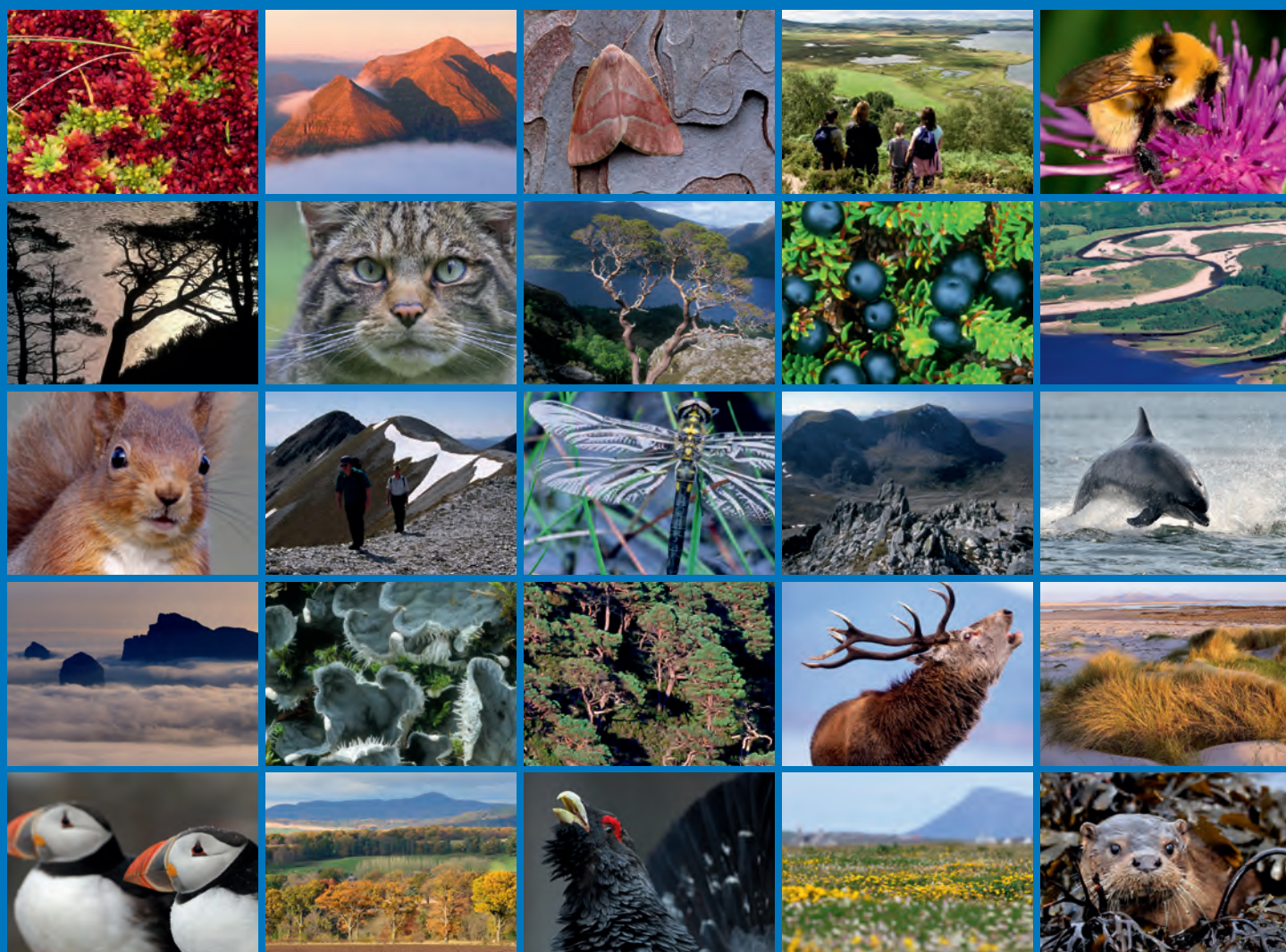


Review of available statistical approaches to help identify Marine Protected Areas for cetaceans and basking shark





Scottish Natural Heritage
Dualchas Nàdair na h-Alba

All of nature for all of Scotland
Nàdar air fad airson Alba air fad

COMMISSIONED REPORT

Commissioned Report No. 573

Review of available statistical approaches to help identify Marine Protected Areas for cetaceans and basking shark

For further information on this report please contact:

Morven Carruthers
Scottish Natural Heritage
Great Glen House
INVERNESS
IV3 8NW
Telephone: 01463 725018
E-mail: morven.carruthers@snh.gov.uk

This report should be quoted as:

Paxton, C.G.M., Scott-Hayward, L.A.S. & Rexstad, E. 2014. Review of available statistical approaches to help identify Marine Protected Areas for cetaceans and basking shark. *Scottish Natural Heritage Commissioned Report No. 573.*

This report, or any part of it, should not be reproduced without the permission of Scottish Natural Heritage. This permission will not be withheld unreasonably. The views expressed by the author(s) of this report should not be taken as the views and policies of Scottish Natural Heritage.

© Scottish Natural Heritage 2014.



COMMISSIONED REPORT

Summary

Review of available statistical approaches to help identify Marine Protected Areas for cetaceans and basking shark

Commissioned Report No. 573

Project no: 13926

Contractor: Paxton, C.G.M., Scott-Hayward, L.A.S. & Rexstad, E.

Year of publication: 2014

Keywords

Statistical modelling; Marine Protected Areas; Risso's dolphin; white-beaked dolphin; minke whale; basking shark; Scottish territorial waters.

Background

The Marine (Scotland) Act 2010 makes provision for the designation of Nature Conservation Marine Protected Areas (hereafter MPAs). In response to this Marine Scotland established the Scottish MPA Project to develop the Scottish MPA network. Here we consider relevant habitat modelling methods and available survey data to help inform identification of MPAs for four charismatic megafaunal species: Risso's dolphin (*Grampus griseus*), white-beaked dolphin (*Lagenorhynchus albirostris*), minke whale (*Balaenoptera acutorostrata*) and basking shark (*Cetorhinus maximus*). Our aims were to:

1. review the appropriate habitat modelling techniques for the identification of marine protected areas,
2. evaluate the quality, quantity and relevance of both the available dependent and explanatory data in evaluating Scottish MPAs (at different spatio-temporal scales),
3. recommend appropriate modelling techniques for each species given the available data, and
4. consider methods for delineating MPAs given the potential results.

Recommendations

1. Preparing sightings data and explanatory covariate data for habitat modelling will take considerable time, even building upon efforts stemming from the Joint Cetacean Protocol (JCP) project. The cost in time and effort to organise these data should be considered along with benefits that might be derived from additional data.
2. The following currently available dependent data should be considered:
 - a. Risso's dolphin: available data collated to inform the JCP project from Scottish territorial waters possibly augmented with JCP data from the Isle of

Man. If only the west coast is of interest for this species then data should be restricted to this spatial extent.

- b. White-beaked dolphin: available data collated to inform the JCP project from Scottish territorial waters initially. If the influence of sandeel presence is negligible (i.e. sandeel presence is not chosen as a predictor), then Scottish shelf waters (i.e. to 200 m depth) should be considered. Sandeel data are not available for the entire shelf.
- c. Minke whale: available data collated to inform the JCP project from Scottish territorial waters but omitting winter data.
- d. Basking shark: available data provided for the JCP project (where basking shark were recorded) from Scottish territorial waters, augmented with the Speedie data, possibly additionally augmented with data from the Isle of Man but omitting winter data.

In all cases a small buffer zone may be applied to the area from which input data are collated, to avoid edge effects in the predictions.

- 3. Additional data from Cetacean Research and Rescue Unit (CRRU), Whale and Dolphin Conservation (WDC) and Hebridean Wildlife and Dolphin Trust (HWDT) may prove useful although some work will be required to integrate these data sets into the existing JCP data resource framework.
- 4. GAMs should be used to create predicted relative density surfaces. It is likely that mixed model GAMs or GEE-GAMs will be used to manage the presumed spatial correlation in the data. It is possible for the data-sparse species (i.e. Risso's dolphin) that a model cannot be fitted, in which case an empirical approach to the identification of regions of relatively higher density could be undertaken.
- 5. Delineation of MPA proposals could be performed by drawing polygons using predicted relative animal densities for individual species. The resulting areas can then be considered by SNH alongside other contextual information (e.g. on behaviour) to inform their advice on areas to be considered for designation as Nature Conservation MPAs.

For further information on this project contact:

Morven Carruthers, Scottish Natural Heritage, Great Glen House, Inverness, IV3 8NW.

Tel: 01463 725018 or morven.carruthers@snh.gov.uk

For further information on the SNH Research & Technical Support Programme contact:

Knowledge & Information Unit, Scottish Natural Heritage, Great Glen House, Inverness, IV3 8NW.

Tel: 01463 725000 or research@snh.gov.uk

Table of Contents	Page
1. INTRODUCTION	1
2. POTENTIAL MODELLING METHODS	2
2.1 Empirical approach	3
2.2 Presence-only modelling	4
2.3 Geostatistical approaches	4
2.4 Regression based approaches	5
2.4.1 Linear models	5
2.4.2 Generalized Linear Models (GLMs)	5
2.4.3 Generalized Additive Models (GAMs)	5
2.4.4 Generalized Estimating Equations (GEEs)	6
2.4.5 Zero-inflated and hurdle models	6
2.4.6 Advanced adaptations to GAMs	6
2.4.7 Classification and regression trees	7
2.5 Further comments on modelling methods for survey data with effort	7
3. DATA EVALUATION	9
3.1 Dependent or response data	9
3.1.1 Risso's dolphin	14
3.1.2 White-beaked dolphin	14
3.1.3 Minke whale	14
3.1.4 Basking shark	14
3.2 Predictor data	14
4. RECOMMENDATIONS FOR DATA SETS AND MODELLING	19
4.1 General recommendations	19
4.1.1 Risso's dolphin	22
4.1.2 White-beaked dolphin	23
4.1.3 Minke whale	24
4.1.4 Basking shark	24
4.2 Modelling outputs	26
4.3 Reliability of model outputs	27
5. POTENTIAL METHODS TO DELINEATE MPAS	28
5.1 MPA selection guidelines	28
5.2 Identifying cells	28
5.3 Joining cells	30
5.4 Consideration of additional data	30
5.5 Summary on MPA delineation	31
6. RECOMMENDATIONS	32
REFERENCES	33

Acknowledgements

Our thanks to all those individuals and groups associated with collecting, compiling, organizing and providing data to the JCP analyses that were used in this analysis, especially Mick Baines for Phase I and Tim Dunn for Phases II & III of the JCP and Morven Carruthers for this project. The individual contributors/organisations included in this project were: Colin Macleod and Sarah Bannon (University of Aberdeen); DECC; EDP Renewables and Repsol Nuevas Energias UK; European Seabirds at Sea data providers; the UK Government and the Oil and Gas Industry for funding Seabirds at Sea Surveys; Hebridean Whale and Dolphin Trust; Henrik Skov, (DHI) & Jan Durinck (Marine Observers); Irish Whale and Dolphin Group; Mainstream Renewable Power; Marine Conservation Research Ltd & IFAW; MARINELife/BDRP; Northern North Sea Cetacean Ferry Surveys (NORCET); ORCA & Company of Whales; RWE npower plc.; Sea Watch Foundation (& Manx Whale and Dolphin Trust) especially Peter Evans; the partner organizations and funders associated with SCANS I 1994; the partner organizations and funders associated with SCANS-II 2005; Paul Thompson & Kate Brookes and colleagues at the University of Aberdeen Lighthouse Field Station, including for the provision of data collected through projects funded by DECC and Talisman Energy (UK) Ltd.; Whale and Dolphin Conservation; WWT Consulting; Alison Gill; Colin Speedie; and Cetacean Research and Rescue Unit. Also our thanks to Lorenzo Millazo for programming of the online data checking tool; and to Morven Carruthers, Katie Gillham, John Baxter and anonymous reviewers for comments.

1. INTRODUCTION

The Marine (Scotland) Act 2010 makes provision for the designation of Nature Conservation Marine Protected Areas (hereafter MPAs). In response to this Marine Scotland established the Scottish MPA Project to select MPAs and develop the Scottish MPA network. The MPA Selection Guidelines (Marine Scotland, 2011b) set out the process for identifying MPAs in Scottish waters and include a list of MPA search features (species and habitats of conservation importance for which spatial protection is considered appropriate) that are being used to help drive the selection of MPAs. Three species of cetaceans (Risso's dolphin *Grampus griseus*, minke whale *Balaenoptera acutorostrata* and white-beaked dolphin *Lagenorhynchus albirostris*) as well as basking shark, *Cetorhinus maximus*, are included on the list of MPA search features (Marine Scotland, 2011b).

Evaluation of cetacean and elasmobranch distribution and habitat assessment of these species has been suggested in the designation of MPAs both in Scotland (e.g. Bailey & Thompson, 2009; Embling *et al.*, 2010) and elsewhere (e.g. Cañadas *et al.*, 2005).

Considerable research has been undertaken to identify the habitat preferences of large marine megafauna in Scottish local waters (e.g. Hastie *et al.*, 2003; Canning *et al.*, 2008; Bailey and Thompson, 2009; Marubini *et al.*, 2009; Anderwald *et al.*, 2012) and elsewhere (e.g. Pierce *et al.*, 2010; Azzellino *et al.*, 2012; Becker *et al.*, 2012; Thorne *et al.*, 2012) from visual surveys and other methods. Surveys can lead to relative estimates of distribution and abundance. The exact form of these estimates depends on the data collected and the analytical methods undertaken. Modelling of (relative) abundance by habitat characteristics and density allows predictions to be made into regions of low effort and even into the future (Becker *et al.*, 2012).

Here we review habitat modelling methods available to identify regions of high density for these four species with the intention of providing evidence to support SNH's advice on MPAs for these species. The modelling methods are considered in light of the data available on animal locations as well as potential habitat data with which to predict their abundance.

There are four aims of this report:

1. to review the appropriate habitat modelling techniques to support identification of marine protected areas,
2. to evaluate the quality, quantity and relevance of both the available dependent and explanatory data in evaluating Scottish MPAs (at different spatio-temporal scales),
3. to recommend appropriate modelling techniques for each species given the available data, and
4. to consider methods for delineating MPAs given the potential results.

2. POTENTIAL MODELLING METHODS

A habitat model of an animal population may possess three functions:

1. to describe the distribution of animals seen i.e. how the animals vary in distribution in time and space;
2. to predict the distribution of animals (normally by interpolation into regions of low or zero effort); and
3. to explain the observed distribution of animals.

The three functions are not mutually exclusive and any model that could explain the observed distribution of animals would presumably also make excellent predictions. It is possible to have models with good predictive properties but little explanatory power. For example a species' distribution could perhaps be modelled accurately by position and time if it appears persistently in a particular location without any knowledge of what drives it to that location. However, the interpolative and especially extrapolative properties of such models may be limited. In the case of MPA delineation, the primary aims of the models are ultimately description and prediction, specifically the aim is identification of regions of persistent (relative) high density using available predictor data (see Section 4). Explanation is also of potential importance in delineating MPAs, but has been considered as a secondary aim for the purposes of assessing appropriate modelling approaches. A predictive model with both environmental covariates and position predictors (e.g. Long. and Lat.) will allow precise prediction of where the animals occur, and in addition, plots of the response of density to the environmental predictors will also be available as an output. However the response to, for example, depth given that position is in the model may be very different to the response to depth in the absence of position in the model, i.e. position probably explains much of the variation in the observed densities due to depth, so if depth is in the model as well it explains only what was not explained by position.

