Vol. 15: 39–52, 2011
doi: 10.3354/esr00367

ENDANGERED SPECIES RESEARCH
Endang Species Res

Published online October 21

Contribution to the Theme Section: 'Beyond marine mammal habitat modeling'

Spatial models of sparse data to inform cetacean conservation planning: an example from Oman

Peter J. Corkeron[1,2,3,8,*], Gianna Minton,[4,5], Tim Collins[4,6], Ken Findlay[7], Andrew Willson[4], Robert Baldwin[4]

[1]Integrated Statistics, Woods Hole, Massachusetts 02543, USA
[2]Bioacoustics Research Program, Cornell Lab of Ornithology, Ithaca, New York 14850, USA
[3]The New England Aquarium, Central Wharf, Boston, Massachusetts 02110-3399, USA
[4]Environment Society of Oman, Ruwi, Sultanate of Oman
[5]Sarawak Dolphin Project, Institute of Biodiversity and Environmental Conservation, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia
[6]Ocean Giants Program, Wildlife Conservation Society, Bronx, New York 10460-1099, USA
[7]MaRe, Oceanography Department, University of Cape Town, Rondebosch 7701, South Africa
[8]*Present address:* NOAA Northeast Fisheries Science Center, Woods Hole, Massachusetts 02543, USA

ABSTRACT: Habitat models are tools for understanding the relationship between cetaceans and their environment, from which patterns of the animals' space use can be inferred and management strategies developed. Can working with space use alone be sufficient for management, when habitat cannot be modeled? Here, we analyzed cetacean sightings data collected from small boat surveys off the coast of Oman between 2000 and 2003. The waters off Oman are used by the Endangered Arabian Sea population of humpback whales. Our data were collected primarily for photo-identification, using a haphazard sampling regime, either in areas where humpback whales were thought to be relatively abundant, or in areas that were logistically easy to survey. This leads to spatially autocorrelated data that are not amenable to analysis using standard approaches. We used quasi-Poisson generalized linear models and semi-parametric spatial filtering to assess the distribution of humpback and Bryde's whales in 3 areas off Oman relative to 3 simple physiographic variables in a survey grid. Our analysis focused on the spatial eigenvector filtering of models, coupled with the spatial distribution of model residuals, rather than just on model predictions. Spatial eigenvector filtering accounts for spatial autocorrelation in models, allowing inference to be made regarding the relative importance of particular areas. As an exemplar of this approach, we demonstrate that the Dhofar coast of southern Oman is important habitat for the Arabian Sea population of humpback whales. We also suggest how conservation planning for mitigating impacts on humpback whales off the Dhofar coast could start.

KEY WORDS: Spatial eigenvector models · Spatial planning · Marine Protected Area · Generalized linear models · Oman · Whales

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

Habitat models, spatial models, and conservation planning

Habitat models have evolved from Hutchinson's (1957) concept of niche as environmental hyperspace (Basille et al. 2008). In general, habitat models are used to inform cetacean management by attempting to understand the relationships between cetaceans and their environment, from which inference is then drawn on space use (e.g. Johnston et al. 2005, Cañadas & Hammond 2008, Redfern et al. 2008, Stafford et al. 2009). In order to develop spatially-based approaches to cetacean conservation, is this the only way forward?

*Email: peter.corkeron@noaa.gov

© Inter-Research 2011 · www.int-res.com

Science is generally used to mitigate inadvertent anthropogenic mortality of cetaceans in a series of steps: estimating the abundance of the population of interest; determining population structure and boundaries (e.g. Taylor et al. 2000); estimating anthropogenic mortality of the population; then modeling the likely sustainability of this mortality (e.g. Wade 1998). With these scientific inputs, managers and stakeholders can devise measures that should reduce mortality to a level that will allow adequate mitigation within the social and cultural norms of the people using the area over which the cetacean population ranges.

The view of oceans underpinning this process is one in which the marine environment was viewed as generally undisturbed, with patches of impact, which has recently been questioned (e.g. Crowder et al. 2006). This has led, in some nations, to the view that marine zoning, i.e. the marine equivalent of terrestrial conservation planning (Margules & Pressey 2000), is a more appropriate paradigm to adopt (e.g. Fernandes et al. 2005). When providing data on habitat use by cetaceans to inform spatially-explicit conservation planning (e.g. Parra et al. 2006a), knowing the absolute abundance of cetaceans becomes less crucial than in the traditional approach. Note also that knowledge of the ecological (or social) processes driving spatial distribution of whales, although very useful, is not necessarily more important than simply having a well-quantified understanding of what the spatial distribution is. At this point, the interaction between model outputs of space use, and the management and policy milieu of the area of interest, is what matters (e.g. Grech & Marsh 2008).

