tons/day), and maximum chlorophyll-a (mgC/m^3). These data were available for the "summer season" (July, August, September, and October), as climatological means across 25 years (1989–2014) at a 7×7 km spatial scale. An estimate of stratification was calculated using the $\log_{10}$ of depth (H) divided by current speed (u) cubed ($\log_{10}(H/u^3)$; Simpson et al. (1981)). In addition, satellite data on a 1×1 km spatial scale were obtained from NEODAAS (NERC Earth Observation Data Acquisition and Analysis Service) and processed to derive ocean-front metrics (Miller et al., 2015): distance to the closest major ocean front (km), and front side - whether a location is on the cold (negative values) or warm (positive values) side of the closest major front, or directly on that front (zero). These satellite data were provided as weekly means during the survey period; however, due to cloud cover obscuring parts of the study area each week, the mean in August and September of 2010 and 2014 were calculated and joined to the survey data from the corresponding year. Where surveys overlapped between the two years, the mean of August and September in both years was used.

All environmental variables were up or down sampled (depending on spatial resolution) and joined to the 5×5 km grid for inclusion in the models. Variables were also standardised by subtracting the mean value. The above variables were selected not necessarily because they were thought to influence porpoise distribution directly, but often because they are thought to be drivers of the distribution of porpoise prey (e.g. Scott et al., 2010). After inspection of histograms of the data, cells with a vertical shear greater than 0.15 m/s and a front side of < -2 were removed, as data outside these ranges were sparse.

2.3. Data preparation for LGCP analysis

The same 5×5 km grid of environmental data was used for this analysis; however, number of sightings in each grid cell, and length of survey effort in each cell were not joined to this grid. Instead, the exact location of sightings, as well as transect start and end points were used directly.

2.4. GAM analysis

The first model was created using GAM. Only cells in which surveys were performed (e.g. survey effort >0 km) were included during model development. A 10×10 grid of points spaced evenly throughout the study region in which points that did not overlap with land were used for creation of the soap-film smoother (resulting in 27 internal knots; Fig. S1). Smoothers with more (up to 30×30), and fewer (5×5) knots

were also tested, but not found to influence results. Models were tested with the maximum likelihood method using families such as negative binomial, Poisson, and zero-inflated Poisson. The negative binomial likelihood was selected for analysis after inspection of summary plots of models. In every model, an offset of the log of the area surveyed was included. The area surveyed was calculated by multiplying the effort (length surveyed in each cell in km) by the strip width (width of sea surface in view of the camera in km).

All environmental variables were tested for collinearity with each other, and those with a correlation of 0.5 or greater were not included in the same models. After initial investigation, relationships between most covariates and sightings were found to be linear; therefore, for simplicity all covariates were treated as linear in the model except for the soap-film smoother. Backwards selection was used to select the best model, in which a full model was fitted which included all environmental covariates except the soap-film smoother (e.g. without the influence of space), and then each environmental covariate was removed sequentially and the model re-run. This process was repeated consecutively using the model which had the lowest Akaike Information Criterion (AIC; Akaike, 1973) until there was no-longer a decrease in AIC when further covariates were removed. An Auto Correlation Function (ACF) plot was used to check for temporal autocorrelation, which was not found to exist; therefore, use of mixed models was not investigated. The selected model was run with the soap-film smoother and then used to predict relative density of porpoise in all cells within 10 km of the survey transects (Fig. 2). A variogram was created to investigate if spatial autocorrelation was evident in the model (Diggle and Ribeiro, 2007), which was not found to be an issue (Fig. S5). All analysis was performed in R version 3.2.4 and GAMs were modelled using the mgcv package (Wood, 2011, 2017). R code can be seen in Appendix 1 and is available on GitHub.

2.5. Hierarchical Bayesian analysis using INLA

Two models were created using hierarchical Bayesian methods fitted using INLA. Both models were fitted using the inlabru version 2.2.4.9000 and INLA 20.03.17 packages in R version 3.4.1 (Bachl et al., 2019; R Core Team, 2017; Rue et al., 2009). The first model used the same gridded data that were used in the GAM analysis with the addition of the SPDE to account for spatial autocorrelation (Lindgren et al., 2011). The SPDE approach uses a Matérn covariance function to approximate the Gaussian field using a flexible stochastic model which provides a continuous relationship between the response and explanatory variables throughout the study area. Similar to the knots in the soap-film smoother in GAM, the SPDE approach requires a triangulation mesh of the modelled area. The mesh provides a lower bound on the spatial resolution for analysis; therefore, a mesh should be developed which is fine enough so that no further changes in the results can be observed when a finer mesh is used (Lindgren et al., 2011). The triangulation mesh for the SPDE was created bounded by the coastline and the boundary of the study region and allowed to place vertices randomly as needed (Fig. S2). Default uninformative prior specification (Gaussian with mean of 0 and variance of 100) was used for all variables. The same effort offset that was used in the GAM was included here (log of transect length * strip width) and again, environmental covariates were treated as linear. A negative binomial likelihood was selected to address clustering of the animals and to be able to compare more closely with the GAM results. Model selection (excluding the SPDE) was performed in the same way as described for the GAM; however, a Widely Applicable Information Criterion (WAIC) – also known as Watanabe AIC (Watanabe, 2010) – was used instead of AIC. This selected model was run once more with the SPDE included and was then used to predict relative density throughout the study area (Fig. 2).