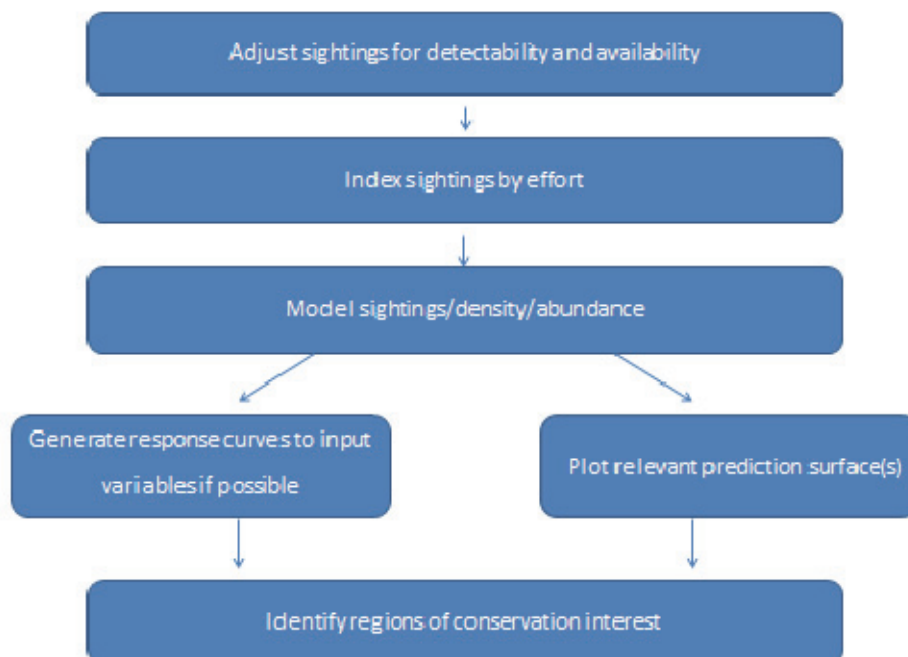


Figure 1. Outline of the process to determine an area of conservation interest from survey data. Note that at the stage where the process splits either method can be employed but the right hand path is the most efficient use of the data and models.

Figure 1 illustrates the process of using habitat data to delineate regions of conservation interest where survey data are available. The sightings are adjusted for detectability and availability (especially if absolute estimate of abundance is desired). Sightings are spatially referenced and effort standardised (unless presence-only models are used, see below). The resultant numbers can then be predicted using other (habitat) data. The habitat model provides outputs that explain the distribution of animals both mathematically (e.g. how Risso's dolphin density fluctuates with depth, day of year, etc.) and with maps of density by location. Locations of conservation interest could be selected on the basis of the formula produced by the model (e.g. by determining most frequent depth, etc.) or by mapping the predictions of the model. The latter is a more economical way of using the model as it allows all the characteristics of the model to be considered simultaneously.

Visual survey data come in a variety of forms: raw sightings, sightings per unit effort and adjusted sightings per unit effort. Sightings can be adjusted for detection (see Buckland *et al.*, 2001) and availability at the surface. Sightings with effort are essential as the amount of effort undertaken to observe animals almost certainly influences the number seen. Non-spatially referenced sightings data from formal surveys with effort (where the broad area of sighting is known) can be used to estimate abundance in a given area empirically (a "design-based" approach). This means that a density can only be estimated as uniform over a broad area. However, if the sightings are more finely spatially referenced, a finer empirical approach can be undertaken or a spatial model may be fitted. Spatial models allow fluctuations in animal density within survey blocks to be identified. For a given locality, such spatial models should provide a more accurate and more precise estimate than an empirical approach, if the relevant predictors are used in the model, but with the risk of bias if the model is misspecified (Hedley and Buckland, 2004). Spatially referenced data are frequently spatially (or temporally) correlated; i.e. a datum from one point in space/ time is correlated with datums from adjoining points in space/ time. The data points cannot necessarily be considered as wholly independent of each other and to treat them as independent leads to an underestimation of the uncertainty in the predictions. Large marine animals are not evenly distributed in the ocean: animals are not found over most of a given region and often found concentrated within widely dispersed areas. This can also contribute to the spatial correlation. Modelling techniques should consider these features of the data. The models suitable for study do not need to be mechanistic/ explanatory, although relevant input variables may help with model interpolation.

We provide a non-technical review of a variety of spatial modelling techniques that could be used to determine relative and absolute densities of large aquatic fauna, after an initial consideration of an empirical approach to the problem. Note that Booth and Hammond (2011), and Macleod *et al.* (2008) also review habitat modelling approaches by comparing analyses of the same data sets and Redfern *et al.* (2006) offer an overarching review without analysis of a dataset.

2.1 Empirical approach

One way of determining regions of conservation interest is simply by empirical assessment of regions of higher density. If a location is visited four times and three of those four times the species of interest is sighted, then this region is of greater interest than another location where the species is not seen in four or five visits. Any comparable spatially referenced index of abundance could be used to identify regions of interest by simply delineating around areas (see Section 5) where the (relative) density was above a particular threshold. Such a procedure could be valuable if predictor variables are lacking or the region of interest is disjunct and dissimilar in properties from the rest of a surveyed area (e.g. Rockall or St Kilda). Such scores could be estimated for specific time periods to take into account seasonality of animal movements.

If the resultant density surface is very rough it can be smoothed by application of a kernel density estimator or similar to produce a smoothed map of densities. Obviously if all the data are used, the method assumes the population is static but using temporal subsets of the data could allow for dynamic responses.

2.2 Presence-only modelling

Presence-only models have been used to predict the distribution of species including cetaceans (e.g. Thorne *et al.*, 2012). The raw data of such analyses are sightings without a measure of survey effort. These methods assume knowledge of the potential environmental predictors over the region of interest. In a sense “presence-only” is a misnomer as such models:

1. make an implicit assumption about the effort associated with the sightings namely that it is uniform (e.g. BIOCLIM, Busby, 1991; DOMAIN, Carpenter *et al.*, 1993; some versions of Ecological Niche Factor Analysis, Hirzel *et al.*, 2002; Maxent, Phillips *et al.*, 2006) even though no actual search effort is considered. Such models are sometimes referred to as “profile techniques” (Robertson *et al.*, 2003); or
2. require a sample of presence-absence data (some further versions of Ecological Niche factor analysis, GARP) to calibrate the data/ model (e.g. Elith *et al.*, 2006); or
3. generate assumed absences (Ferrier & Watson, 1997; Ferrier *et al.*, 2002) either from outwith or within the range of the presence-only data (the former sometimes referred to as “real absences” and the latter “pseudo-absences” although in neither situation is truth known).

The latter two methods are sometimes referred to as “group discrimination” techniques (Robertson *et al.*, 2001). If effort is not uniform over the region, or the sampled presence/ absence survey range is biased relative to the presence-only data spatial range or the presumed absences are unrepresentative of actual absences, then inaccuracies in prediction will occur. Collecting sightings by boat from a particular area and inferring there are no animals where there has been no survey effort is unlikely to be defensible. There is no merit in the use of presence-only models if effort data are available, as inferences made from effort standardised data will *always* be more accurate than inferences made from sightings alone (e.g. Macleod *et al.*, 2008). This also means that effort should always be recorded by observers.

If only sightings data are available, then the use of presence-only models is still problematic because the resultant biases associated with the assumptions and subsequent fitting of such models are unknown. Presence-only models explain reported presences better than a random model when considered using a cross-validation technique (where a subset of data is used to create a model that is then tested against the remainder of the data) (e.g. Macleod *et al.*, 2008) but this is unsurprising if the same effort biases occur in the training set as the testing set. The only time such results would be trustworthy is the unusual situation where effort was unknown yet known to be uniform.

2.3 Geostatistical approaches

Geostatistical approaches to modelling data were primarily developed for use in geological situations where the initial aim was to predict density of interest with no explanatory power (Haining, 1990). The approaches have the advantage of dealing with the spatial correlations in the data. Kriging is the approach used where there is interpolation between known points to unknown locations between the points. However, universal kriging allows the use of indicator variables to aid the interpolation process (Upton & Cook, 2002; Cressie & Wikle, 2011). As kriging and regression techniques have improved (with the addition of explanatory

variables of greater functional complexity to kriging, and spatial autocorrelation to regression methods) they have become more similar (e.g. Stein & Corsten, 1991). Current regression methods allow more model flexibility. A method of kriging to allow for zero-inflated (see below) count data has been used to describe the distributions of fin whales (*Balaenoptera physalus*) in the Mediterranean (Monestiez *et al.*, 2006; Bellier *et al.*, 2010). One problem with these methods is that they are extremely computationally expensive (Paxton, pers. obs) and only small data sets can be analysed on a desktop machine.

2.4 Regression based approaches

Regression based approaches to the modelling of survey data offer a different approach to geostatistical modelling in that explanatory factors as well as pure spatial referencing can be used to predict densities. Sightings/ sightings per unit effort/ densities are considered as a function of input variables. Models can be considered linear in the parameters, which includes single linear terms or simple curves of the form $x + x^2$ etc. and combinations thereof. Even some responses that are non-linear, in the narrow statistical sense, e.g. functions like the logistic curve $e^x/1+e^x$ (see Augustin *et al.*, 1996) can be considered in this framework. Some methods can even deal with spatial/ temporal correlation between adjoining data points.

2.4.1 Linear models

(General) linear models consider the responses to multiple input variables (predictors) when these input variables can be both factors and continuous variables. Responses to continuous predictors can be a simple linear function or a polynomial function. In the simplest case:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

where Y is the response (say sightings per unit effort) and X_1 and X_2 are the input variables (say depth and sea surface temperature). β_0 , β_1 , and β_2 are parameters estimated by the fitting process. The error structure, ε is assumed normal with mean 0 and variance σ^2 (Faraway, 2005).

2.4.2 Generalized Linear Models (GLMs)

Generalized linear models (McCullagh & Nelder, 1989) are similar to linear models except that the error structure does not have to be normal but can take a variety of functional forms (McCullagh & Nelder, 1989; Faraway, 2006). This means, for example, that data with a presence-absence structure (zeros and ones) can be modelled. Non-linear functional forms that can be linearised by transformation (e.g. logistic curve) can be incorporated. Data that are overdispersed can be modelled assuming a quasi-Poisson or negative binomial error distribution. Further extensions of the generalized linear modelling methods allow modelling of spatial correlation in the data (e.g. see Pinheiro & Bates, 2000; Zuur *et al.*, 2009).

2.4.3 Generalized Additive Models (GAMs)

Generalized additive models (Hastie & Tibshirani, 1990) are similar to GLMs except that the response to the predictors can be a smooth curve composed of linked polynomial segments (Wood, 2006). Therefore curved responses to predictor variables can be fitted allowing more flexibility than when using GLMs. Spatial correlations in the data can also be incorporated. A variety of smooth functions can be fitted of parametric or non-parametric form (i.e. the component parts can be described by a formula or not).

2.4.4 Generalized Estimating Equations (GEEs)

A different way of considering the problem of spatial/ temporal correlation in a regression context is made using Generalized Estimating Equations (GEEs) (Hardin & Hilbe, 2003). Data can appear overdispersed when in fact it is only apparently overdispersed, the relevant predictors are missing (Hardin & Hilbe, 2003). For example there could be clumping in the data caused by intrinsic overdispersion or by the presence of multiple sources of food on which the animals are feeding. GEEs offer a different approach to conventional GLMs in that they are population averaged as opposed to cluster specific, that is the fit is derived from the data rather than the model (Hardin & Hilbe, 2003; Ghisletta & Spini, 2004; Zuur *et al.*, 2009). Such techniques have been used to model cetacean distributions (e.g. Panigada *et al.*, 2008; Pirota *et al.*, 2011; Paxton *et al.*, 2013).

2.4.5 Zero-inflated and hurdle models

These constitute a class of regression type model where the fitted values can be thought of as a function of two different processes that generate the observed density of animals in a location (Lambert, 1992). For example there could be two reasons why a minke whale is not recorded on a survey in the River Ness (assuming perfect detectability/availability). It might not have been there because the River Ness is not a suitable habitat for a minke whale. No healthy minke whale would ever find itself in the River Ness. Alternatively minke whale could be found there, but did not happen to be present when the survey was made. There are potentially two processes to be modelled differently. Zero-inflated and hurdle models seek to model density/abundance with a presence-absence component which models absolute absence versus presence, and an additional component which models the abundance if present/possible. In practice, a probability of presence is calculated as well as an estimate of density.

Hurdle models (also known as zero-altered models) model the probability of presence of the animals given a set of variables and then model the non-zero presences in a separate modelling process. No zeros can be generated by this second stage as the data are modelled assuming a zero-truncated distribution (e.g. Goetz *et al.*, 2012; Read *et al.*, submitted). In contrast zero-inflated models evaluate whether the zeros are true permanent zeros (the animal will never be in the location) or simply transient zeros (the animal will be in the location some of the time) by modelling the presence/absence and numbers if present at the same time and estimating the allocation of zeros to one type or another (Lambert, 1992).

Given the distribution of large marine animals is undoubtedly a function of avoiding wholly unacceptable regions and random, albeit infrequent appearance in favourable regions then zero-inflated/hurdle type of modelling closely matches the decisions animals may ultimately take. While hurdle models can be readily fitted in a two stage process, the zero-inflated models, as currently implemented, do not have the robustness of some of the other modelling methods and there appear to be fitting and prediction problems (Paxton pers. obs.). However, Choudhary *et al.* (2012) modelled river dolphins in the Ganges using this method.

2.4.6 Advanced adaptations to GAMs

This section covers both spatially adaptive and spatially complex modelling. Global smoothing methods (i.e. methods that smooth equally over the range of the data, such as GAMs) are sub-optimal, particularly when the underlying spatial surface is smooth in some areas and highly structured in others. Thus, local features in the data will be over-smoothed (estimated to be too flat) and far reaching trends under-smoothed (estimated to be too wiggly). Spatially adaptive methods allow the smoothness of the response to vary depending on where in the predictor range you are. There are several methods for one-dimensional smooths (e.g. Ruppert, 2000; Pintore *et al.*, 2006; Crainiceanu *et al.*, 2007; Walker *et al.*,

2010) but fewer tested methods for two-dimensional smooths (Krivobokova *et al.*, 2007; Yue *et al.*, 2010; Walker *et al.*, in prep). Some methods are not general or well automated, whilst others require many estimated parameters and are computationally expensive.

The second advancement of GAMs is spatially complex modelling. Spatially complex refers to the topography of the area of interest. For example, a study region containing a topographically complex coastline including, perhaps, some islands. The issue is that many modelling methods allow the spatial estimate to change smoothly with Euclidean (straight-line) distance. This is not always realistic, as a straight path between points may cross a boundary (e.g. coastline), so the nearest distance does not reflect “how the whale swims”, resulting in ‘leakage’ in model predictions: high or low densities in one body of water can influence across boundaries into another area where in fact connectivity would be low e.g. around a peninsula. There are several methods available for dealing with this situation: Complex Region Spatial Smoother (CReSS; Scott-Hayward *et al.*, in press); Soap film smoothing (SOAP; Wood *et al.*, 2008); Geodesic Low Rank Thin Plate Splines (GLTPS; Wang & Ranalli, 2007) and Finite Element splines (FEL-splines; Ramsay, 2002). Of these methods, GLTPS cannot be used in areas with islands, FEL-splines have a strong boundary condition which causes problems and SOAP is difficult to parameterise.