Spatial autocorrelation

Spatial autocorrelation (i.e. the closer samples are, the more similar they are) introduces challenges when making inference from models, as standard errors of fixed effects from linear models are likely underestimated (Dormann 2009). Recently, several modeling techniques that can account for spatial autocorrelation have been brought to the attention of ecologists (Dormann et al. 2007). Even when working with data from surveys specifically designed to produce distance-sampled estimates of abundance (e.g. Gómez de Segura et al. 2007, Redfern et al. 2008), spatial autocorrelation arising from niche-related or social factors must be considered. However, when working from vessels of opportunity, or from field data where the principal aims did not require systematic or random sampling for survey coverage, another source of autocorrelation needs consideration, viz. that introduced by the haphazard sampling regime.

Note that here we are using 'haphazard' technically to refer to sampling that is not explicitly randomized, nor a fixed sampling regime starting from a randomized point. It is generally assumed in these instances that simply accounting for effort is sufficient (e.g. Macleod et al. 2004). However, papers analyzing haphazardly-collected data that then account for effort rarely analyze model residuals to demonstrate that spatial autocorrelation has been handled satisfactorily by the model used (e.g. Macleod et al. 2004).

Arabian Sea humpback whales

Here we address this problem by modeling the distribution of the Arabian Sea population of humpback whales *Megaptera novaeangliae* off the coast of the Sultanate of Oman (hereafter, Oman). The Arabian Sea humpback whales are the only known non-migratory population of humpback whales, and were designated as an Endangered subpopulation in the 2008 revision of the IUCN Red List for cetacean species (Minton et al. 2008). The current population, estimated to number less than 100 individuals, does not appear to be recovering from depletion due to whaling in the 1960s (Minton et al. in press). Data from photo-identified individuals (Minton et al. 2010) and genetics (Rosenbaum et al. 2009) demonstrate that this population is isolated from the nearest neighboring Indian Ocean populations.

The distribution of the Arabian Sea population of humpback whales is assumed to include the waters of other nations (Fig. 1), particularly the Islamic Republic of Pakistan, India, the Islamic Republic of Iran, and the Republic of Yemen (Minton et al. 2010), but dedicated survey effort in these nations' waters is either absent or very limited (e.g. Braulik et al. 2010). Thanks to the combination of historical records (Mikhalev 2000), continued research efforts in recent years, a history of attendance at International Whaling Commission meetings, and the establishment of a well networked non-government organization as a platform to support conservation-based research, Oman has become the range state of primary importance for the research and protection of the Arabian Sea population of humpback whales.