The second, continuous space model created using HBM used the point process likelihood (Simpson et al., 2016) which has been adapted to line transect surveys (Yuan et al., 2017b). This approach represents
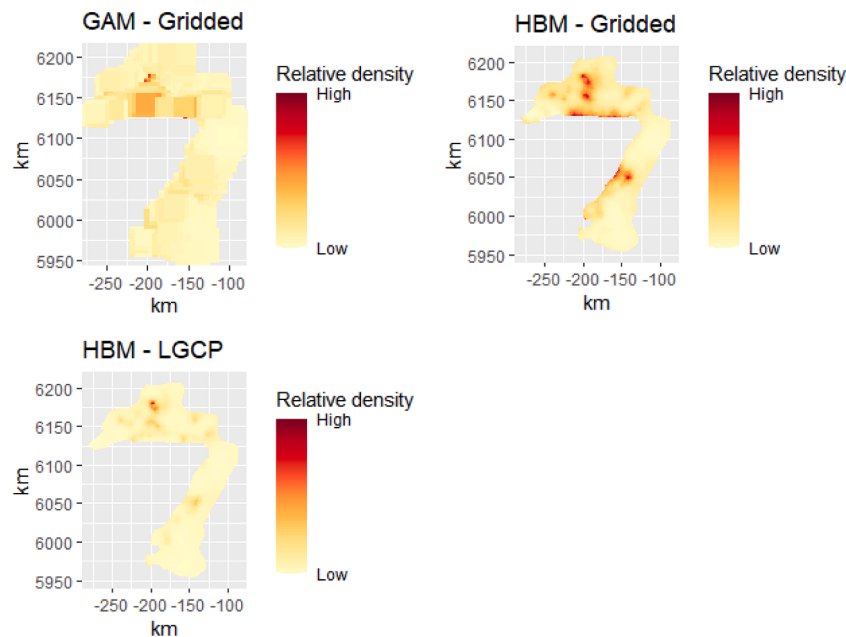
Fig. 2. Predicted relative porpoise density for the best GAM model (top left), HBM of gridded data (top right), and HBM using the LGCP (bottom left).

the locations of animal sightings as realisations of LGCP observed at the area covered by the transects. As opposed to the gridded approach, where data are aggregated, the LGCP allows the locations of observations and transects to be used directly. Taking the actual transect locations into account allows us to distinguish between areas where we did not find any animals because no survey was performed in that location, compared to those areas where we did not find anything, but they were also surveyed. This is especially useful when transects are not placed randomly throughout a study area. Information about the relative and absolute animal location is thereby retained, which allows for a higher precision in estimating the relationship between covariates and density, as well as the residual spatial correlation structure. The LGCP model was created using the same set of covariates and the SPDE described above. Similarly, backwards selection using WAIC was used based on the environmental covariates without the influence of space and once the best environmental variables were identified the SPDE was included, and relative density was predicted throughout the study area. R code used to create the HBM using INLA and the LGCP can be seen in Appendices 2 and 3.

Both HBMs were also predicted onto the same grid used in the GAM models to facilitate comparison between methods. Spearman's correlation coefficients were calculated between GAM and HBM results. Moreover, for visual comparison, results of each model were divided by their maximum value (to standardise) and HBM results were then subtracted from GAM results to investigate if GAM or HBMs predicted higher relative densities (Fig. 4).

3. Results

The negative binomial likelihood was selected for GAM analysis after inspection of summary plots of the models (Fig. S3). A negative binomial likelihood was also selected for the HBMs using INLA on gridded data for comparison to the GAM. The best models selected using each of the three methods included many of the same covariates (Table 2); however, potential energy anomaly, net primary productivity, vertical shear, and front distance were not retained in any of the best models. This best GAM model had a soap-film smooth term that included latitude, longitude, and the boundaries of the study region to account for spatial autocorrelation. Spatial correlation was not included as a random effect in this model because a variogram of model results did not show any

Table 2
Comparison of GAM and HBM results from models with the lowest AIC. A * indicates the variable was significant – for the GAM this means $p < 0.05$, and for HBM that the 95% CI did not cross 0.

	GAM	HBM - gridded	HBM - LGCP
Environmental variables	Estimate ± SD	Mean ± SD	Mean ± SD
Intercept	−8.49 ± 253.06	−2.68 ± 0.36 *	−1.48 ± 22.36
Depth (m)	0.02 ± 0.20 *	0.01 ± 0.01	NA
Mixing (H/u³)	0.55 ± 8.27	0.69 ± 0.39	0.16 ± 0.32
Chlorophyll-a (mgC/m³)	−0.06 ± 1.05	−0.07 ± 0.08	−0.01 ± 0.07
Bottom temperature (ºC)	0.53 ± 19.44	NA	NA
Current speed (m/s)	27.04 ± 404.38	NA	NA
Vertical shear (m/s)	−6.29 ± 297.20	NA	−5.87 ± 8.01
Thermal front side (unitless)	0.92 ± 8.21 *	NA	NA
Slope (º)	0.63 ± 6.84	NA	0.48 ± 0.41
Proportion of sediment that was sand or gravelly sand (%)	NA	NA	−1.48 ± 22.36

relationship with distance (Fig. S5). This model predicted the highest relative density of porpoise on the Smith Bank, and an area of high density along the southern coast of the Moray Firth. This model explained 33.2% of deviance (Fig. 2 and S4).