The spatially adaptive method of Walker *et al.* (SALSA, in prep); and spatially complex CReSS method have both been recently applied to the Joint Cetacean Protocol data resource (Paxton *et al.*, 2013).

2.4.7 Classification and regression trees

An alternative regression-like method is the use of classification and regression trees (CARTs). Classification trees are a hierarchical decision making process for classifying data (Breiman *et al.*, 1984; Upton and Cook, 2002). Regression trees integrate regression models in the decision making process. As the position of megafauna is a function of avoiding wholly unsuitable regions and random appearance in favourable regions then it is conceivable such areas could be distinguished by different constants (or indeed regression lines) for different combinations of the explanatory variables. For example if the relationship between basking shark surface density is different on the east coast of Britain compared to the west coast of Britain then this could be described by two constants or, if there is an additional predictor, two distinct regression lines describing the relationship between for example density and depth. The rules split the data hierarchically like a tree. The disadvantage in such methods is that no uncertainty can be generated for the decision process, unlike the use of an interaction in conventional regression. The end result is a tree-like descriptive breakdown of the habitat preference. Examples of regression trees in estimating cetacean responses to environmental parameters include Panigada *et al.* (2008), Goetz *et al.* (2007) and Macleod *et al.* (2007). Regression trees can perform very well in prediction but can suffer from bias (e.g. Opper *et al.*, 2012).

2.5 Further comments on modelling methods for survey data with effort

All of the regression models up to Section 2.4.3 can be thought of as elaborations of earlier forms. For example Generalized Additive Models can do all the operations of Generalized Linear Models and Generalized Linear Models can do all the things that linear models can do. GEE are regression models, but have a slightly different approach to modelling noise. The methods from Section 2.4.5 onwards are effectively mutually exclusive. However, there is no reason in the future that they cannot be combined, as they all refine different aspects of the model fitting process with the exception of regression trees, which is different modelling philosophy. Geostatistical approaches are robust, but computationally expensive and typically lack the flexibility of regression methods. CARTs can be useful in understanding the data, but the lack of uncertainty in the predictions is unhelpful. Regression models are

computationally cheap (with the exception of some of the more complex smoothers) and produce predictions with uncertainty but there can be model-sensitivity issues (Paxton, pers. obs.).

3. DATA EVALUATION

The relevant data are either dependent data: the actual sightings and associated effort or predictor data: the various potential predictors of the density variation. Section 3.1 describes the dependent data. Section 3.2 describes the predictor data and Section 3.3 considers the realized data availability given both predictor and dependent data ranges.

3.1 Dependent or response data

Most dependent data under consideration here were submitted to the Joint Cetacean Protocol (JCP) data resource, a collection of platform of opportunity and formal survey data collated by the Joint Nature Conservation Committee (JNCC) and analysed in Paxton *et al.* (2013). This consists of shipboard and aerial visual survey data (sightings and effort). Acoustic data are not included, nor are land based data, as the integration of these types of data with vessel-based visual survey data requires a number of often unmet assumptions or additional information (e.g. the distribution of the animals away from the shore in the case of static land-based survey data). The JCP data resource is useful in that the data have already been checked and segmented and in the case of the three cetacean species under consideration a corrected sightings per unit effort index is already available. Dependent data not included in the JCP data resource are basking shark survey data collected by Colin Speedie in the Inner Hebrides from 2003 to 2008, hereafter the “Speedie data”. Here we will review the available dependent data (i.e. available survey data) for their:

1. applicability for modelling;
2. availability for modelling (i.e. the spatial/temporal relevance of the data to Scottish territorial waters);
3. quantity (e.g. number of 10 km segments).

Other survey data may be available in the future such as that from the Cetacean Research and Rescue Unit (CRRU, Robinson, pers. comm.) from 2009 – 2012; Whale and Dolphin Conservation (WDC) data from north-east Lewis from 2010 – 2012; and additional Hebridean Whale and Dolphin Trust (HWDT) data from 2011 – 2012. These data will need to be formally segmented and the sightings will need to be allocated to those segments prior to modelling. The data will then have to be augmented with the existing JCP data. After that it will be necessary to estimate detection functions and hence detection probabilities prior to spatial modelling.

Data were considered from a variety of potential areas:

1. solely from the 12 nm Scottish territorial limit (green polygon in Figure 2);
2. the Scottish MPA project area (northern red polygon in Figure 2) also the same as the Scottish part of the UK continental shelf;
3. the UK continental shelf (the larger all inclusive red polygon in Figure 2);
4. OSPAR (The Convention for the protection of the marine environment of the North-East Atlantic) regions II and III (the blue polygon in Figure 2);
5. Scottish shelf waters (Scottish waters to 300 m depth, not shown).

Any model predictions would be for Scottish territorial waters as this is the area where MPAs need to be identified for cetaceans and basking shark (SNH, 2012), nevertheless it may be useful to include data from outwith this region to construct models as these might illuminate the potential role of environmental predictors.

Many surveys were submitted to the JCP data resource. All of these data consist of sightings with associated segmented effort. The data submitted to the JCP were only used in the current project where specific permission was given. For some of the JCP surveys, either basking shark sightings were not always recorded or permission has not been granted to use the basking shark material from the surveys. Therefore the quantity of data varies greatly dependent on species (Table 1). Data were segmented into lengths of approximately 8 km (mean 8.1 km). The data could be analysed at a finer resolution but if this were to happen the proportion of non-zero to zero segments would decrease. This is undesirable as models of data with a large proportion of zeros can be difficult to fit (see Section 4.1). Consequently modelling would become more difficult. The Speedie data also consist of segmented effort with associated sightings (mean segment length 11.3 km).

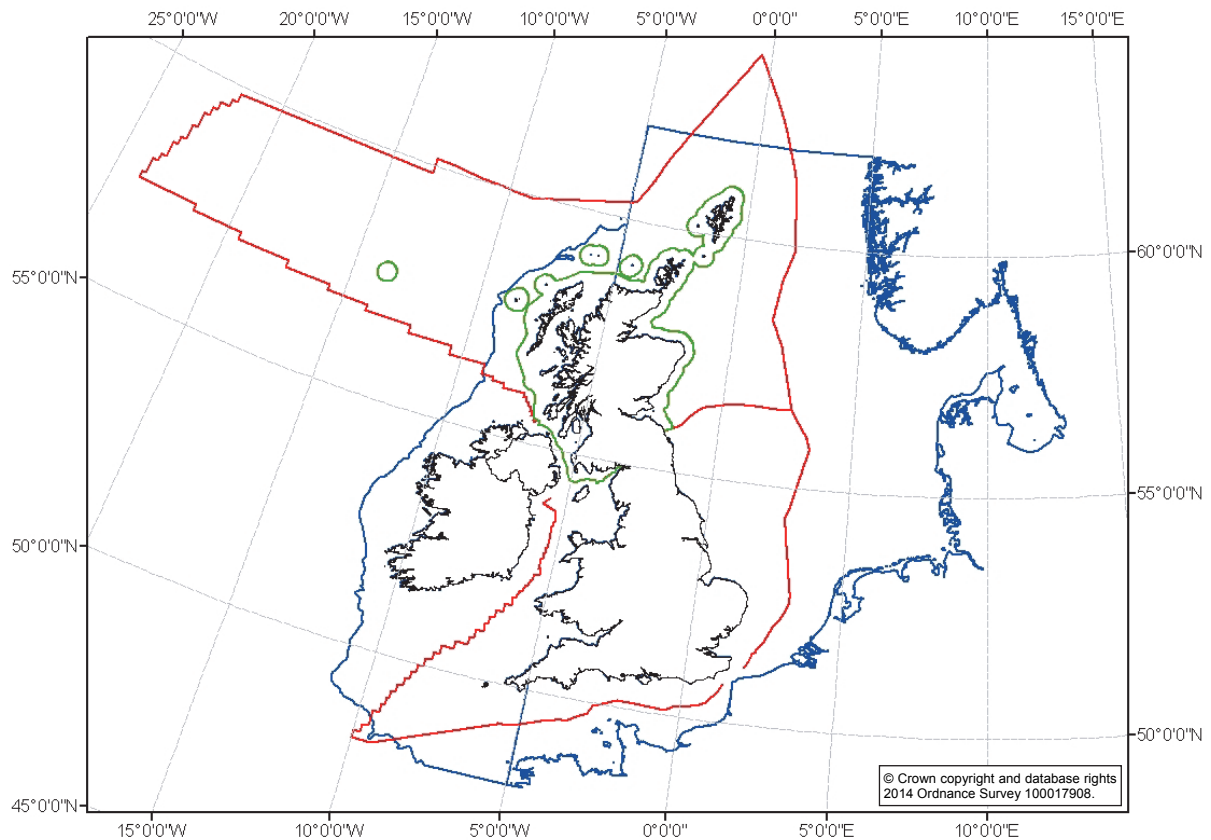


Figure 2. Potential spatial ranges of dependent data considered in this analysis. Red: UK Continental Shelf (northern: Scottish (also Scottish MPA Project area), southern: English, Welsh, Northern Irish and Manx combined), Green: Scottish territorial waters limit. Blue: Boundaries of OSPAR regions II (east) and OSPAR III (west). Not shown: Scottish shelf waters from the blue OSPAR III western boundary then the southern, eastern and northern boundaries of the Scottish MPA project areas within 300 m depth contour. N.B. Rockall shelf area not included in the relevant ranges of the Scottish territorial waters and Scottish geographic shelf waters for the purposes of this project.

Sightings associated with effort can be readily modelled as raw sightings per unit effort (SPUE) or be adjusted for detection in a variety of ways (see Buckland *et al.*, 2001, 2004). The data can then be considered as spatially referenced points corresponding to small segments of effort as per Hedley & Buckland (2004) and Hedley *et al.* (2004). Another way in which this sort of data has been considered in the past (e.g. Macleod and Zuur, 2005; Monestiez *et al.*, 2006) is by binning the relevant effort into a regular grid and then modelling the SPUEs or equivalent spatially referenced points according to the grid. However, the choice of grid resolution is essentially arbitrary and spatial information can be lost by this

method. Because survey data are dynamic the data have to be further binned by time. Given these constraints direct modelling of segments seems the most appropriate response where environmental variables are allocated to segments directly.

Table 1 gives the maximum amount of effort available for each species in each area under consideration, for all seasons and Tables 2 and 3 show the contrasting amount of effort available in winter and summer.

Table 1. Maximum available data for each species. The temporal ranges for the data sets are JCP cetaceans (1982 – 2011), JCP basking shark (1982 – 2010) and Speedie data (2003 – 2006)

Species	Data source	Data range (spatial)	Amount of effort (km)	Number of segments*	Percent of non-zero segments
Risso's dolphin	JCP (where agreement from data providers)	OSPAR2 & OSPAR3	1,339,382	164,174	0.1
		Scottish MPA Project Area	339,824	48,208	0.2
		Scottish Territorial Limit	203,832	31,441	0.3
		Scottish Waters <300m	300,262	43,675	0.2
		UK Continental Shelf	96,7814	118,657	0.1
White beaked dolphins	JCP (where agreement from data providers)	OSPAR2 & OSPAR3	1,341,947	164,446	0.5
		Scottish MPA Project Area	339,824	48,208	1.1
		Scottish Territorial Limit	203,832	31,441	0.9
		Scottish Waters <300m	300,262	43,675	1.2
		UK Continental Shelf	969,883	118,880	0.6
Minke whale	JCP (where agreement from data providers)	OSPAR2 & OSPAR3	1,341,947	164,446	0.8
		Scottish MPA Project Area	339,824	48,208	1.7
		Scottish Territorial Limit	203,832	31,441	1.8
		Scottish Waters <300m	300,262	43,675	1.8
		UK Continental Shelf	969,883	118,880	0.9
Basking shark	JCP (where agreement from data providers and basking shark were recorded)	OSPAR2 & OSPAR3	1,303,754	159,063	0.2
		Scottish MPA Project Area	304,838	43,234	0.7
		Scottish Territorial Limit	177,472	27,666	1.1
		Scottish Waters <300m	266,221	38,833	1.0
		UK Continental Shelf	931,690	113,497	0.3
Basking shark	Colin Speedie data	OSPAR2 & OSPAR3	7,930	700	22.8
		Scottish MPA Project Area	7,930	700	22.8
		Scottish Territorial Limit	7,930	700	22.8
		Scottish Waters <300m	7,930	700	22.8
		UK Continental Shelf	7,930	700	22.8

*Mean of c. 8.1 km for JCP data and 11.3 km for Speedie basking shark survey data.

Table 2. Maximum available data for each species in winter (January, February and March). The temporal ranges for the data sets are JCP cetaceans (1982 – 2011), JCP basking shark (1982 -2010) and Speedie data (2003 – 2006)

Species	Data source	Data range (spatial)	Amount of effort (km)	Number of segments*	Percent of non-zero segments
Risso's dolphin	JCP (where agreement from data providers)	OSPAR2 & OSPAR3	232,978	25,600	0.02
		Scottish MPA Project Area	32107	3,891	0.05
		Scottish Territorial Limit	18,901	2,308	0.04
		Scottish Waters <300m	28,182	3,455	0.01
		UK Continental Shelf	184,366	19,715	0.01
White beaked dolphins	JCP (where agreement from data providers)	OSPAR2 & OSPAR3	235,543	25,872	0.2
		Scottish MPA Project Area	32,107	3,891	0.8
		Scottish Territorial Limit	18,901	2,308	0.6
		Scottish Waters <300m	28,182	3,455	0.9
		UK Continental Shelf	186,434	19,938	0.2
Minke whale	JCP (where agreement from data providers)	OSPAR2 & OSPAR3	235,543	25,872	<0.001
		Scottish MPA Project Area	32,107	3,891	0
		Scottish Territorial Limit	18,901	2,308	0
		Scottish Waters <300m	28,182	3,455	0
		UK Continental Shelf	186,434	19,938	<0.001
Basking shark	JCP (where agreement from data providers and basking shark were recorded)	OSPAR2 & OSPAR3	232,608	25,554	0
		Scottish MPA Project Area	32,106	3,891	0
		Scottish Territorial Limit	18,901	2,308	0
		Scottish Waters <300m	28,248	3,463	0
		UK Continental Shelf	183,996	19,669	0
Basking shark	Colin Speedie data	OSPAR2 & OSPAR3	0	0	-
		Scottish MPA Project Area	0	0	-
		Scottish Territorial Limit	0	0	-
		Scottish Waters <300m	0	0	-
		UK Continental Shelf	0	0	-

*Mean of c. 8.1 km for JCP and 11.3 km for Speedie basking shark survey data.

Table 3. Maximum available data for each species in summer (July, August and September). The temporal ranges for the data sets are JCP cetaceans (1982 – 2011), JCP basking shark (1982 – 2010) and Speedie data (2003 – 2006).