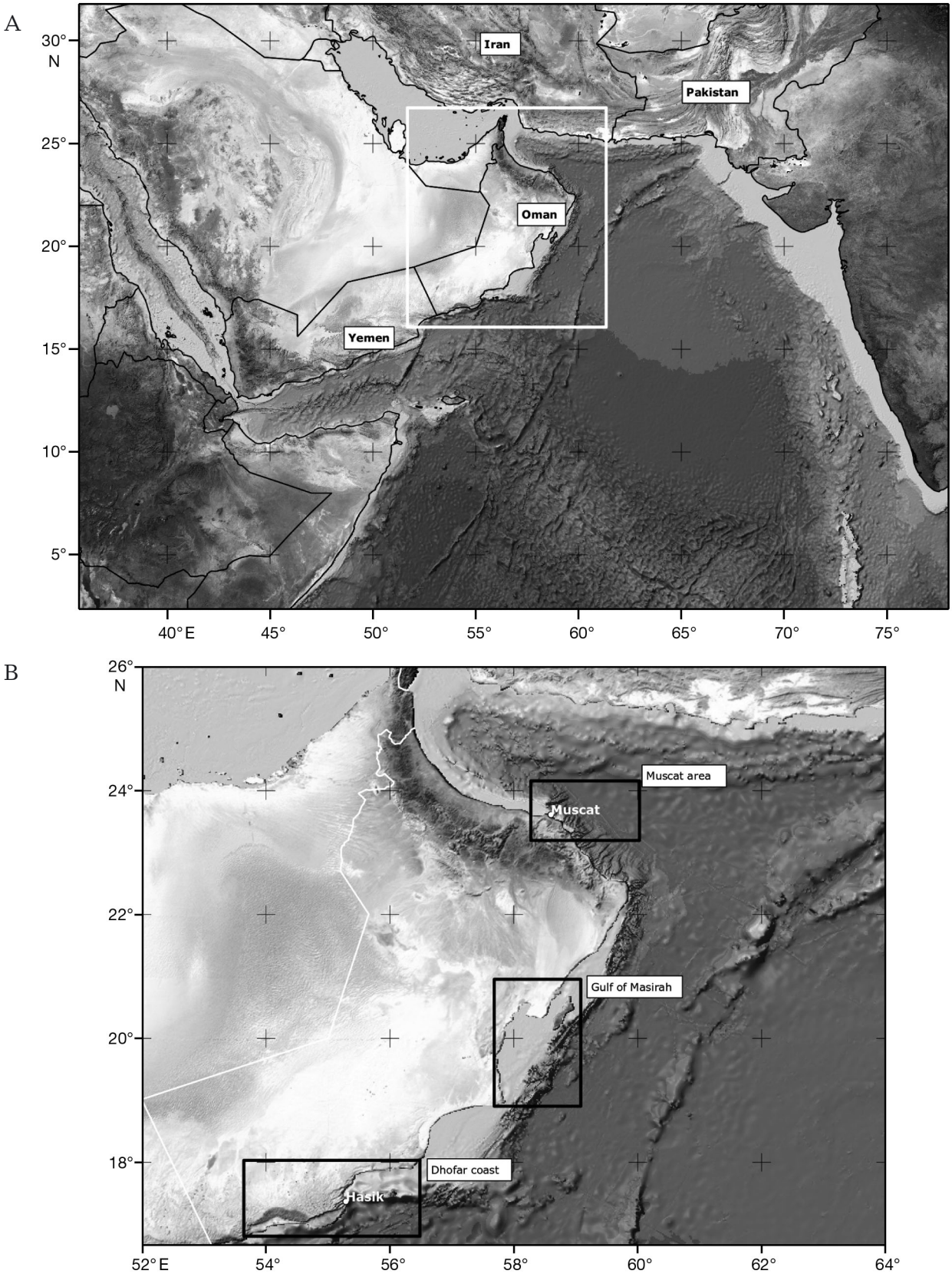


Fig. 1. Study area, showing bottom topography. (A) Arabian Sea and environs, showing countries named in the text. White box delineates area shown in detail in (B). (B) Study area in the waters off the Sultanate of Oman in detail. Black boxes delineate the 3 study areas described in the text

The available data present several challenges for modeling. As the Arabian Sea population of humpbacks is small, there are few sightings of individuals, and those sightings are clustered. The data were collected by small boat, primarily for photo-identification and genetic sampling, so surveys were haphazard, with coverage affected by logistical constraints, and with a concentration on areas where whales were likely to be encountered. Furthermore, as surveys were conducted by researchers on a volunteer basis, survey timing had to fit around researchers' normal occupations, and were spread out over 4 yr (although they were timed to coincide with the likely presence of humpback whales as indicated by historic whaling records; Mikhalev 2000).

For these reasons, previous publications from these surveys have either been descriptive (e.g. Minton et al. 2011), or provided results from photo-identification (e.g. Minton et al. 2010, in press) or genetic studies (e.g. Rosenbaum et al. 2009). Our aim in this paper was to develop a spatial model from the Oman sightings data in order to identify the areas of greatest relative abundance of humpback whales off the Oman coast. We could not use biological oceanographic predictors for our model, given the timing over which data were collected and the relatively few sightings in each year (Minton et al. 2011), and because oceanographic processes off Oman are driven largely by monsoonal conditions. The timing of the monsoon, and its strength, varies between years (Burkill 1999). That being so, we chose simple physiographic predictor variables for our model.

This means that although we are developing *spatial* models, we are not constructing *habitat* models from our data. Instead, we show how relatively simple spatial models, based on data that violate most models' assumptions of spatial independence, can still provide the scientific foundation for management action. In doing so, we aim to demonstrate how others with similar sightings data and issues of spatial autocorrelation can extract statistically robust, meaningful results that can be used to inform conservation measures. As we have survey data for several cetacean species, we used the differences between humpback and Bryde's whales *Balaenoptera* sp. in model results to begin to differentiate between spatial autocorrelation caused by species-specific ecological factors from those due to haphazard sampling.