HBM models used the SPDE approach to account for spatial autocorrelation. When comparing the HBM of gridded data with the GAM, the same areas of highest density were apparent with an additional third area of high density along the east coast of Scotland (Fig. 2). In addition, much finer-scale clustering was apparent in the HBM results. The predicted distribution of porpoise from the GAM showed broad areas of high density. This contrasted with the HBM on a grid, particularly in the southern half of the study area where the area of high density was much more restricted than in the GAM. Comparing the HBM using LGCP (Fig. 2) with the GAM, we again saw much finer localised clustering,

with only the Smith Bank highlighted as a high-density area. This suggested that the model was identifying areas where animals were either found in groups, or they used very fine-scale areas (1–2 km), which were not apparent in the gridded GAM analysis.

Maximum coefficients of variation for the models fitted using HBMs were much lower than those created using GAM (GAM CV = 33, HBM - gridded CV = 8, HBM - LGCP CV = 11, Fig. 3), which showed higher confidence in the models. The influence of environmental covariates on models can be seen in Figs S4, S7, and S8. Effects of environmental covariates were similar between the three models, with the sign of each variable being the same even though the exact numbers varied (Table 2).

The GAM also had quite large areas with high Coefficient of Variation (CV), whereas both HBMs had a maximum CV less than half that of the GAM, but also the highest areas of CV in the HBMs were very localised and only at the edges of the study area in areas with very few data (Fig. 3). Correlation with GAM results was higher for the HBM of gridded data (Spearman's ρ = 0.76) than for the HBM using LGCP (Spearman's ρ = 0.68). The GAM consistently predicted higher relative porpoise density just south of the Smith Bank and both HBMs predicted higher density on the Smith Bank and along the east coast (Fig. 4).

Scatterplots were created to investigate how well predicted relative densities matched observed data (number of sightings divided by the survey effort; Fig. S6). Correlation between these ranged between 0.09 and 0.51.

3.1. Discussion

All three models had similar predicted distributions of porpoise (Fig. 2), with important differences. Firstly, compared to GAMS, HBMs (particularly when using LGCP), revealed more fine-scale areas of porpoise higher relative density, with an additional fine-scale (<5 km) area, that was not identified in GAMs. Secondly, compared to GAMS, HBM LGCP predicted a more realistic patchier porpoise distribution. This trend was not apparent in the other two models, which used the negative binomial distribution and gridded data, which smoothed out the highest-density areas. Finally, the best models from each of these three techniques included tidal mixing and chlorophyll-*a*, while two models also included depth, slope, and vertical shear (Table 2), suggesting that

these may be key drivers of porpoise distribution (Brookes et al., 2013; Embling et al., 2012; Sadykova et al., 2017; Scott et al., 2010; Williamson et al., 2016).

In terms of relating these model outputs to real porpoise ecology, animals in the field tended to be observed singly or in small groups of 2–3 animals. Clustering observed in models can likely be explained in terms of animal use of fine-scale environmental features that influenced prey availability (Table 2; e.g. Scott et al., 2010; Williamson et al., 2017), wider social factors – such as cooperative hunting behaviour (Ortiz et al., 2021) – or simply porpoise group spacing on survey sampling days. In terms of oceanography, thermal fronts are fine-scale, transient features which aggregate fish (Hofmann and Powell, 1998; Olson and Backus, 1985), and consequently have been related significantly to porpoise and other cetacean detections (e.g. Mendes et al., 2002; Philpott, 2013). Finally, the scale at which modelling was performed may have Influenced the impact/importance of variables between models (Mod et al., 2020).

3.1.1. Variability

Compared to GAM, variability for both HBMs was less than half, showing twice the level of certainty in the results. Some values of high CV in the GAM were in areas where there were survey data, whereas in the HBMs, areas of highest CV were restricted to the perimeter of the study area, where there was no survey effort. The GAM predicted much larger areas (25+ km) of high density than HBMs because the GAM smoothed out some spatial trends by gridding the data and using the negative binomial distribution, whereas the HBMs took the spatiality explicit autocorrelation into account.