Species	Data source	Data range (spatial)	Amount of effort (km)	Number of segments*	Percent of non-zero segments
Risso's dolphin	JCP (where agreement from data providers)	OSPAR2 & OSPAR3	537,899	69,802	0.2
		Scottish MPA Project Area	171,636	25,190	0.3
		Scottish Territorial Limit	99,368	16,052	0.4
		Scottish Waters <300m			
		UK Continental Shelf	378,998	50,070	0.2
White beaked dolphins	JCP (where agreement from data providers)	OSPAR2 & OSPAR3	537,899	69,802	0.7
		Scottish MPA Project Area	171,636	25,190	1.4
		Scottish Territorial Limit	99,368	16,052	1.2
		Scottish Waters <300m	153,535	22,989	1.5
		UK Continental Shelf	378,998	50,070	0.9
Minke whale	JCP (where agreement from data providers)	OSPAR2 & OSPAR3	537,899	69,802	1.2
		Scottish MPA Project Area	170,636	25,190	2.2
		Scottish Territorial Limit	99,368	16,052	2.5
		Scottish Waters <300m	153,535	22,989	2.3
		UK Continental Shelf	378,998	50,700	1.4
Basking shark	JCP (where agreement from data providers and basking shark were recorded)	OSPAR2 & OSPAR3	520,542	67,359	0.4
		Scottish MPA Project Area	155,614	22,922	1.0
		Scottish Territorial Limit	86,735	14,247	1.6
		Scottish Waters <300m	137,512	20,721	c. 1.2
		UK Continental Shelf	361,641	47,627	0.5
Basking shark	Colin Speedie data	OSPAR2 & OSPAR3	5,286	469	15.6
		Scottish MPA Project Area	5,286	469	15.6
		Scottish Territorial Limit	5,286	469	15.6
		Scottish Waters <300m	5,286	469	15.6
		UK Continental Shelf	5,286	469	15.6

*Mean of c. 8.1 km for JCP and 11.3 km for Speedie basking shark survey data.

3.1.1 *Risso's dolphin*

Risso's dolphin are seldom recorded (less than 200 sightings in the whole JCP database), especially in winter (Table 1 vs. Table 2) making them by far the most sparsely recorded of all the species of interest. Therefore the proportion of non-zero to zero segments is extremely low. Modelling this species in the JCP proved difficult (authors pers. obs., see Paxton *et al.*, 2013) and only simple models of distribution could be fitted. The proportion of non-zero to zero density segments is higher in the Scottish territorial limit waters compared to other areas, although more non-zero data outwith Scottish waters are available by using data from around the Isle of Man.

3.1.2 *White-beaked dolphin*

Of all the available areas for consideration, Scottish shelf waters (to 300 m depth) has the most favourable ratio of non-zero to zero segments (Table 1).

3.1.3 *Minke whale*

Minke whales occur at slightly higher density in Scottish territorial waters compared to other regions (Table 1). Minke whale are seldom seen in winter surveys (Table 2). From Paxton *et al.* (2013) it appears this is not a detectability issue and the species is genuinely elsewhere in the winter.

3.1.4 *Basking shark*

The data for this species from the JCP are much reduced compared to the effort available for the cetacean species. Nevertheless basking sharks are recorded at greatest frequency within the Scottish territorial limit compared to other British waters. Basking sharks also occur at high frequency around the Isle of Man.

The Speedie data are associated with a much higher reported density of basking shark, suggesting that a dedicated basking shark survey has a far higher detection rate for basking sharks than cetacean surveys (although the same areas have not been directly compared). This suggests that any subsequent modelling of these data should take survey type as an explanatory variable.

Basking shark are not recorded in the available data for winter which is consistent with a long held observation that basking sharks are not at the surface around the British Isles in the winter (Parker & Boeseman, 1954; Witt *et al.*, 2012). One way to make the data more tractable for modelling is to exclude winter survey data for basking shark thus increasing the ratio of non-zero to zero segments.

3.2 **Predictor data**

Potential explanatory data are given in Table 4 along with their spatial and temporal range, resolution and source. Some are dynamic (e.g. sea surface temperature) whilst others are static (e.g. depth). All have been considered as potential or actual predictors of megafaunal distribution in the past or advocated as such (Table 5). The table suggests that compared to cetaceans, we know little of the habitat preferences of basking shark.

Predictors vary in their temporal and spatial extent with static variables having the best spatial coverage. It is possible that, for variables with small spatial gaps (e.g. chlorophyll a), interpolative techniques such as kriging could be used to fill in missing values at the risk of adding additional uncertainty to the modelling process.

Some variables are really measuring the same thing so both should not be placed in the same final model i.e. distance to coast and depth (Figure 3). It may be that the presence of oceanographic fronts will be correlated with tide/current related variables.

Resolution varies considerably for the explanatory variables, however, with the exception of sea surface temperature, the available resolution of the explanatory variables are higher than the segment resolution associated with sighting data, hence data can be readily supplied to each. The lower resolution of the sea surface temperature data means that the response to temperature might not be quite as clear as it otherwise might be.

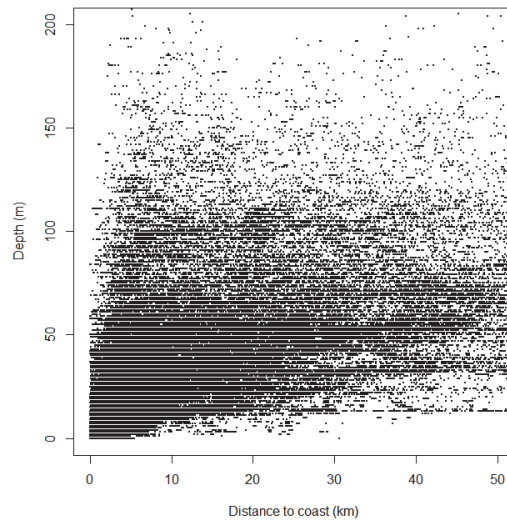


Figure 3. Relationship of depth and distance to coast in the JCP data resource.

Table 4. Spatial predictors considered in this review.

Predictor	Description	Spatial range	Temporal range (if relevant) Complete years only	References
Position	Either as Long, Lat or Easting, Northing	Global	-	Collected on survey
Depth	ETOPO1: One arc minute depths (m) on a regular grid	Global	-	http://www.ngdc.noaa.gov/mgg/global/global.html
Distance to coast	Coastline from NOAA, distance to nearest point on coastline.	Global	-	http://oceancolor.gsfc.nasa.gov/SeaWiFS/
Slope	Calculated from above, rounded to 5000 m Easting, Northing grid	Global	-	Standard formula used see Burroughs and McDonnell (1998)
Time in tidal cycle	Fractional indicator of tie since load tide (%)	Inshore waters only	1982 - 2010	POLTIPS http://www.pol.ac.uk/appl/poltips3.html
Tidal range	Depth from lowest to highest spring tide	Inshore waters only	1982 - 2010	POLTIPS http://www.pol.ac.uk/appl/poltips3.html
Sea Surface Temperature	OI SST at 1/4 degree daily resolution	Global	1982 - 2010	ftp://eclipse.ncdc.noaa.gov/pub/OI-daily-v2/NetCDF/
Wave energy & current energy	0.1 to 35 km	British and Irish legal shelf waters and North Sea	Mechanistic model	EU SeaMap Consortium, as described in ABP Marine Environmental Research (2010)
Chlorophyll a	SeaWiFS at 9 km resolution	North Atlantic with lacunae	1997 - 2010	http://oceancolor.gsfc.nasa.gov/SeaWiFS/
Probability of presence <i>Eledone cirrhosa</i>	1 km grid	Scottish territorial waters	2008 - 2012	Macleod <i>et al.</i> (2014)

Table 4 Continued – Spatial predictors considered in this review.

Predictor	Description	Spatial range	Temporal range (if relevant) Complete years only	References
CPUE of <i>Eledone cirrhosa</i>	1 km grid	Scottish territorial waters	2008 - 2012	Macleod <i>et al.</i> (2014)
Sandeel presence	Irregular positions	Scottish territorial waters	Based on data collected over last 2 decades assumed static	P. Wright, pers. comm. (Wright <i>et al.</i> , 2000)
Sediment	Positions on a 0.006° grid with lacunae	OSPAR 2 & 3	Assumed static	EUSeamap: http://jncc.defra.gov.uk/euseamap UKSeaMap: http://jncc.defra.gov.uk/ukseamap , as described in Cooper <i>et al.</i> (2010)
Oceanographic fronts	Positions on a 0.0025° grid	UKCS (fronts based on SST)	Assumed static, derived from period 1998 - 2008 (based on SST)	Miller <i>et al.</i> (2010)
		Scottish part of EEZ (fronts based on colour)	Assumed static, derived from 2009 - 2011 (based on colour)	Miller <i>et al.</i> (2014)

Table 5. Examples of predictors considered in megafaunal habitat modelling

Species	Predictor	Considered as a potential predictor
Risso's dolphin	Depth	Azzellino <i>et al.</i> (2012), Cañadas <i>et al.</i> (2005)
	Slope	Azzellino <i>et al.</i> (2012), Paxton <i>et al.</i> (2013)
	Temperature	Paxton <i>et al.</i> (2013)
	Curled octopus abundance	MacLeod <i>et al.</i> (2014)
	Chlorophyll a	de Segura <i>et al.</i> (2008)
White-beaked dolphins	Depth	Paxton <i>et al.</i> (2013)
	Slope	Paxton <i>et al.</i> (2013)
	Temperature	Paxton <i>et al.</i> (2013)
Minke whale	Depth	Paxton <i>et al.</i> (2013)
	Slope	Paxton <i>et al.</i> (2013)
	Temperature	Anderwald <i>et al.</i> (2012)
	Chlorophyll a	Anderwald <i>et al.</i> (2012)
	Sandeel presence	Anderwald <i>et al.</i> (2012)
Other cetaceans	Distance to coast	Marubini <i>et al.</i> (2009)
	Time in tidal cycle	Booth and Hammond (2011)
	Tidal range	Isojunno <i>et al.</i> (2012), Embling <i>et al.</i> (2010), Marubini <i>et al.</i> (2009)
	Sediment	Embling <i>et al.</i> (2010), Booth & Hammond (2011)
	Oceanographic fronts	De Boer (2010)
Basking shark	Temperature	Cotton <i>et al.</i> (2005), Skomal <i>et al.</i> (2004)
	Oceanographic fronts	Sims (2008), Sims <i>et al.</i> (2000)

4. RECOMMENDATIONS FOR DATA SETS AND MODELLING

4.1 General recommendations

As the sightings considered in this report are all associated with effort, presence-only methods can be immediately set aside. The methods must account for any spatial correlation in the data as a failure to do so would result in an underestimate of the uncertainty associated with any predictions. Geostatistical techniques do not allow the opportunity to explore the interactions of potential predictors and are computationally expensive, so this leaves regression type modelling. Smooths allow more flexible responses than polynomials suggesting that some GAM based approach should be used over a GLM. The sightings data will likely be overdispersed (e.g. Figure 4), i.e. the animals, where found, are clustered in areas of preferred habitat and spatially correlated, so the modelling method should be able to account for these features.

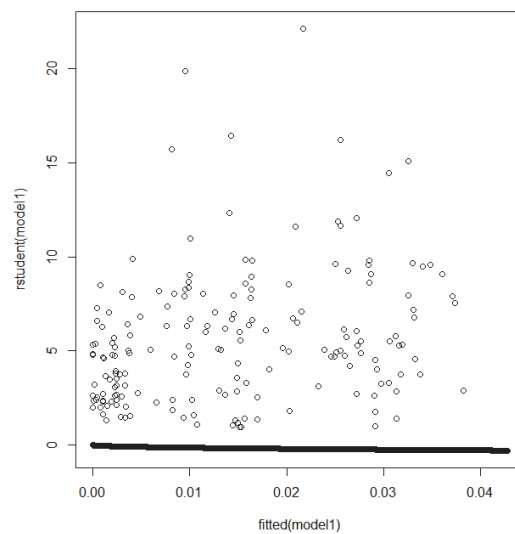


Figure 4. Fits residual plot from a simple Poisson GLM of density of Risso's dolphins in response to location and day of year. Residuals are more spread out at higher fitted values indicating the presence of overdispersion. If there was no overdispersion a roughly symmetrical cloud would be seen. The solid line along the bottom of the graph is caused by the presence of numerous zeros in the data.

Therefore we recommend that the default modelling approach should be either GAMs assuming a spatially corrected error structure, a GAM hurdle model, or GEE-GAMs. As the immediate aim here is prediction rather than biological explanation, position is considered as an explanatory variable.

The spatial and temporal range to consider for the available data requires several considerations:

1. Spatial relevance: locations closer to the Scottish territorial waters are more relevant for modelling Scottish territorial waters than locations further afield.
2. Temporal relevance: the most recently collected data are more relevant than more historical records, although past records can allow estimation of potential future fluctuations and cycling on a decadal scale etc.
3. Best signal: by excluding data from regions and times where no animals were seen modelling becomes more tractable.

Based on these principles, Table 6 provides a breakdown of potential data sets based on the most favourable ratio of non-zero to zero data in Tables 1-3 and the ranges of potential predictor data seen in Table 4. There is a case for including a small buffer around the selected region to avoid strange predictions near the edge of the region.

One further consideration is the spatial scale of any habitat modelling. No formal method for evaluating the appropriate scale of segmentation for such spatial analyses has been developed. However, several practical considerations are apparent:

1. increased resolution is desirable to produce more detailed understanding of the process under consideration;
2. increasing resolution means smaller segments which increases the ratio of zero to non-zero segments and increases the correlation between segments;
3. segment resolution should not be less than the available explanatory variables;
4. ideally habitat should not vary significantly within a segment;
5. prediction cells should not be smaller than the effective area of segments;
6. ideally (but doubtfully in practice) segments should be independent.

Previous modelling experience suggests that a target segment length of 10 km is a useful compromise between these six criteria over a large region like the Scottish territorial waters. If conclusions were being made from data where there is a high density over a very local area then perhaps segments could be shortened.

We now discuss variables to be considered in each specific species model (Table 6). It is assumed all models should initially consider position (as the primary aim is prediction), year, day of year (or season), depth (as opposed to distance to coast), sea surface temperature at the highest resolution possible (noting that it will be correlated with day of year), chlorophyll a (but availability of these data is restricted to some years), sediment type (discrete rather than continuous measurement) and the oceanographic front index (as indicator of general marine productivity).

Species specific recommendations for the starting models are given below.

Table 6. Potential data sets for spatial habitat modelling. All models include position (Long. or Lat. or Easting or Northing), depth, SST (at 0.25 degrees), sediment type and a front index. The proposed spatial extent is always Scottish territorial waters except in the case of white beaked dolphin where details are given for Scottish shelf waters (to 300 m depth contour).

Species	Data source	Data range (temporal)	Amount of effort (km)	Number of current segments	Percent of non-zero segments	Potential Predictors***
Risso's dolphin	JCP (where agreement from data providers)	1982 - 2011	203,832	31,441	0.3	Position, Depth, SST, Sediment type & Front index
		1998 - 2011	122,352	21,998	0.2	Position, Depth, SST, Sediment type, Front index and Chlorophyll a
White-beaked dolphin	JCP (where agreement from data providers)	1982 - 2011	300,262	43,675	1.2	Position, Depth, SST, Sediment type, Front index & Sandeels
		1998 - 2011	122,352	21,998	0.5	Position, Depth, SST, Sediment type, Front index, Sandeels and Chlorophyll a
Minke whale	JCP (where agreement from data providers)	1982 - 2011	203,832	31,441	1.8	Position, Depth, SST, Sediment type, Front index & Sandeels
		1998 - 2011	122,352	21,198	1.9	Position, Depth, SST, Sediment type, Front index, Sandeels and Chlorophyll a
Basking shark*	JCP (where agreement from data providers and basking shark were recorded) + Speedie data	1982 - 2010 (not winter)	158,570 + 7,930	25,358 + 7,930	4.5	Position, Depth, SST, Sediment type & Front index
		1998 - 2010 (not winter)	88,853 + 7,930	16,468 + 7,930	c. 4.5	Position, Depth, SST, Sediment type, Front index & Chlorophyll a

*There are additional data from the Isle of Man which would usefully supplement these data. **Sandeel presence/absence. ***Distance to coast and tidal variables also to be included dependent on the presence of other variables.