MATERIALS AND METHODS

Study area and field techniques

Table 1 shows a summary of survey dates and distances covered on effort. Fig. 2 shows effort within the study grid (see below). Full details of the survey design and field methods are given by Minton et al. (2011). Small boat surveys (most frequently a 6.5 m rigid-hulled inflatable boat) were run between January 2000 and October 2003 in 3 areas: off Muscat, the Gulf of Masirah, and the Dhofar coast (Fig. 1). Surveys were generally conducted on a monthly basis in the Muscat region through most of the study period; the Gulf of Masirah was surveyed in October and November, and the Dhofar coast was surveyed in February and March. As the research focus was on humpback whales, areas of known or suspected humpback whale distribution were targeted, based on historical data (e.g. Wray & Martin 1983, Mikhalev 2000) and anecdotal reports. The excep-

Table 1. Dates and locations of small boat surveys in Oman. Effort indicates time spent actively searching for whales and excludes time spent working with whales, in transit, or on breaks

Survey area	Survey dates	Effort hours
Muscat		
Monthly surveys	15 Mar 2001 – 15 Jul 2003	104.21
Dhofar		
Hallaniyat Islands	15–24 Jan 2000, 8–21 Feb 2000	63.5
Dhofar	9–22 Feb 2001	34.26
Dhofar	10 Feb – 2 Mar 2002	62.37
Hasik Bay	24–26 Jun 2002	4.32
Sharbitat and Hallaniyats	17–20 Nov 2002	36.83
Dhofar	24 Feb – 19 Mar 2003	116.31
Dhofar (Hasik only)	15–17 May 2003	2.17
Total		319.76
Gulf of Masirah		
N. Gulf of Masirah	15–17 Oct 2000	11
Gulf of Masirah	4–27 Oct 2001	83.15
Gulf of Masirah	24 Oct – 16 Nov 2002	58.2
Total		152.35
Other areas		
Ras al Hadd	30 Mar – 2 Apr 2001	8.13
Shore-based observations		
Duqm	10–13 Jun 2001	25

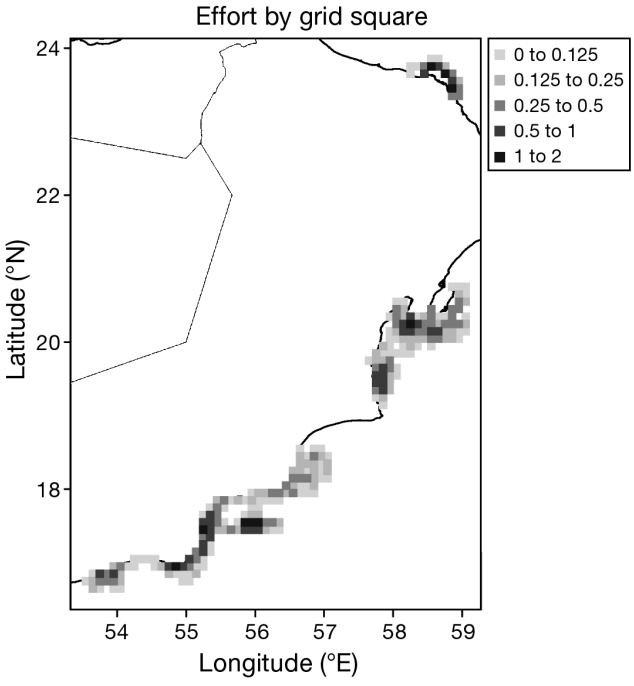


Fig. 2. On-effort survey (in decimal degrees) for cetaceans surveyed off Oman coastal waters, 2000 to 2003. Grid is that used for all analyses

tion to this was the area around Muscat, as authors who ran the field surveys lived there. Within each of the 3 survey areas, tracks were designed to provide as much coverage of the area as possible within the logistic and safety limitations of daily excursion small boat surveys.

Survey tracklines generally followed an irregular saw-tooth pattern along the coast, and were traversed at speeds of 12 to 15 knots (22 to 28 km h–1). Search effort was suspended when cetaceans were sighted and groups were approached to confirm species identity and collect data (e.g. photographs to identify individual animals, biopsy samples). Searching stopped in Beaufort states of 4 or higher. Sighting positions and other positional data were recorded using Garmin 12 or 12XL GPS units. Tracks were logged, with the vessel's position recorded every 30 to 45 s. Georeferenced data were imported into ArcView® 3.2a (ESRI: www.esri.com) and checked at the end of each day. Sightings data were stored in an MS Access® database.