3.1.2. Discrete vs. continuous space

The continuous modelled data using the SPDE/INLA, offered clear advantages because it eliminated information loss associated with data gridding, as performed in traditional GAMs. This is expected to be similar when segmenting data and can result in more accurate predictions of distribution. Both the HBM of gridded data (analysed in continuous space), and the HBM using LGCP (which used exact locations of sightings and transects) have smaller areas of high density (5 km) with a patchier distribution.
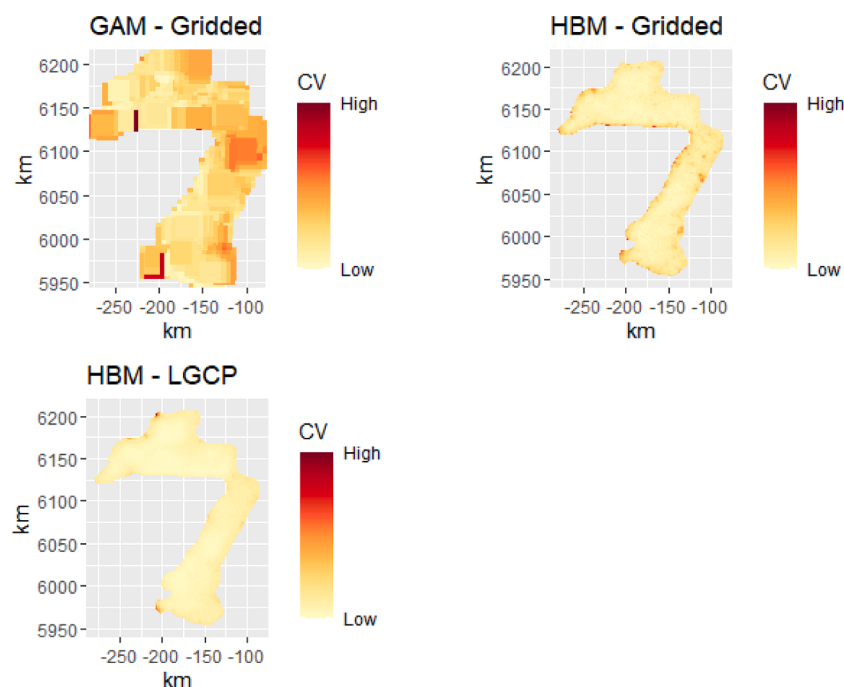

Fig. 3. Coefficient of variation (CV) for the best GAM model (top left), HBM of gridded data (top right) and HBM using the LGCP (bottom left).
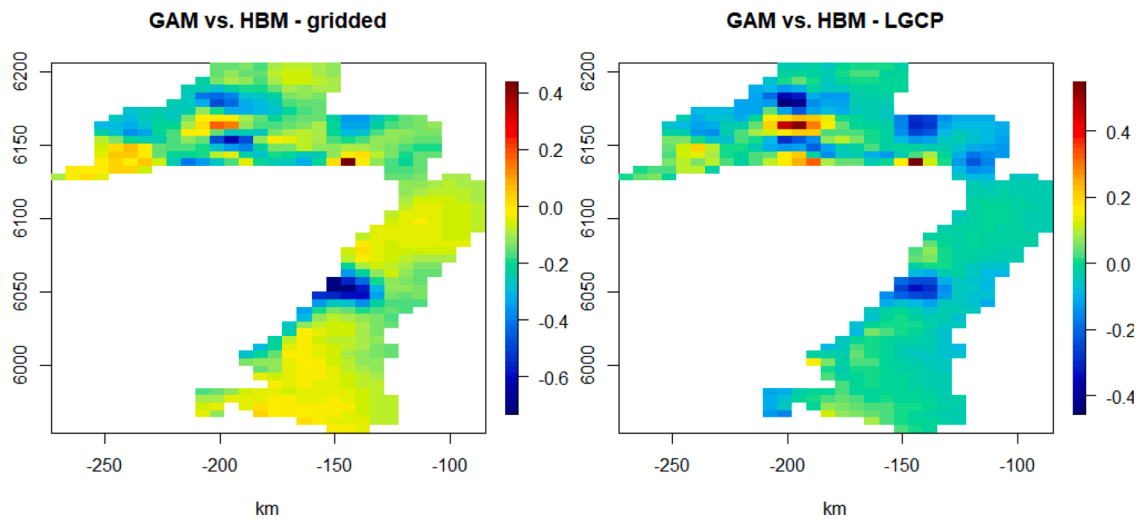
Fig. 4. Comparison of GAM and HBM results. Positive (red) indicates GAM predicted higher relative density, negative (blue) indicates GAM predicted lower density and zero is the same density. GAM vs. HBM of gridded data is on the left and GAM vs. HBM using LGCP is on the right.

In general, the continuous representation and triangulation using the mesh provided a more flexible way of approximating a spatial area. This becomes relevant both when the observation area is complex (with a lack of data, *etc.*), and if the study area represents a large area of the earth's surface and a fixed grid distorts special relationships. In this case, models can be defined on the sphere without the need to project into two-dimensional space and the resulting distortion, which is inevitable when working with a grid.

3.1.3. Spatial clustering

Results of the HBM using LGCP showed a slightly different pattern to either of the other models fitted to the gridded data. This LGCP model showed more localised areas of high density, with overall low density everywhere else. Again, this was because animals were observed in clusters. Using the negative binomial likelihood (as in both the GAM and HBM using gridded data) is an arbitrary fix for this grouping problem, which smooths data within the grid, but there is not a straightforward interpretation of how the negative binomial likelihood reflects overdispersion. Consequently, if appropriate for the data and species being studied, using Bayesian modelling with the LGCP provides the best advantage because it does not smooth out patchy areas; however, porpoise is a highly mobile species, and performing the same survey on a different day would likely give different results, especially because the animals' distribution is related strongly to environmental covariates, which can change on timescales of minutes. With repeated surveys, the LGCP approach offers potential to differentiate between fine-scale clustering caused by grouping of animals, and large-scale clustering caused by the underlying distribution pattern, but it was not possible to investigate this here due to a lack of data from repeat surveys. Nonetheless, the exercise based on survey data available, produced a much clearer and more flexible approach than using the negative binomial likelihood, but was not as straightforward to implement when using GAMs. Without several repeated surveys, it is unclear if these very fine-scale trends reflect the true distribution of porpoise. A single survey of this type might be appropriate for sessile organisms, but this appears to be modelling at a finer resolution than is realistic for data from a single survey of mobile organisms.