4.1.1 Risso's dolphin

The default modelling method should be as outlined in Section 4.1. Even with the data restricted to Scottish territorial waters the ratio of non-zero to zero segments is very low with only 82 non-zero (out of 31,441) segments in the period 1982 – 2011 and only 37 non-zero (out of 21,198) segments in the period 1998 – 2011. If there is not enough signal in the distribution of this species to successfully model the density, a more empirical approach should be considered with a density surface generated from adjusting sightings per unit effort rather than mediated through a model (e.g. see Figure 9). Addition of further data from the Isle of Man may aid modelling by increasing the proportion of non-zero segments. Data from the east coast could also be omitted, in order to improve the model fit by increasing the ratio of non-zero to zero segments. The suggested available data are plotted in Figure 5. Additional data may be available from Cetacean Research and Rescue Unit (CRRU), Hebridean Whale and Dolphin Trust (HWDT), and Whale and Dolphin Conservation (WDC). Whilst the data coverage is considerable overall, for particular seasons and years coverage can be very sparse (this is the case for all species).

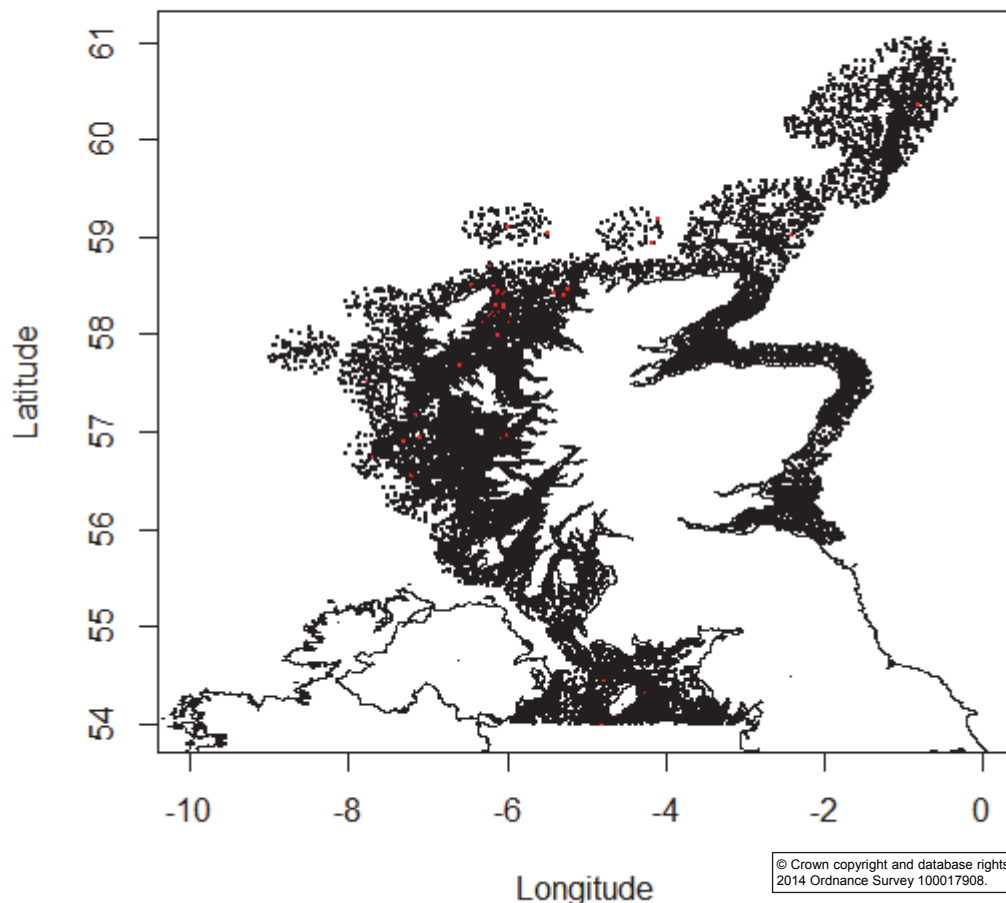


Figure 5. Proposed Risso's dolphin data for analysis. Each point represents a midpoint of a JCP segment within the Scottish territorial limit. Black indicates an empty segment and red indicates a segment where one or more Risso's dolphins were seen (1982 - 2011). Note the paucity of segments with sightings.

Probability of presence or abundance of the curled octopus *Eledone cirrhosa* predictions (from Macleod *et al.*, 2014) were considered, but all inputs into the habitat models for this species can be included in the Risso's dolphin models directly, making the process simpler. Furthermore, Macleod *et al.* (2014) did not find a relationship between Risso's dolphin space use and the predicted highest densities of *Eledone*. Sandeel presence/ absence was not considered as this species is not known to consume sandeels (Baird, 2009).

4.1.2 White-beaked dolphin

The default modelling method should be as outlined in Section 4.1 using data from Scottish shelf waters, possibly with additional datasets from CRRU, HWDT and WDC, as this combination has the best ratio of zero to non-zero segments (Table 1, Table 6). Sandeel presence/ absence should also be included, as sandeels are a noted feature of their diet and have been implicated in affecting their distribution (Canning *et al.*, 2008; Okka *et al.*, 2010). However, sandeel data are only available for inshore regions. Therefore it might be better to consider Scottish territorial waters initially, evaluate the potential influence of sandeel presence/ absence, and then consider using the larger data set if sandeels turn out to be unimportant.

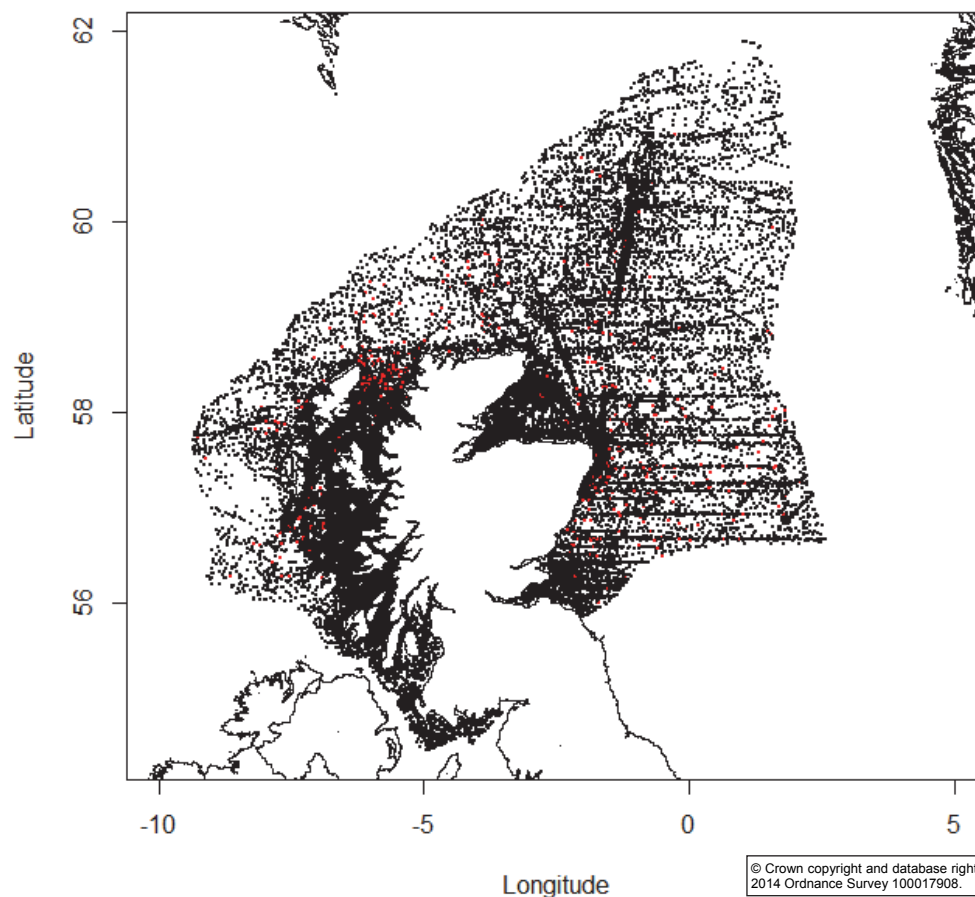


Figure 6. Proposed white-beaked dolphin data for analysis. Each point represents a midpoint of a JCP segment within the Scottish territorial limit. Black indicates an empty segment and red indicates a segment where one or more white-beaked dolphins were seen (1982 – 2011).

4.1.3 Minke whale

The default modelling method should be as outlined in Section 4.1, using data from Scottish territorial waters alone, possibly augmented with additional datasets from CRRU, HWDT and WDC. If the ratio of zero to non-zero segments makes data fitting problematic, winter survey data could be removed, unless the response to sea surface temperature was of specific biological interest. Sandeels should also be included in the modelling of these data and, as minke whale may switch diets through the year (e.g. Anderwald *et al.*, 2012), then the interaction of sandeel presence/ absence and day of year should also be considered.

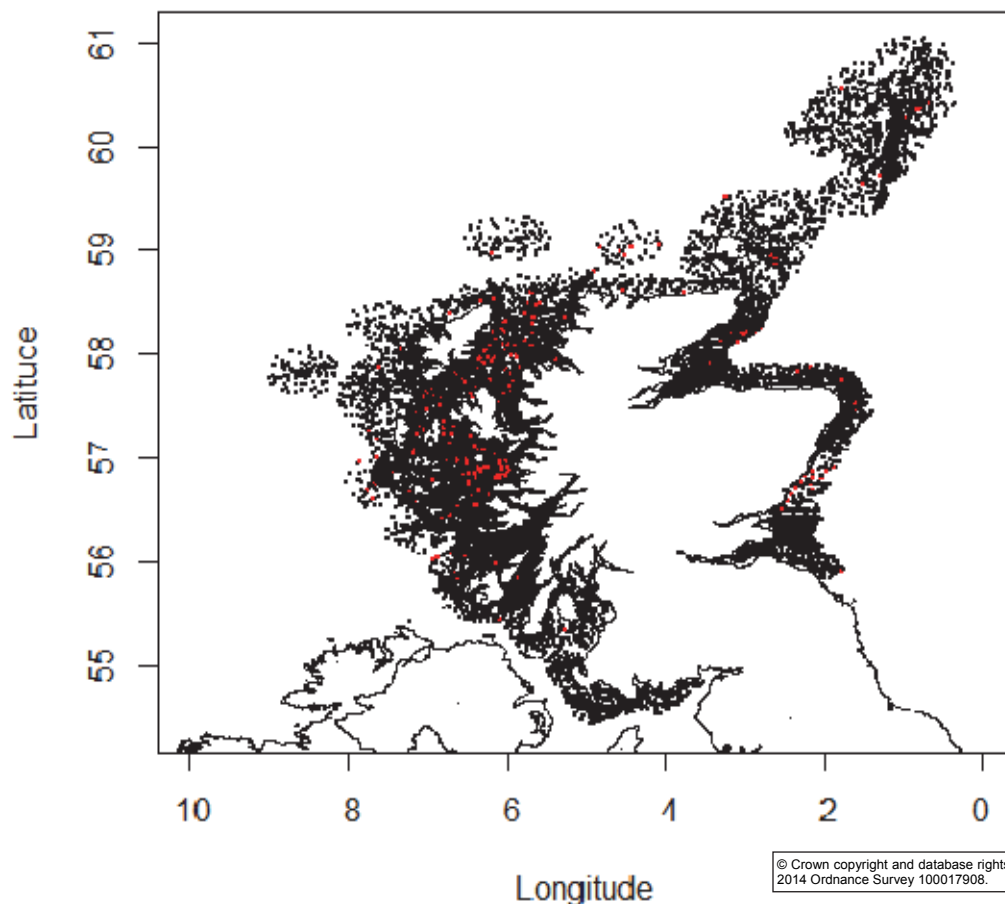


Figure 7. Proposed minke whale data for analysis. Each point represents a midpoint of a JCP segment within the Scottish territorial limit. Black indicates an empty segment and red indicates a segment where one or more minke whales were seen (1982 – 2011).

4.1.4 Basking shark

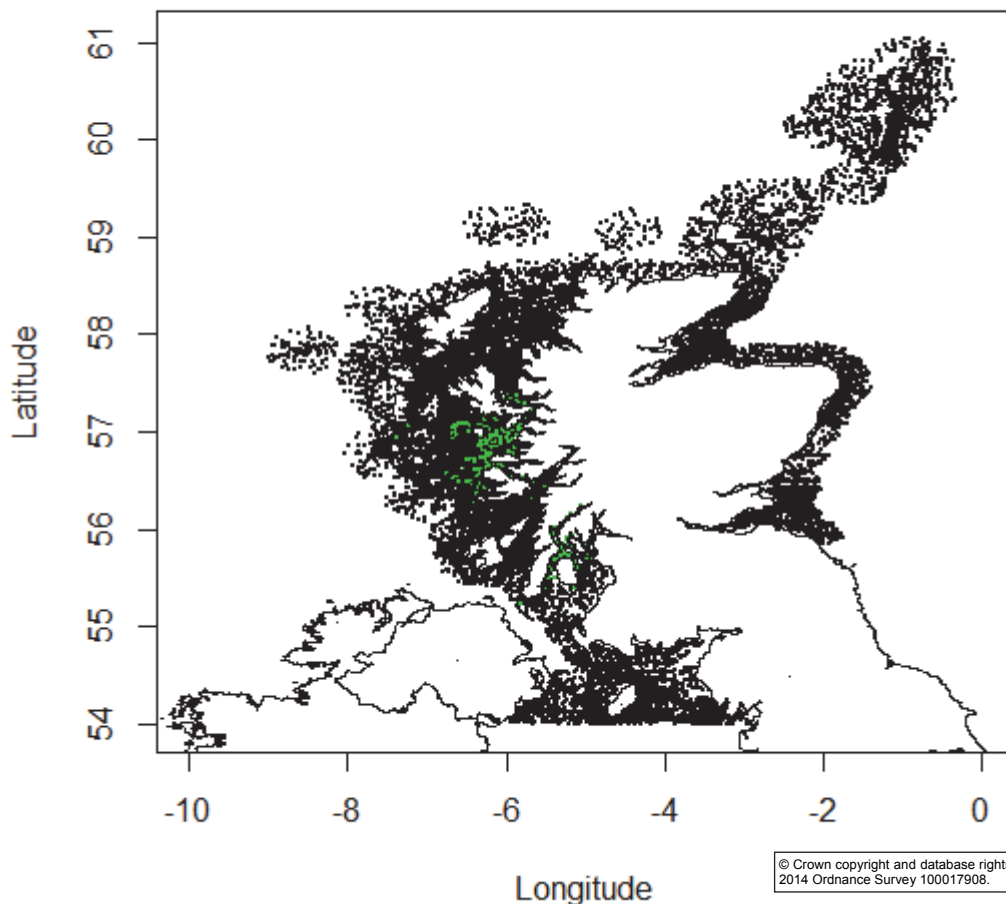
Basking sharks occur in few segments in the JCP data set, especially compared to the similarly segmented data from Speedie. There are two obvious reasons for this:

1. The combination of covariate values associated with the Speedie survey locations is conducive to surface sightings of basking shark. (i.e. the area of the Speedie survey

contains a greater concentration of basking shark than the overall JCP data resource, presumably because of the environmental characteristics of this area).

2. Dedicated basking shark surveys are rather better at detecting sharks than dedicated cetacean surveys and platforms of opportunity.

The two data sets can be integrated. If they were modelled as the default above, an additional variable indicating whether the data came from a dedicated shark survey should be included in the analysis. The basking shark data are not shown, as the sightings per segment have not yet been calculated, but the spatial coverage of effort is given (Figure 8).