Geoprocessing

Sightings data were overlaid onto a 0.1 × 0.1° lat./long. grid (at these latitudes, approximately 11 × 11 km). Grid cell size was determined as a compromise between accuracy in classifying habitat characteristics within grid cells and the need for sufficient encounters within each cell to yield usable results (e.g. Hamazaki 2002). On-effort portions of survey tracks were imported into ArcGIS (ESRI, WGS84 projection) and converted into shape files, one for each day's effort. The geo-processing 'intersect' and 'dissolve' functions of ArcGIS were then used to calculate the total distance (in decimal degrees) surveyed on-effort in each cell. The 'spatial join' function of ArcGIS was used to calculate the total number of cetacean groups, by species, in each cell, from the MS Access® database.

Digitized depth files were generated for each survey area using rasterized nautical charts (British Admiralty Raster Chart Series, British Admiralty chart nos. 2851, 2828, 2896, 3519, 3522, 3784, and 3785). Depth files were interpolated using ArcGIS Spatial Analyst to generate depth rasters for the grid, with a mask applied to exclude terrestrial surfaces from grid cells overlapping the coast. Minimum and maximum values for slope and depth were calculated from the rasters for each grid cell. All geoprocessing was conducted by G. Minton.

The ArcGIS shape file of the Oman coast (WGS84 projection) was imported into R (R Development Core Team 2010) using the maptools v0.7-34 (package Lewin-Koh et al. 2010), and converted into a SpatialLines object. The center point of each grid square was also read into R as a SpatialPoints object, using the same projection. Both objects were then projected to UTM (zone 40Q). In order to calculate the distance to shore for the center of each grid square, the SpatialPoints and SpatialLines objects were transformed into spatial point patterns and line segment patterns, respectively, using the spatstat v1.21-2 (package Baddeley & Turner 2005). The 'nncross' command was used to calculate distances.

Model construction

Our data were counts, and although the humpback data were approximately Poisson distributed, the Bryde's whale data were not. As we needed a modeling approach that would be consistent across both species, we used quasi-Poisson generalized linear models (GLMs) with log-link (Venables & Ripley 2002). To account for survey effort differing across grid cells, the natural log of on-effort distance for each cell was included as an offset. Mapping the

residuals of GLMs (see below) showed spatial patterning, so further analysis was undertaken. We used spatial eigenvector mapping (SEVM; Dormann et al. 2007) to account for residual spatial autocorrelation (SAC), as SEVM builds on GLM results. Also, as Dormann et al. (2007, p. 612) noted, it is a method that 'could thus be very useful for data with SAC stemming from larger scale observation bias,' and we know that biases due to haphazard design confound our data.

SEVM works by 'whitening out' residual spatial autocorrelation in a model, rather than incorporating it into the model. We used the 'ME' command from the spdep 0.5-16 (package Bivand et al. 2010), which takes a brute force approach to finding the smallest possible subset of eigenvectors that removes residual spatial autocorrelation from a GLM. The residual autocorrelation is then accounted for by refitting the original model with the eigenvectors included as covariates (for further details, see Dormann et al. 2007, Bivand et al. 2008). Hereafter, we refer to these as SEVM-GLMs. A flow diagram outlining the process of model construction and listing commands used is provided in Fig. 3.

Analyses were run using R 2.11.1 (R Development Core Team 2010) through rgedit 0.7.0.1 on an x86 computer running Ubuntu 9.04.

RESULTS

Of the slope and depth values, maximum depth and minimum slope were the least correlated and so were selected for inclusion in the model. Distance from shore was not strongly correlated with any other physiographic variable. Of the 3 separate study areas, the Muscat and Dhofar coasts are relatively similar in topography, with deep water within 1 grid cell of shore. The Gulf of Masirah, on the other hand, includes one of the largest areas of shallow waters anywhere off the Arabian Sea coast of Oman (approximately 80 km at the widest, see Fig. 1B), with a gently sloping shelf extending to the outer edge of the survey area. Coefficients for the GLMs for both species are shown in Table 2.