Relatively low correlation was observed between predicted relative densities and observed data (Fig. S6). A potential reason for this arises from a spatial mismatch between where surveys or sightings fell within the 5 × 5 km grid cell and the centre of the cell where predictions were calculated. Also, models considered information from neighbouring cells when calculating predictions, which could be increasing/decreasing predictions and would not be reflected by survey data within

a particular cell. In terms of relating this to empirical observations of porpoise spatial clustering of animals in the field, the situation is more complex. For example, in energetically active tidal flows, there are linkages between porpoise presence and small-scale heterogeneity amongst different environmental covariates (e.g., tidal phase, time of day) and significant spatiotemporal variability in site use at scales of hundreds of metres and hours (Benjamins et al., 2017).

3.1.4. Future developments

`Inlabru` (inlabru.org; Bachl et al., 2019), the R package used here to implement HBM using INLA (Yuan et al., 2017b), is still in development, but aims to make these techniques more accessible to users and expand versatility. `Inlabru` makes it easier to use data from transects, as well as actual locations of sightings using the LGCP method. Moreover, because `inlabru` was motivated originally by the need to incorporate distance sampling with HBM, it is possible to implement information relevant to sighting conditions and a detection function into modelling. This was not used here, as data analysed were strip transects in which the detection function was assumed to be uniform throughout the strip width; however, this could be a benefit for line-transect surveys (e.g. boat-based surveys for marine mammals or birds), where sighting conditions (e.g. seastate, visibility, glare, etc.) influence observations strongly.

Given that models fitted with `inlabru` are Bayesian, the user can include prior information about parameters. Choosing appropriate priors is still an area of ongoing research and the Penalised Complexity (PC) prior approach discussed in Simpson et al. (2017) and Sørbye et al. (2019) aims at making prior choice transparent and user-driven based on the distance and strength of spatial correlation. Default uninformative priors available in `inlabru` (and `R-INLA`) were used here, but different approaches to prior choice should be assessed in the future.

3.1.5. Practicalities of use

The increased computing time, power, statistical complexity, and high end-user capability of fitting models using HBM with INLA, combined with the fact that they are still under development by statisticians, has meant that their uptake within the ecological community has been slow. Moreover, these newer methods have limited user-friendly documentation available (especially regarding use of the `inlabru` package), which may further hinder implementation (Norberg et al., 2019). Nonetheless, compared to GAMS, models incorporating INLA methodology are more flexible and can assess additional complex relationships. Moreover, HBMs are capable of modelling complex spatio-temporal data and patterns, as well as modelling interspecific interactions and species

characteristics (e.g. joint species distribution modelling: Illian et al., 2012; Sadykova et al., 2017) which is not straight-forward in GAMs.

3.2. Management implications

In terms of translating differences between GAM vs. HBM-INLA into applied, real-world scenarios and case studies, this work has demonstrated that the loss of spatial detail in a traditional GAM, could have important implications for development of protected areas, or mitigation during fine-scale offshore operations such as pile driving, drilling, dredging, *etc.* For example, any elevated fine-scale clustering of porpoise at scales <1 km due to, inter alia localised bathymetric features (*e.g.* Smith Bank), adjacent to low-use areas, could be overlooked with a GAM, which might misinform any fine-scale designation of protected areas. Moreover, relevant to this study, for two offshore windfarms in the Moray Firth, GAMs were used to assess porpoise distribution at a minimum 4 × 4 km spatial resolution (Brookes et al., 2013; Williamson et al., 2016), which provided a quick overview of distribution or coarse estimates of relative density (e.g. for initial impact assessment or analysis at the scale of national jurisdictions); however, if fine-scale trends in animal distribution at the sub-km resolution was required to develop a plan of where to place the individual Moray Firth turbines, then LGCP analysis could have disentangled relationships between chance and feature-related clustering.

This study has demonstrated that modelling gridded data using a GAM and the negative binomial distribution highlighted two areas of high density within the Moray Firth, Scotland. Modelling the same gridded data using HBM-INLA (also with a negative binomial distribution) highlighted the same two areas, with an additional third area along the east coast that was not apparent when using GAMs. Using HBM-INLA, this time with the LGCP, (which takes spatial clustering of animals into account), showed that the Smith Bank had the area of highest relative density. Consequently, we conclude that there are three parts of the study area which host higher relative densities of harbour porpoise (one of which is missed by GAMs).

4. Author contributions

LW, KB, and BS conceived the study; KB collected data; PM provided satellite data; LW, JBI, and ML performed modelling; LW, BS, and VLGT led the writing of the manuscript. All authors contributed critically to drafts and gave final approval for publication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ecolmodel.2022.110011.

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In B. N. Petran & F. Csaaki (Eds.). International Symposium on Information Theory, 2nd ed. Akadeemiai Kiadi, Budapest, Hungary, pp. 267–281.

Bachl, F.E., Lindgren, F., Borchers, D.L., Illian, J.B., 2019. inlabru: an R package for Bayesian spatial modelling from ecological survey data. Methods Ecol. Evol. 10 (6), 760–766. https://doi.org/10.1111/2041-210x.13168.

Bailey, H., Thompson, P., 2010. Effect of oceanographic features on fine-scale foraging movements of bottlenose dolphins. Mar. Ecol. Prog. Ser. 418, 223–233. https://doi.org/10.3354/meps08789.