*Figure 8. Proposed basking shark data for analysis (1982 – 2010). Each point represents a segment in the Scottish territorial limit. Black indicates a JCP segment and green a segment from the Speedie data set. **N.B. Sightings are not shown.***

The basking shark data are likely to be overdispersed and/ or spatially correlated as in the case of the cetacean data (see Paxton *et al.*, 2013). The best ratio of zero to non-zero segments is from Scottish territorial waters. We recommend just this region should be considered, as most explanatory variables are available for these localities. If necessary, sightings from around the Isle of Man might be brought into the analysis, which would improve the ratio of zero to non-zero sightings. Like minke whale there is little to be gained

by including winter effort in the model unless responses to sea surface temperatures were of specific interest.

Using surface observations of basking shark to evaluate potential MPAs is different from considering cetacean observations. Cetaceans have to surface to breathe, basking shark do not. Inferences from habitat models of basking shark based on visual surveys indicate where they choose to surface not necessarily where they congregate in the greatest numbers.

4.2 Modelling outputs

The statistical modelling process involves taking the (adjusted) sightings per unit effort and modelling these data in terms of the predictors given above. Once the model is fitted there are four types of output:

1. Predictions for particular regions over Scottish territorial waters on a cellular grid for particular points in time (if dynamic predictors are used). The exact resolution of the cells would depend on the initial segmentation used in the modelling (see above). Here, by way of example, we have depicted the map using the same scale as the JCP data analysis, each cell is 25km² (Figure 10). Given a segmentation of c. 10 km, a much finer resolution would not be possible. The problem of objectively choosing a conservation region becomes a problem of recognising cells of interest. These prediction surfaces will have uncertainty associated within them, typically in the form of lower and upper confidence surfaces (e.g. Paxton *et al.*, 2013).
2. In addition, for particular regions, time series of (relative) abundances can be produced over seasons or years.
3. Some indication of the responses of density to the predictors (component plots or partial plots) so the habitat preferences of the animals can be visualised.
4. In the case of Risso's dolphin, if a model cannot be fitted, an adjusted sightings per unit effort surface could be produced to allow identification of regions of high density (e.g. Figure 9).

It might be thought that (3) alone is the most important output from the point of view of MPA designation. However, this would not use the full power of the models. Essentially, as the models produce a formula for describing the distribution of the megafauna, that formula can be used to make predictions not only from locations where data were collected but also interpolations into other areas/ times. Therefore for a given point in time, the density plot incorporates all the information from (3) plus purely locational information (i.e. from predicting from spatial variables like latitude and longitude), and so can capture variations in density that cannot currently be coupled with any particular environmental predictor (such as the seasonal appearance of animals off a particular island). If the model contains dynamic variables, this density surface can be generated for different seasons/ years, etc. The identification of optimum locations for Marine Protected Areas then becomes a delineation problem (see below) for a given density surface or surfaces (if dynamic variables are considered).

The role a given location plays in the life history of a particular species of marine megafauna (e.g. calving areas, courtship areas, etc.) is *not* elucidated by habitat modelling, unless the density of calves is explicitly modelled or the proportion of mother/ calf pairs is identified (see Cañadas & Hammond (2006) for an example with bottlenose dolphins), or a critical time period is identified so the distribution during those periods could be predicted. Where the information exists these types of consideration may be more usefully taken into account qualitatively, alongside the quantitative outputs from habitat modelling.

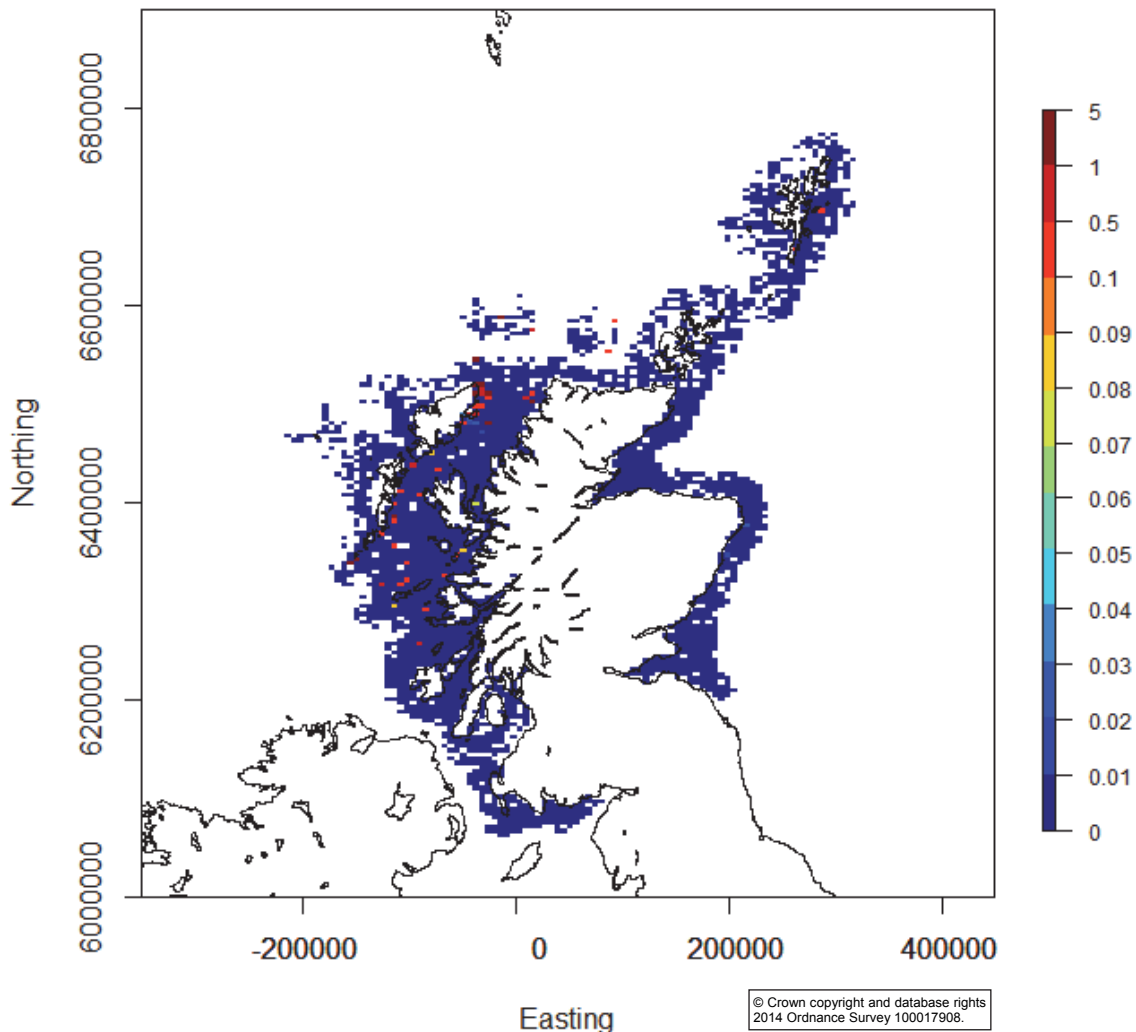


Figure 9. Risso's dolphin adjusted sightings per unit effort (km^2) for the territorial waters of Scotland based on survey effort 1994 – 2010. Effort is binned into cells 5 by 5 km therefore empty cells indicate zero effort or extraterritorial waters.

4.3 Reliability of model outputs

Uncertainty in the predictions is supplied by confidence intervals on the prediction surfaces typically in the form of lower and upper confidence surfaces (e.g. Paxton *et al.*, 2013) i.e. the lower surface represents a surface of cells each of which is the 2.5th percentile of the range of predictions made for that cell based on the uncertainty in the models. Likewise the upper surface consists of a surface of cells each of which is the 97.5th percentile of the range of predictions made for that position.

5. POTENTIAL METHODS TO DELINEATE MPAS

5.1 MPA selection guidelines

The MPA selection guidelines (Marine Scotland, 2011b) outline the way in which Nature Conservation MPAs might contribute to the protection of biodiversity and geodiversity in Scotland's seas. For mobile species MPAs may be appropriate for protecting:

- a) significant aggregations or communities of important marine species in Scottish waters;
- b) essential areas for key life cycle stages of important mobile species that persist in time, including habitats known to be important for reproduction and nursery stages; and/or
- c) areas contributing to the maintenance of ecosystem functioning in Scottish waters.

It should be stressed that habitat modelling does *not* directly address all of the above three points. Density modelling can identify regions of high density for each species of interest, which when considered together does address point (a). As mentioned in Section 4.2 unless mothers/ calves are modelled (and the data for this are really lacking), spatial modelling does not identify essential areas for key life history stages, just regions of highest density of the species at particular times (which may not be the same thing). However, the modelling could provide plots of the response of density to the various environmental predictors tested. Alongside other ecological information and more qualitative data it may be possible to interpret the outputs of these plots in terms of points (b) and (c).

Other criteria are important as well: identification of areas that are meaningful to the conservation of the species concerned and defining a boundary in such a way that a physical feature can be identified for potential "users" (e.g. a submarine feature that can be detected on sonar i.e. depth or substrate) is desirable (SNH, pers. comm).

We approach this topic presuming the major focus is to identify, delineate and protect predicted regions of high megafaunal abundance or persistence from the models. It is not necessarily the case that the identification of predicted cells of interest and the subsequent joining of them to form a larger area should be seen as distinct processes, as some methods for identifying cells of conservation interest will be coupled with boundary considerations. The results of the model will be considered by SNH alongside other contextual information.

5.2 Identifying cells

The most obvious and simplest action would be to have some threshold, defining the cells of interest. For example in Figure 10, one could choose all cells that are associated with a density of greater than 0.1 animals/km² or include 90% of the local population (based on the above surface), or relative density (e.g. Embling *et al.*, 2010) or a set population size. Only if all the assumptions of the analysis are met could the estimates be considered an estimate of absolute density. Furthermore, density for a designated time period would presumably reflect usage over that time period as well. Also there are potential issues with specifying threshold densities before the data have been analysed, in that the definition of 'high density' may vary significantly between species. Similarly setting percent population targets may not be straightforward for cetaceans and basking shark as our knowledge of local population abundance is often insufficient to assess targets confidently. Therefore, of these methods, choosing cells based on relative density or abundance (i.e. the estimated abundance in a cell or area as the quotient of the abundance in a wider area e.g. Scottish territorial waters) is likely to be the most appropriate.

The example in Figure 10 is for a single point in time. If persistence or some measure of temporal use is to be incorporated, then seasonal/ yearly densities would need to be combined. If the aim is protection of many species then weighted standardised densities could be combined from different species. This approach may be appropriate, for example, in the area of the Skye to Mull MPA search location, which is proposed for both basking shark and minke whale, or in the area of the Southern Trench MPA search location, proposed for minke whale and white-beaked dolphin (SNH & JNCC, 2012). Any designation of an MPA should take into account uncertainty (e.g. Gerrodette & Eguchi, 2011). Uncertainty could be incorporated by considering a prediction surface that is based on some confidence boundary. For example having an inclusion threshold for cells based on whether they had a density >0.1 on the upper 97.5% confidence prediction surface.

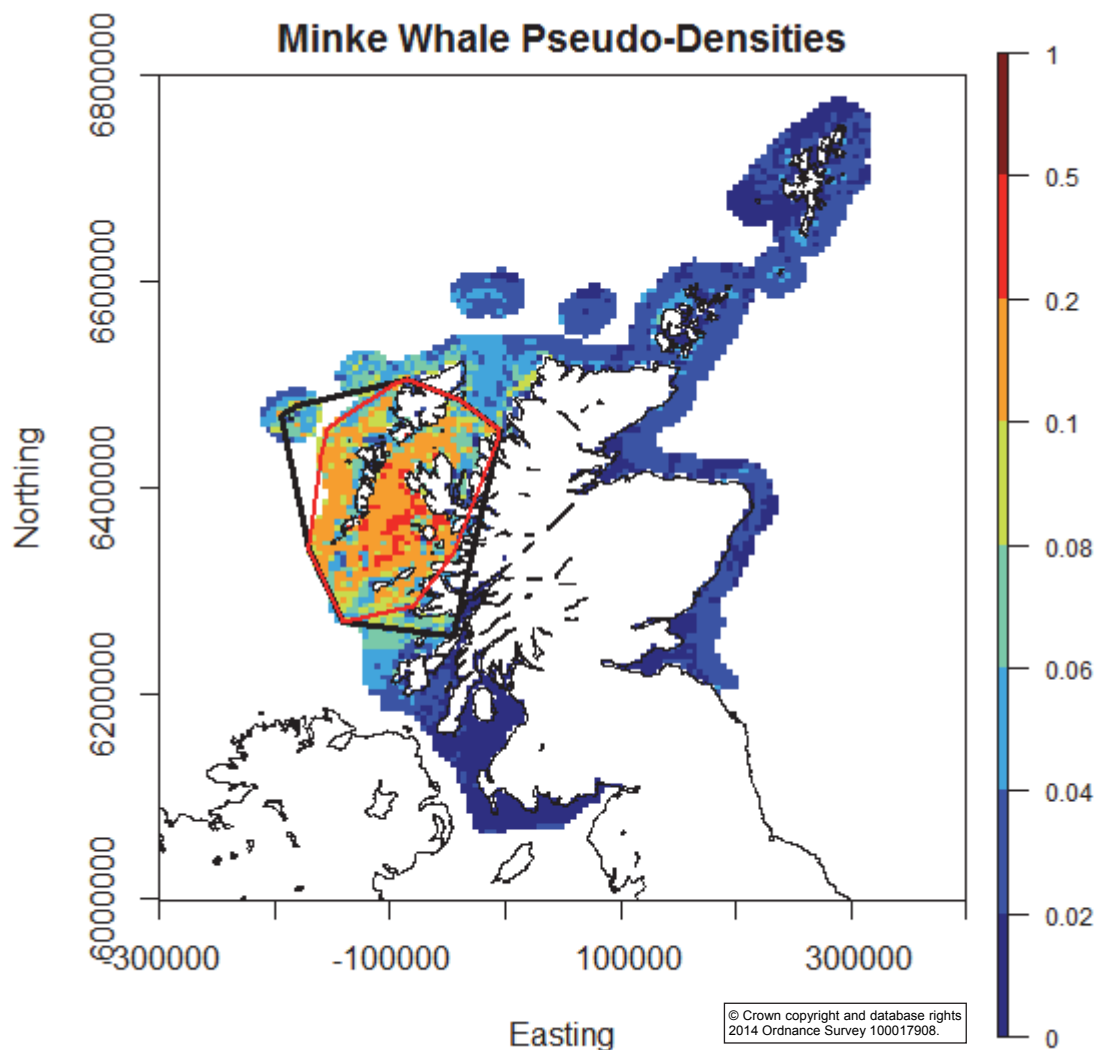


Figure 10. Hypothetical density (animals per km²) predictions on a 25 km² resolution grid of minke whale in summer in Scottish territorial waters. Note that such results would most probably come from a model with a variety of temporal/ spatial predictors. **N.B. This has been drawn for illustrative purposes and is not a formal proposal for an MPA.**

5.3 Joining cells

The next action is simply to have a rule for joining those cells together. There are a variety of methods for doing this. Within the UK there are several guidelines on marine boundary delineation good practice (Marine Scotland, 2011b; Natural England/ JNCC, 2010).