Becker, E.A., Carretta, J.V., Forney, K.A., Barlow, J., Brodie, S., Hoopes, R., Hazen, E.L., 2020. Performance evaluation of cetacean species distribution models developed using generalized additive models and boosted regression trees. Ecol. Evol. 10 (12), 5759–5784. https://doi.org/10.1002/ece3.6316.

Benjamins, S., van Geel, N., Hastie, G., Elliott, J., & Wilson, B. (2017). Harbour porpoise distribution can vary at small spatiotemporal scales in energetic habitats. *Deep Sea Research Part II: topical Studies in Oceanography,* 141, 191–202. doi:10.1016/j.dsr2.2016.07.002.

Blangiardo, M., Cameletti, M., Baio, G., Rue, H., 2013. Spatial and spatio-temporal models with R-INLA. Spat Spatiotemporal Epidemiol 7, 39–55.

Booth, C.G., Embling, C., Gordon, J., Calderan, S.V., Hammond, P.S., 2013. Habitat preferences and distribution of the harbour porpoise *Phocoena phocoena* west of Scotland. Mar. Ecol. Prog. Ser. 478, 273–285. https://doi.org/10.3354/meps10239.

Brookes, K.L., Bailey, H., Thompson, P.M., 2013. Predictions from harbor porpoise habitat association models are confirmed by long-term passive acoustic monitoring. J. Acoust. Soc. Am. 134 (3), 2523–2533. https://doi.org/10.1121/1.4816577.

BSH, 2013. Standard Investigation of the impacts of offshore wind turbines on the marine environment (StUK4). Hamburg and Rostock: Bundessamt für Seeschifffahrt Und Hydrographie (BSH). Federal Maritime and Hydrographic Agency.

Buckland, S.T., Burt, M.L., Rexstad, E.A., Mellor, M., Williams, A.E., Woodward, R., 2012. Aerial surveys of seabirds: the advent of digital methods. J. Appl. Ecol. 49 (4), 960–967. https://doi.org/10.1111/j.1365-2664.2012.02150.x.

Cameletti, M., Lindgren, F., Simpson, D., & Rue, H. (2012). Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis,* 97, 109–131. doi:10.1007/s10182-012-0196-3.

Carvalho, F.J., de Santana, D.G., Sampaio, M.V., 2020. Modeling overdispersion, autocorrelation, and zero-Inflated count data via generalized additive models and Bayesian statistics in an aphid population study. Neotrop. Entomol. 49 (1), 40–51. https://doi.org/10.1007/s13744-019-00729-x.

Culik, B.M., 2004. Review of small cetaceans: distribution, behaviour and threats *Marine Mammal Action Plan.* In: Regional Seas Reports and Studies, 177. UNEP, CMS Regional Seas, Bonn, Germany;, pp. 218–225.

Diggle, P.J., Moraga, P., Rowlingson, B., Taylor, B.M., 2013. Spatial and spatio-temporal log-Gaussian Cox processes: extending the geostatistical Pparadigm. Stat. Sci. 28, 542–563. https://doi.org/10.48550/arXiv.1312.6536.

Diggle, P.J., Ribeiro, P.J., 2007. Model-based Geostatistics. Springer, New York, NY.

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Wilson, R., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30 (5), 609–628. https://doi.org/10.1111/j.2007.0906-7590.05171.x.

Edwards, K.P., Barciela, R., Butenschön, M., 2012. Validation of the NEMO-ERSEM operational ecosystem model for the North West European Continental Shelf. Ocean Sci. 8 (6), 983–1000. https://doi.org/10.5194/os-8-983-2012.

Embling, C.B., Gillibrand, P.A., Gordon, J., Shrimpton, J., Stevick, P.T., Hammond, P.S., 2010. Using habitat models to identify suitable sites for marine protected areas for harbour porpoises (*Phocoena phocoena*). Biol. Conserv. 143 (2), 267–279. https://doi.org/10.1016/j.biocon.2009.09.005.

Embling, C.B., Illian, J., Armstrong, E., van der Kooij, J., Sharples, J., Camphuysen, K.C.J., Scott, B.E., 2012. Investigating fine-scale spatio-temporal predator–prey patterns in dynamic marine ecosystems: a functional data analysis approach. J. Appl. Ecol. 49 (2), 481–492. https://doi.org/10.1111/j.1365-2664.2012.02114.x.

Fithian, W., Hastie, T., 2013. Statistical models for presence-only data: finite-sample equivalence and addressing observer bias. Ann. Appl. Stat. 7, 1917–1939. https://doi.org/10.48550/arXiv.1207.6950.

Folk, R.L., 1954. The distinction between grain size and mineral composition in sedimentary-rock nomenclature. J. Geol. 62 (4), 344–359.

Hammond, P., Lacey, C., Gilles, A., Viquerat, S., Börjesson, P., Herr, H., . . . Øien, N. (2021). *Estimates of cetacean abundance in European Atlantic waters in summer 2016 from the SCANS-III aerial and shipboard surveys.*

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman & Hall/CRC, New York/Boca Raton.

Hofmann, E.E., Powell, T.M., 1998. Environmental variability effects on marine fisheries: four case histories. Ecol. Appl. 8 (sp1), S23–S32. https://doi.org/10.1890/1051-0761(1998)8[s23:eveomf]2.0.co;2.

Hopkins, P.J., 2011. Exploited fish and shellfish species in the Moray Firth. Proc. Royal Soc. B. Biol. Sci. 91, 57–72. https://doi.org/10.1017/S0269727000009234.

Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. Statistical Analysis and Modelling of Spatial Point Patterns. John Wiley & Sons, Ltd, West Sussex, England.

Illian, J., Sørbye, S.H., Rue, H., 2012. A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). Ann. Appl. Stat. 6 (4), 1499–1530. https://doi.org/10.1214/11-AOAS530.

Jones, E.L., McConnell, B.J., Smout, S., Hammond, P.S., Duck, C.D., Morris, C.D., Matthiopoulos, J., 2015. Patterns of space use in sympatric marine colonial predators reveal scales of spatial partitioning. Mar. Ecol. Prog. Ser. 534, 235–249. https://doi.org/10.3354/meps11370.

Law, R., Illian, J., Burslem, D.F.R.P., Gratzer, G., Gunatilleke, C.V.S., Gunatilleke, I.A.U. N., 2009. Ecological information from spatial patterns of plants: insights from point process theory. J. Ecol. 97 (4), 616–628. https://doi.org/10.1111/j.1365-2745.2009.01510.x.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc., B Stat. Methodol. 73 (4), 423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x.

MacLeod, C.D., Santos, M.B., Reid, R.J., Scott, B.E., Pierce, G.J., 2007. Linking sandeel consumption and the likelihood of starvation in harbour porpoises in the Scottish North Sea: could climate change mean more starving porpoises?. In: Proceedings of the Royal Society Biology Letters, 3, pp. 185–188. https://doi.org/10.1098/rsbl.2006.0588.

Madec, G., & the NEMO team. (2008). *NEMO ocean engine*.

Mendes, S., Turrell, W., Lütkebohle, T., Thompson, P., 2002. Influence of the tidal cycle and a tidal intrusion front on the spatio-temporal distribution of coastal bottlenose dolphins. Mar. Ecol. Prog. Ser. 239, 221–229. https://doi.org/10.3354/meps239221.

Miller, D.L., Glennie, R., Seaton, A.E., 2019. Understanding the stochastic partial differential equation approach to smoothing. J. Agric. Biol. Environ. Stat. https://doi.org/10.1007/s13253-019-00377-z.

Miller, P.I., Scales, K.L., Ingram, S.N., Southall, E.J., Sims, D.W., 2015. Basking sharks and oceanographic fronts: quantifying associations in the north-east Atlantic. Funct. Ecol. 29 (8), 1099–1109. https://doi.org/10.1111/1365-2435.12423.

Mod, H.K., Chevalier, M., Luoto, M., Guisan, A., 2020. Scale dependence of ecological assembly rules: insights from empirical datasets and joint species distribution modelling. J. Ecol. 108 (5), 1967–1977. https://doi.org/10.1111/1365-2745.13434.

Møller, J., Waagepetersen, R.P., 2007. Modern statistics for spatial point processes. Scand. J. Stat. 34 (4), 643–684. https://doi.org/10.1111/j.1467-9469.2007.00569.x.

Norberg, A., Abrego, N., Blanchet, F.G., Adler, F.R., Anderson, B.J., Anttila, J., Ovaskainen, O., 2019. A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. Ecol. Monogr. 89 (3), e01370. https://doi.org/10.1002/ecm.1370.

O'Dea, E.J., Arnold, A.K., Edwards, K.P., Furner, R., Hyder, P., Martin, M.J., Liu, H., 2012. An operational ocean forecast system incorporating NEMO and SST data assimilation for the tidally driven European North-West shelf. J. Oper. Oceanogr. 5 (1), 3–17. https://doi.org/10.1080/1755876X.2012.11020128.

Olson, D.B., Backus, R.H., 1985. The concentrating of organisms at fronts: a cold-water fish and a warm-core Gulf Stream ring. J. Mar. Res. 43 (1), 113–137. https://doi.org/10.1357/002224085788437325.

Ortiz, S.T., Stedt, J., Midtiby, H.S., Egemose, H.D., Wahlberg, M., 2021. Group hunting in harbour porpoises (*Phocoena phocoena*). Can. J. Zool. 99 (6), 511–520. https://doi.org/10.1139/cjz-2020-0289.

Philpott, E., 2013. *Examining Potential Effects of Marine Renewable Energy Developments On Top Predators.* (PhD). University of Aberdeen.

R Core Team, 2017. R: a Language and Environment For Statistical Computing (Version 3.4.1). R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org.

Redfern, J.V., Ferguson, M.C., Becker, E.A., Hyrenbach, K.D., Good, C., Barlow, J., Werner, F., 2006. Techniques for cetacean-habitat modeling. Mar. Ecol. Prog. Ser. 310, 271–295. https://doi.org/10.3354/meps310271.

Renner, I.W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S.J., Warton, D.I., 2015. Point process models for presence-only analysis. Methods Ecol. Evol. 6 (4), 366–379. https://doi.org/10.1111/2041-210x.12352.

Rue, H., Held, L., 2005. Gaussian Markov random fields: Theory and Applications. Chapman and Hall/CRC.

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc., B Stat. Methodol. 71 (2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x.