The simplest method is the generation of a convex hull that simply draws a line around the outer boundary cells of interest (e.g. the black line around the 0.1 animals/km² in Figure 10) to create a convex polygon thus ensuring all relevant sites are contained with a single boundary (as suggested by Natural England/ JNCC, 2010). There are a variety of methods of drawing convex hulls around data (see Avis *et al.*, 1997 for a review). Alpha hulls, a form of convex hull, have been recommended for use in IUCN Red List conservation assessments (IUCN, 2013) and may be appropriate in the current analysis. Algorithms are also available to build concave hulls (e.g. Edelsbrunner *et al.*, 1983; Park & Oh, 2012). This may be more desirable if, as per Marine Scotland (2011b) (Section 11.9), the aim is to “Draw boundaries as closely as possible around the feature(s) to support the MPA acting as a functional whole for the conservation of the features concerned.” Whilst this may be useful for defining an area based on predicted densities, to be consistent with the Scottish MPA Selection Guidelines, other contextual information will also need to be reviewed to determine whether the area is considered to be essential to key life cycle stages (SNH, pers. comm).

This may be complicated by the fact that some cells may be too disjunct from the core area to be included in the area of interest. So for example the cells to the north of Jura could be omitted, as well as the cells around St Kilda, to produce the convex hull seen in red in Figure 10. N.B. This discussion of connectivity is on a different scale from a conventional discussion about connectivity between MPAs (e.g. Marine Scotland, 2012). Here we are discussing cells within a proposed MPA. Of course there is no reason if there is a disjunct area that it could not be considered for designation as a separate MPA.

Other approaches to this problem are also possible. Gerrodette & Eguchi (2011) estimated the detection and density of long-beaked common dolphin *Delphinus capensis* at a set of gridded localities off the coast of Baja California. Simultaneously as part of a Bayesian model they estimated the number of dolphins within a depth contour increasing from the coast. There is no reason why the processes (as above) cannot be split up with the initial estimate of density followed by estimating the number of animals enclosed by some environmental feature. For example, a zone of interest could be designated around Lewis on the basis that 10 animals would be on average within a depth of 100 m of Lewis whereas 25 animals would be within 200m depth from Lewis etc. Uncertainty in such estimates is incorporated by considering the uncertainty in the predictions of abundance within the contour or less desirably by extending the region of interest. The seaward boundary does not have to be a depth contour but could instead be a given distance from the coast chosen to enclose a given percentage of the local population. For example in the ornithological literature a 1% population threshold is used to designate wetlands of interest i.e. 1% of an identified population regularly utilises the site (Jackson *et al.*, 2004). Although in the case of marine megafauna identifying distinct populations can be problematic.

Other similar methods to enclose a proportion of the population could include the use of maximum curvature (e.g. O'Brien *et al.*, 2012a,b for sea birds), a method that considers the gradient in density to identify the boundary at which the biggest change in the population occurs.

5.4 Consideration of additional data

The consideration of regions of conservation interest does not necessarily consider purely population/ biological criteria. There can be species diversity, sociological and economic

considerations as well. In a Scottish context socio-economic factors are considered after biological considerations (Marine Scotland, 2011a) and so will not be taken into account through habitat modelling or by SNH at the stage of providing advice on MPAs for cetaceans and basking shark. Marine Scotland will consider socio-economic information when deciding whether to designate an area as an MPA (SNH, pers.comm).

The methods outlined at Sections 5.2 and 5.3 could be adapted to take additional data into account by weighting any predictions for a particular cell with other features. There are a variety of “off the shelf” packages that act similarly but based on rather more formal theory. Alternatively this problem may not be seen as one that can be adequately dealt with by the application of an algorithm and must be considered as an iterative qualitative process with government agencies and stakeholders and the quantitative input is limited to the identification of areas of high density as given above. However, for completeness here, we briefly outline some of the leading quantitative approaches to this problem.

Marxan (Ball *et al.*, 2009) is a package that uses ecological and other criteria to designate whether units are in a reserve. Potential reserve units are added until a target is met (e.g. a population of 10,000) or units can be designated and a score calculated.

Zonation (Moilanen *et al.*, 2009) is another similar tool, in this case candidate cells are removed from a set of cells in such a way as to maintain conservation value (especially in a biodiversity context) whilst accounting for connectivity and other values. Like *Marxan*, *Zonation* has scholarly support (Moilanen *et al.*, 2005, 2009; Moilanen, 2007; Leathwick *et al.*, 2010). The consideration of connectivity may be different in the construction of marine reserves than terrestrial reserves, as large mobile marine animals may be able to move more readily between unsuitable areas than terrestrial animals. Other packages such as *C-Plan* (Pressey *et al.*, 2009) and *ConsNet* (Ciarlegio *et al.*, 2009) undertake similar analyses but the emphasis is on the preservation of biodiversity across networks rather than delineating regions of high density.

The methods have been formally compared (Delavenne *et al.*, 2012). *Marxan* and *Zonation* produce similar results, with *Zonation* more sensitive to connectivity (Delavenne *et al.*, 2012). Similarly *C-Plan* and *Marxan* have been compared with similar results (Carwardine *et al.*, 2007).

5.5 Summary on MPA delineation

Considerable thought needs to be given to MPA delineation. In the context of considering modelling approaches, an important decision is whether MPA delineation should be a qualitative process with quantitative inputs or a wholly algorithm driven process.

As expertise in spatial modelling is distinct from expertise in MPA delineation, it may be better to take the prediction surfaces and/ or response curves from the spatial models and apply a simple delineation criteria as detailed above based on density alone and leave the inclusion of other criteria to a later stage in the process, to be considered qualitatively or quantitatively as appropriate. This approach would allow SNH to consider quantitative information from habitat modelling alongside other ecological knowledge for each species in context of applying the MPA Selection Guidelines and providing advice on MPAs.

However, even using something as simple as relative density as a criterion will require expert judgement to inform the appropriate relative density threshold and provide guidance as to how uncertainty should be incorporated in the calculation. This in turn means it is difficult to indicate the potential scales involved before viewing the outputs of the models for the different species.

6. RECOMMENDATIONS

1. Preparing sightings data and explanatory covariate data for habitat modelling will take considerable time, even building upon efforts stemming from the JCP project. The cost in time and effort to organise these data should be considered along with benefits that might be derived from added data.
2. The following currently available dependent data should be considered:
 - a. Risso's dolphin: available data collated to inform the JCP project from Scottish territorial waters possibly augmented with JCP data from the Isle of Man. If only the west coast is of interest for this species then data should be restricted to this spatial extent.
 - b. White-beaked dolphin: available data collated to inform the JCP project from Scottish territorial waters initially. If the influence of sandeel presence is negligible (i.e. sandeel presence is not chosen as a predictor), then Scottish shelf waters (i.e. to 200 m depth) should be considered. Sandeel data are not available for the entire shelf.
 - c. Minke whale: available data collated to inform the JCP project from Scottish territorial waters but omitting winter data.
 - d. Basking shark: available data provided for the JCP project (where basking shark were recorded) from Scottish territorial waters, augmented with the Speedie data possibly additionally augmented with data from the Isle of Man but omitting winter data.

In all cases a small buffer zone may be applied to the area from which input data are collated, to avoid edge effects in the predictions.

3. The following additional data may prove useful CRRU, WCD and more recent HWDT data, although some work will be required to integrate these data sets into the existing JCP data resource framework.
4. GAMs should be used to create predicted relative density surfaces. It is likely that mixed model GAMs or GEE-GAMs will be used to manage the presumed spatial correlation in the data. It is possible for the data sparse species (i.e. Risso's dolphin) a model cannot be fitted, in which case an empirical approach to the identification of regions of relatively higher density should be undertaken.
5. Delineation of areas to be assessed as MPAs should be performed by drawing polygons using predicted relative animal densities or abundances for individual species. The resulting areas can then be considered by SNH alongside other contextual information (e.g. on behaviour) to inform their advice on areas to be considered for designation as Nature Conservation MPAs.

REFERENCES

- ABP Marine Environmental Research 2010. Seabed Kinetic Energy. EUSeamap. Report to JNCC.
- Anderwald, P., Evans, P.G.H., Dyer, R., Dale, A., Wright, P.J. & Hoelzel, A.R. 2012. Spatial scale and environmental determinants in minke whale habitat use and foraging. *Marine Ecology Progress Series*, **450**, 259-274.
- Augustin, N.H., Muggleston, M.A. & Buckland, S.T. 1996. An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology*, **33**, 339-347.
- Avis, D. Bremner, B. & Seidel R. 1997. How good are convex hull algorithms? *Computational Geometry*, **7**, 265-301.
- Azzellino, A., Panigada, S., Lanfredi, C., Zanardelli, M., Aioldi, S., & Notarbartolo di Sciara, G. 2012. Predictive habitat models for managing marine areas: Spatial and temporal distribution of marine mammals within the Pelagos Sanctuary (Northwestern Mediterranean sea). *Ocean and Coastal Management*, **67**, 63-74.
- Bailey, H. & Thompson P.M. 2009. Using marine mammal habitat modelling to identify priority conservation zones within a marine protected area. *Marine Ecology Progress Series*, **378**, 279-287.
- Baird, R.W. 2009. Risso's dolphin *Grampus griseus*. In: Perrin, W.F., Würsig, B. & Thewissen, J.G.M. (eds) *Encyclopedia of Marine Mammals*. London: Elsevier, pp. 975-976.
- Ball, I.R., Possingham, H.P. & Watts, M. 2009. *Marxan* and relatives: Software for spatial conservation prioritisation. In: Moilanen, A., Wilson, K.A. & Possingham, H.P. (eds) *Spatial conservation prioritisation: Quantitative methods and computational tools*. Oxford: Oxford University Press, pp. 185-195.
- Becker, E.A., Foley, D.G., Forney, K.A., Barlow, J., Redfern, J.V., & Gentemann, C.L. 2012. Forecasting cetacean abundance patterns to enhance management decisions. *Endangered Species Research*, **16**, 97-112.
- Bellier, E., Monestiez, P. & Guinet, C. 2010. Geostatistical Modelling of Wildlife Populations: A Non-stationary Hierarchical Model for Count Data. In: Atkinson, P.M. and Lloyd, C.D. (eds.), *geoENV VII – Geostatistics for Environmental Applications. Quantitative Geology and Geostatistics*, **16**, Dordrecht: Springer, pp. 1 - 12.
- Booth, C.G. & Hammond, P.S. 2011 *A comparison of different techniques for mapping cetacean habitats*. Report to the Joint Nature Conservation Committee under Project Specification C0902070240
- Breiman, L. Friedman, J., Stone, C.J. & Olshen, R.A. 1984. *Classification and Regression Trees*. London: Chapman & Hall.
- Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. & Thomas, L. 2001. *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*. Oxford: Oxford University Press.

- Buckland, S.T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L. & Thomas L. 2004. *Advanced Distance Sampling*. Oxford: Oxford University Press.
- Burroughs, P.A. & McDonell, R.A., 1998. *Principles of Geographical Information Systems*. New York: Oxford University Press.
- Busby, J.R. 1991 BIOCLIM - A Bioclimatic Analysis and Prediction System. *In: Margules, C.R. & Austin M.P. (eds) Nature Conservation: Cost Effective Biological Surveys and Data Analysis*. Canberra: CSIRO. pp. 64-68.
- Cañadas, A. & Hammond, P.S. 2006. Abundance and habitat preferences of the short-beaked common dolphin *Delphinus delphis* in the southwestern Mediterranean: implications for conservation. *Endangered Species Research*, **309**, 309-331.
- Cañadas, A., Sagarminaga, R., Urquiola, E. & Hammond, P.S. 2005. Habitat preference modelling as a conservation tool: proposals for marine protected areas for cetaceans in southern Spanish waters. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **15**, 495-521.
- Canning, S.J., Santos, M.B., Reid, R.J., Evans, P.G.H., Sabin, R.C., Bailey N. & Pierce, G.J. 2008. Seasonal distribution of white-beaked dolphins (*Lagenorhynchus albirostris*) in UK waters with new information on diet and habitat use. *Journal of the Marine Biological Association of the United Kingdom*, **88**, 1159-1166.
- Carpenter, G., Gillison, A.N., Winter, J. 1993. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation*, **2**, 667-680.
- Carwardine J., Rochester W.A., Richardson K.S., Williams K.J., Pressey R.L. & Possingham H.P. 2007. Conservation planning with irreplaceability: does the method matter? *Biodiversity and Conservation*, **16**, 245-258.
- Choudhary, S., Dey, S., Dey, S., Sagar, V., Nair, T. & Kelkar, N. 2012. River dolphin distribution in regulated river systems: implications for dry-season flow regimes in the Gangetic basin. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **22**, 11-25.
- Ciarleglio, M., Wesley Barnes, J. & Sarkar, S. 2009. *ConsNet*: new software for the selection of conservation area networks with spatial and multi-criteria analyses. *Ecography*, **32**, 205-209.
- Cooper, R., Long, D., Doce, D., Green, S. & Morando, A. 2010. Creating and assessing a sediment data layer for UKSeaMap 2010. *British Geological Survey Commercial Report*, CR/09/168. 29pp.
- Cotton, P.A., Sims, D.W., Fanshawe, S. & Chadwick, M. 2005. The effects of climate variability on zooplankton and basking shark relative abundance off southwest Britain. *Fisheries Oceanography*, **14**, 151-155.
- Crainiceanu, C., Ruppert, D., Carroll, R. Joshi, A. & Goodner, B. 2007. Spatially adaptive Bayesian penalized splines with heteroscedastic errors. *Journal of Computational and Graphical Statistics*, **16**, 265-288.
- Cressie, N. & Wikle, C.K. 2011. *Statistics for Spatio-Temporal Data*. Oxford: Wiley-Blackwell.

De Boer M.N. 2010 Spring distribution and density of minke whale *Balaenoptera acutorostrata* along an offshore bank in the central North Sea. *Marine Ecology Progress Series*, **408**, 265-274.

de Segura, A.G., Hammond, P.S. & Raga, J.A. 2008. Influence of environmental factors on small cetacean distribution in the Spanish Mediterranean. *Journal of the Marine Biological Association of the UK*, **88**, 1185-1192.

Delavenne, J., Metcalfe, K., Smith, R.J., Vaz, S., Martin, C.S., Dupuis, L., Coppin, F. & Carpentier, A. 2012. Systematic conservation planning in the eastern English Channel: comparing the *Marxan* and *Zonation* decision-support tools. *ICES Journal of Marine Science*, **69**, 75-83.

Edelsbrunner, H., Kirkpatrick, D.G. & Seidel, R. 1983. On the shape of a set of points in the plane. *IEEE Transactions on Information Theory*, **29**, 551-559.

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. Overton, J., Townsend Peterson, A., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

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*). *Biological Conservation*, **143**, 267-279.

Faraway, J.J. 2005. *Linear Models with R*. London: Chapman & Hall.

Faraway, J.J. 2006. *Extending the Linear Model with R*. London: Chapman & Hall.

Ferrier, S. & Watson, G. 1997. *An evaluation of the effectiveness of environmental surrogates and modelling techniques in predicting the distribution of biological diversity*. Canberra: Environment Australia.

Ferrier, S., Drielsma, M., Manion, G. & Watson, G. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity: the north-east New South Wales experience. I. Species-level modelling. *Biodiversity & Conservation*, **11**, 2275–2307.

Gerrodette, T. & Eguchi, T. 2011. Precautionary design of a marine protected area based on a habitat model. *Endangered Species Research*, **15**, 159-166.

Ghisletta, P. & Spini, D. 2004. An Introduction to generalized estimating equations and an application to assess selectivity effects in a longitudinal study on very old individuals. *Journal of Educational & Behavioral Statistics*, **29**, 421-437.