Sadykova, D., Scott, B.E., De Dominicis, M., Wakelin, S.L., Sadykov, A., Wolf, J., 2017. Bayesian joint models with INLA exploring marine mobile predator–prey and competitor species habitat overlap. Ecol. Evol. 7 (14), 5212–5226. https://doi.org/10.1002/ece3.3081.

Santos, M.B., Pierce, G.J, 2003. The diet of harbour porpoise (*Phocoena phocoena*) in the northeast Atlantic. Oceanogr. Mar. Biol.:Ann. Rev. 41, 355–390.

Scott-Hayward, L.A.S., Mackenzie, M.L., Donovan, C.R., Walker, C.G., Ashe, E., 2014. Complex region spatial smoother (CReSS). J. Comput. Graph. Statist. 23 (2), 340–360. https://doi.org/10.1080/10618600.2012.762920.

Scott, B.E., Sharples, J., Ross, O.N., Wang, J., Pierce, G.J., Camphuysen, C.J., 2010. Sub-surface hotspots in shallow seas: fine-scale limited locations of top predator foraging habitat indicated by tidal mixing and sub-surface chlorophyll. Mar. Ecol. Prog. Ser. 408, 207–226. https://doi.org/10.3354/meps08552.

SeaZone Solutions Ltd. (2005a). *1:250,000 scale tiles: NW25600020, NW25600040, NW25600060, NW25800040*.

SeaZone Solutions Ltd. (2005b). *6 arcsecond grid tiles: NW25600020, NW25600040, NW25600060, NW25800040*.

Simpson, D., Illian, J.B., Lindgren, F., Sørbye, S.H., Rue, H., 2016. Going off grid: computationally efficient inference for log-Gaussian Cox processes. Biometrika 103 (1), 49–70. https://doi.org/10.1093/biomet/asv064.

Simpson, D., Rue, H., Riebler, A., Martins, T.G., Sørbye, S.H., 2017. Penalising model component complexity: a principled, practical approach to constructing priors. Stat. Sci. 32, 1–28. https://doi.org/10.1214/16-STS576.

Simpson, J.H., Crisp, D.J., Hearn, C., Swallow, J.C., Currie, R.I., Gill, A.E., Simpson, J.H., 1981. The shelf-sea fronts: implications of their existence and behaviour. Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences 302 (1472), 531–546. https://doi.org/10.1098/rsta.1981.0181.

Sørbye, S.H., Illian, J.B., Simpson, D.P., Burslem, D., Rue, H., 2019. Careful prior specification avoids incautious inference for log-Gaussian Cox point processes. J. R. Stat. Soc., C Appl. Stat. 68 (3), 543–564. https://doi.org/10.1111/rssc.12321.

Soriano-Redondo, A., Jones-Todd, C.M., Bearhop, S., Hilton, G.M., Lock, L., Stanbury, A., Illian, J.B., 2019. Understanding species distribution in dynamic populations: a new approach using spatio-temporal point process models. Ecography 42 (6), 1092–1102. https://doi.org/10.1111/ecog.03771.

Waagepetersen, R., Schweder, T., 2006. Likelihood-based inference for clustered line transect data. J. Agric. Biol. Environ. Stat. 11 (3), 264. https://doi.org/10.1198/108571106X130557.

Watanabe, S., 2010. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. J. Mach. Learn. Res. 11, 3571–3594.

Williamson, L.D., Brookes, K.L., Scott, B.E., Graham, I.M., Bradbury, G., Hammond, P.S., Thompson, P.M., 2016. Echolocation detections and digital video surveys provide reliable estimates of the relative density of harbour porpoises. Methods Ecol. Evol. 7 (7), 762–769. https://doi.org/10.1111/2041-210X.12538.

Williamson, L.D., Brookes, K.L., Scott, B.E., Graham, I.M., Thompson, P.M., 2017. Diurnal variation in harbour porpoise detection - potential implications for management. Mar. Ecol. Prog. Ser. 570, 223–232. https://doi.org/10.3354/meps12118.

Wood, S.N., 2006. Generalized Additive models: an Introduction With R. Chapman & Hall/CRC, Boca Raton, Florida, USA.

Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. R. Stat. Soc. B 73 (1), 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x.

Wood, S.N., 2017. Generalized Additive Models: An Introduction With R, 2 ed. Chapman and Hall/CRC.

Wood, S.N., Bravington, M.V., Hedley, S.L., 2008. Soap film smoothing. J. R. Stat. Soc., B Stat. Methodol. 70 (5), 931–955. https://doi.org/10.1111/j.1467-9868.2008.00665.x.

Yuan, Y., Bachl, F.E., Lindgren, F., Borchers, D.L., Illian, J.B., Buckland, S.T., Gerrodette, T., 2017a. Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. Ann. Appl. Stat. 11 (4) https://doi.org/10.1214/17-aoas1078.

Yuan, Y., Bachl, F.E., Lindgren, F., Borchers, D.L., Illian, J.B., Buckland, S.T., Gerrodette, T., 2017b. Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. Ann. Appl. Stat. 11 (4), 2270–2297. https://doi.org/10.1214/17-AOAS1078.

Further reading

Mackenzie, M.L., Scott-Hayward, L.A.S., Oedekoven, C.S., Skov, H., Humphreys, E., & Rexstad, E.A. (2013). *Statistical modelling of seabird and cetacean data: guidance document*.