Goetz, K.T., Rugh, D.J., Read, A.J., & Hobbs, R.C. 2007. Habitat use in a marine ecosystem: beluga whales *Delphinapterus leucas* in Cook Inlet, Alaska. *Marine Ecology Progress Series* **330**, 247-256.

Goetz, K.T., Montgomery, R.A., VerHoef, J.M., Hobbs R.C. & Johnson, D.S. 2012. Identifying essential summer habitat of the endangered beluga whale *Delphinapterus leucas* in Cook Inlet, Alaska. *Endangered Species Research*, **16**, 135-147.

- Haining, R. 1990. *Spatial data analysis in the social and environmental sciences*. Cambridge University Press. Cambridge.
- Hardin, J.W. & Hilbe, J.M. 2003. *Generalized estimating equations*. London: Chapman & Hall.
- Hastie, G.D., Wilson, B. & Thompson, P.M. 2003. Fine-scale habitat selection by coastal bottlenose dolphins: application of a new land-based video-montage technique. *Canadian Journal of Zoology*, **81**, 469–478
- Hastie, T.J. & Tibshirani, R.J. 1990. *Generalized Additive Models*. London: Chapman & Hall.
- Hedley, S.L. & S.T. Buckland. 2004. Spatial models for line transect sampling. *Journal of Agricultural, Biological and Environmental Statistics*, **9**, 181-199
- Hedley, S.L., Buckland, S.T. & Borchers. D.L. 2004. Spatial distance sampling models. In: Buckland, S.T., D.R. Anderson, D.R. Burnham, K.P. Laake, D.L. Borchers & L. Thomas (eds) *Advanced Distance Sampling*. Oxford: Oxford University Press, pp. 48-70.
- Hirzel, A.H., Hausser, J., Chessel, D. & Perrin, N. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology*, **83**, 2027-2036.
- Isojunno S., Matthiopoulos J. & Evans P.G.H. 2012. Harbour porpoise habitat preferences: robust spatio-temporal inferences from opportunistic data. *Marine Ecology Progress Series*, **448**, 155-170.
- IUCN Standards and Petitions Subcommittee. 2013. Guidelines for Using the IUCN Red List Categories and Criteria. Version 10.1. Prepared by the Standards and Petitions Subcommittee.
- Jackson, S.F. Kershaw, M. & Gastona, K.J. 2004. The performance of procedures for selecting conservation areas: waterbirds in the UK. *Biological Conservation*, **118**, 261-270.
- Krivobokova, T., Crainiceanu, C.M. & Kauermann.G. 2007. Fast adaptive penalized splines. *Journal of Computational and Graphical Statistics*, **17**, 1-20.
- Lambert, D. 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**, 1-14.
- Leathwick, J.R., Moilanen, A., Ferrier, S. & Julian. K. 2010. Complementarity-based conservation prioritization using a community classification, and its application to riverine ecosystems. *Biological Conservation*, **143**, 984-991.
- MacLeod, C.D., Weir, C.R., Pierpoint, C. & Harland, E.J. 2007. The habitat preferences of marine mammals west of Scotland (UK). *Journal of the Marine Biological Association of the United Kingdom*, **87**, 157-164.
- MacLeod, C.D., Mandleberg, L., Schweder, C., Bannon S.M. & Peirce, G.J. 2008. A comparison of approaches for modelling the occurrence of marine animals *Hydrobiologia*, **612**, 21-32.
- MacLeod, C.D., Santos, M.B. & Pierce, G.J. 2014. Can habitat modelling for the octopus

Eledone cirrhosa help identify key areas for Risso's dolphin in Scottish waters? *Scottish Natural Heritage Commissioned Report No. 530*.

Macleod, C.D. & Zuur, A. 2005. Habitat utilization by Blainville's beaked whales off Great Abaco, northern Bahamas, in relation to seabed topography. *Marine Biology*, **147**, 1-11.

Marine Scotland. 2011a. *A strategy for marine nature conservation in Scotland's seas*. Marine Scotland. Scottish Government.

Marine Scotland. 2011b. *Marine Protected Areas in Scotland's seas. Guidelines on the selection of MPAs and development of the MPA network*. Scottish Government.

Marine Scotland. 2012. *Connectivity within the Scottish MPA network*. The 4th Marine Protected Areas Workshop Working Paper.

Marubini, F., Gimona, A., Evans, P.G.H. Wright, P.J., & Pierce, G.J. 2009. Habitat preferences and interannual variability in occurrence of the harbour porpoise *Phocoena phocoena* off northwest Scotland. *Marine Ecology Progress Series*, **381**, 297-310.

McCullagh, P. & Nelder, J.A. 1989. *Generalized Linear Models*. London: Chapman & Hall.

Miller, P.I., Xu, W. & Lonsdale, P. 2014. Seasonal shelf-sea front mapping using satellite ocean colour to support development of the Scottish MPA network. *Scottish Natural Heritage Commissioned Report No. 538*.

Miller, P.I., Christodoulou, S. & Saux-Picart, S. 2010. *Oceanic thermal fronts from Earth observation data – a potential surrogate for pelagic diversity*. Report to the Department of Environment, Food and Rural Affairs. Defra Contract No. MB102. Plymouth Marine Laboratory, subcontracted by ABPmer, Task 2F, Report No. 20., p. 24.

Moilanen, A. 2007. Landscape zonation, benefit functions and target-based planning: Unifying reserve selection strategies. *Biological Conservation*, **134**, 571-579.

Moilanen, A., Franco, A.M.A., Early, R., Fox, R., Wintle, B.A. & Thomas C.D. 2005. Prioritising multiple-use landscapes for conservation: methods for large multi-species planning problems. *Proceeding of the Royal Society of London B.*, **272**, 1885-1891.

Moilanen, A., Kujala, H. & Leathwick, J. 2009. The *Zonation* framework and software for conservation prioritization. In: Moilanen, A., Wilson, K.H. & Possingham, H.P. (eds) *Spatial Conservation Prioritization*. Oxford: Oxford University Press pp.196-210.

Monestiez, P., Dubroca, L., Bonninc, E., Durbecc, J.-P., & Guinet, C. 2006. Geostatistical modelling of spatial distribution of *Balaenoptera physalus* in the Northwestern Mediterranean Sea from sparse count data and heterogeneous observation efforts. *Ecological Modelling*, **193**, 615-628.

Natural England/ Joint Nature Conservation Committee 2010. *The Marine Conservation Zone Project: Ecological Network Guidance*. Natural England and JNCC: Sheffield & Peterborough.

O'Brien, S.H., Win, I., Parsons, M., Allcock, Z. & Reid, J.B. 2012a. The numbers and distribution of inshore waterbirds along the south Cornwall coast during winter. Draft *JNCC Report to Natural England*, Peterborough, UK.

- O'Brien, S.H., Webb, A., Brewer, M.J. & Reid J.B. 2012b. Use of kernel density estimation and maximum curvature to set Marine Protected Area boundaries: Identifying a Special Protection Area for wintering red-throated divers in the UK. *Biological Conservation*, **156**, 15-21.
- Oppel, S., Meirinho, A., Ramírez, I. Gardner, B., O'Connell, A.F. Miller, P.I. & Louzao, M. 2012. Comparison of five modelling techniques to predict the spatial distribution and abundance of seabird. *Biological Conservation*, **156**, 94 - 104.
- Okka, E. J., Leopold, M.F., Meesters, E.H.W.G. & Smeenk, C. 2010. Are white-beaked dolphins *Lagenorhynchus albirostris* food specialists? Their diet in the southern North Sea. *Journal of the Marine Biological Association of the United Kingdom*, **90**, 1501-1508.
- Panigada, S., Zanardelli, M., MacKenzie, M., Donovan C., Mélin, F. & Hammond, P.S. 2008. Modelling habitat preferences for fin whales and striped dolphins in the Pelagos Sanctuary (Western Mediterranean Sea) with physiographic and remote sensing variables. *Remote Sensing of Environment*, **112**, 3400-3412.
- Parker, H.W. & Boeseman, M. 1954. The basking shark (*Cetorhinus maximus*) in winter. *Proceedings of the Zoological Society London*, **124**, 185-194.
- Park, J.-S. & Oh, S.-J. 2012. A new concave hull algorithm and concaveness measure for *n*-dimensional datasets. *Journal of Information Science and Engineering*, **28**, 587-600
- Paxton, C.G.M., Scott-Hayward, L., Mackenzie, M., Rexstad, E. & Thomas, L. 2013. *Revised Phase III Data Analysis of Joint Cetacean Protocol Data Resource*. Report to JNCC (unpublished).
- Philips, S.J., Anderson, R.P. & Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231-259.
- Pierce, G.J., Caldas, M., Cedeira, J., Santos, M.B., Llavona, A., Covelo, P., Martinez, G., Torres, J., Sacau, M. & Alfredo López, A. 2010. Trends in cetacean sightings along the Galician coast, north-west Spain, 2003–2007, and inferences about cetacean habitat preferences. *Journal of the Marine Biological Association of the UK*, **90**, 1547-1560.
- Pinheiro, J.C. & Bates, D.M. 2000. *Mixed-Effects Models in S and Splus*. New York: Springer.
- Pintore, A., Speckman, P. & Holmes, C. 2006. Spatially adaptive smoothing splines. *Biometrika*, **93**, 113-125.
- Pirotta, E., Matthiopoulos, J., MacKenzie, M., Scott-Hayward, L. & Rendell, R. 2011. Modelling sperm whale habitat preference: a novel approach combining transect and follow data. *Marine Ecology Progress Series*, **436**, 257-272.
- Pressey, R.L., Watts, M.E., Barrett, T.W. & Ridges, M.J. 2009. The C-plan conservation planning system: origins, applications, and possible futures. In: Moilanen, A., Wilson, K.H. & Possingham, H.P. (eds) *Spatial Conservation Prioritization*. Oxford: Oxford University Press pp. 211-234.
- Ramsay, T.O. 2002. Spline smoothing over difficult regions. *Journal of the Royal Statistical Society B.*, **64**, 307-319.

Read, A.J., Borchers, D.L., Cummings, E.W., McAlarney, R.J., McLellan, W.A., Nilsson, P., Pabst, D.A., Paxton, C.G.M. Urian, K.W. & Waples, D.M. (submitted). Occurrence, Distribution and Abundance of Cetaceans in Onslow Bay, North Carolina, USA. *Marine Ecology Progress Series*.

Redfern, J.V., Ferguson, M.C., Becker, E.A., Hyrenbach, K.D., Good, C., Barlow, J., Kaschner, K., Baumgartner, K.M.F., Forney, K.A., Ballance, L.T., Fauchald, P., Halpin, P., Hamazaki, T., Pershing, A.J., Qian, S.S., Read, A., Reilly, S.B., Torres, L. & Werner, F. 2006. Techniques for cetacean-habitat modelling. *Marine Ecology Progress Series*, **310**, 271-295.

Robertson, M.P., Caithness, N. & Villet M.H. 2001. A PCA-based modelling technique for predicting environmental suitability for organisms from presence records. *Diversity and Distributions*, **7**, 15-27.

Robertson, M.P., Peter, C.L., Villeta, M.H. & Ripley, B.S. 2003 Comparing models for predicting species' potential distributions: a case study using correlative and mechanistic predictive modelling techniques. *Ecological Modelling*, **164**, 153-167.

Ruppert, D. 2000. Spatially-adaptive penalties for spline fitting. *Australian & New Zealand Journal of Statistics*, **42**, 205-223.

Scott-Hayward, L. A. S., Mackenzie, M., Donovan, C., Walker, C. & Ashe, E. In press. Complex Region Spatial Smoother (CRSS). *Journal of Computational & Graphical Statistics*

Scottish Natural Heritage and the Joint Nature Conservation Committee. 2012. Advice to the Scottish Government on the selection of Nature Conservation Marine Protected Areas (MPAs) for the development of the Scottish MPA network. *Scottish Natural Heritage Commissioned Report No. 547*.

Scottish Natural Heritage. 2012. *Marine Protected Areas – Cetaceans and Basking Shark*. The 4th Marine Protected Areas Workshop Working Paper.

Sims, D.W. 2008. Sieving a living: a review of the biology, ecology and conservation status of the plankton-feeding basking shark *Cetorhinus maximus*. *Advances in Marine Biology*, **54**, 171-220.

Sims, D.W., Southall, E.J., Quayle, V.A. & Fox, A.M. 2000. Annual social behaviour of basking sharks associated with coastal front areas. *Proceedings of the Royal Society of London B.*, **267**, 1897-1904.

Skomal, G.B., Wood, G. & Caloyianis, N. 2004. Archival tagging of a basking shark, *Cetorhinus maximus*, in the western North Atlantic. *Journal of the Marine Biological Association of the UK*, **84**, 795-799.

Stein, A. & Corsten, L.C.A. 1991 Universal kriging and cokriging as a regression procedure. *Biometrics*, **47**, 575-587.

Thorne, L.H., Johnston, D.W., Urban, D.L., Tyne, J., Bejder, L., Baird, R.W., Yin, S., Rickards, S.H., Deakos, M.H., Mobley, J.R., Pack, A.A. & Hill, M.C. 2012. Predictive modeling of spinner dolphin (*Stenella longirostris*) resting habitat in the main Hawaiian islands. *PLoS ONE*, **7**, e43167.

- Upton, G. & Cook, I. 2002. *Oxford Dictionary of Statistics*. Oxford: Oxford University Press.
- Walker, C., Mackenzie, M., Donovan, C. & O'Sullivan, M. 2010. SALSA - a Spatially Adaptive Local Smoothing Algorithm. *Journal of Statistical Computation and Simulation*, **81**, 179-191.
- Walker, C., Mackenzie, M., Donovan, C., Scott-Hayward, L.A.S. & O'Sullivan, M. In prep. Spatially adaptive smoothing in two/three dimensions: an update to SALSA.
- Wang, H. & Ranalli, M. G. 2007. Low-rank smoothing splines on complicated domains. *Biometrics*, **63**, 209-217.
- Witt, M.J., Hardy, T., Johnson, L., McClellan, C.M., Pikesley, S.K., Ranger, S., Peter B., Richardson, P.B., Solandt, J.-L., Speedie C., Williams R. & Godley, B.J. 2012. Basking sharks in the northeast Atlantic: spatio-temporal trends from sightings in UK waters. *Marine Ecology Progress Series*, **459**, 121-134.
- Wood, S.N. 2006. *Generalized Additive Models: An Introduction with R*. Chapman & Hall. London.
- Wood, S.N., Bravington, M.V. & Hedley, S.L. 2008. Soap film smoothing. *Journal of the Royal Statistical Society B*, **70**, 931-935.
- Wright, P.J., Jensen, H. & Tuck, I. 2000. The influence of sediment type on the distribution of the lesser sandeel, *Ammodytes marinus*. *Journal of Sea Research*, **44**, 243-256.
- Yue, Y.(R.), Loh, J.M. & Lindquist, M.A. 2010. Adaptive spatial smoothing of fMRI images. *Statistics and its Inference*, **3**, 3-13.
- Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. & Smith G.M. 2009. *Mixed Effects Models and Extensions in Ecology in R*. New York: Springer.

www.snh.gov.uk

© Scottish Natural Heritage 2014
ISBN: 978-1-85397-959-0

Policy and Advice Directorate, Great Glen House,
Leachkin Road, Inverness IV3 8NW
T: 01463 725000

You can download a copy of this publication from the SNH website.



Scottish Natural Heritage
Dualchas Nàdair na h-Alba

All of nature for all of Scotland
Nàdar air fad airson Alba air fad