To calculate spatial eigenvectors, first we constructed a neighborhood, then calculated eigenvectors (see Fig. 3 for details). The distribution of Bryde's whales showed little spatial autocorrelation, so we used an alpha value of 0.25 as a stopping rule for eigenvector calculation. All neighborhoods are, by definition, confined within one of the 3 separate study areas, i.e. the Dhofar coast, the Gulf of Masirah, or off

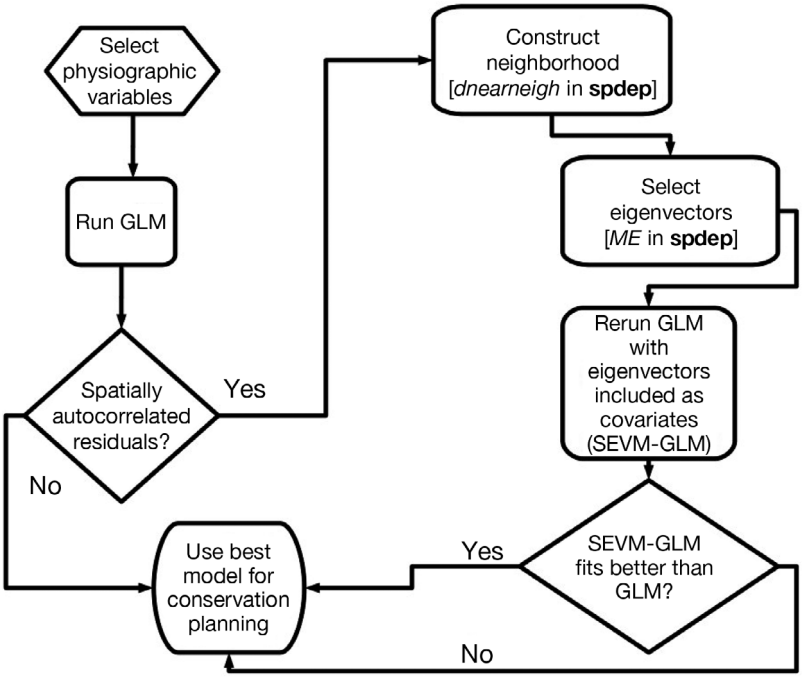


Fig. 3. Process by which the choice to use a generalized linear model (GLM) or a spatial eigenvector mapping GLM (SEVM-GLM) is made. Text in *italics* indicates commands used from the R packages indicated in **bold**

Table 2. Coefficients from the quasi-Poisson generalized linear models of humpback whales *Megaptera novaeangliae* and Bryde's whales *Balaenoptera* sp. surveyed off Oman coastal waters, 2000 to 2003, with results of significance testing. *p < 0.05, ***p < 0.001

	Estimate	SE	*t*	p r(>|*t*|)
Megaptera				
Intercept	-2.42×10^{-1}	2.04×10^{-1}	−1.182	0.238
DepthMax	5.83×10^{-4}	2.62×10^{-4}	2.228	0.0268*
DistShore	-1.46×10^{-5}	1.69×10^{-5}	−0.862	0.390
SlopeMin	-3.74×10^{-6}	1.95×10^{-6}	−1.915	0.0566
Balaenoptera				
Intercept	−2.59	4.13×10^{-1}	−6.256	<0.001***
DepthMax	-2.55×10^{-4}	6.21×10^{-4}	−0.410	0.682
DistShore	4.04×10^{-5}	2.19×10^{-5}	1.842	0.0667
SlopeMin	1.93×10^{-6}	1.36×10^{-6}	1.418	0.157

Muscat. Coefficients for the SEVM-GLMs for both species are shown in Table 3. Tests comparing the fits of GLMs and SEVM-GLMs are given in Table 4.

We plotted the results of GLMs and the SEVM-GLMs side by side in order to display their differences. Plots of the predicted values generated by GLMs and SEVM-GLMs are shown in Fig. 4, with residuals in Fig. 5. Note that predicted values are for counts of groups of cetaceans (the unit of sighting) per grid square over the entire period of the study, so the maps show spatial patterns of relative abundance. For both species, the standard errors estimated for the SEVM-GLMs were greater than those for the GLMs (as expected), and so these results are not mapped. Fig. 6 shows maps of eigenvector values.

For humpback whales, the most important habitat variable identified in the GLM was depth. The 3 eigenvectors extracted fell clearly into the 3 survey areas: off Muscat, the Gulf of Masirah, and the Dhofar coast, respectively (Fig. 6A). The SEVM-GLMs fitted the data much better than the GLM, with slope and depth appearing important. As expected, the SEVM-GLM residuals are substantially smaller and less spatially clustered than the GLM residuals (Fig. 5A).

Bryde's whales produced a very different result from humpbacks. The distribution of sightings across the study area was more even than Poisson (dispersion parameter for the GLM was ~0.6). There was relatively little spatial autocorrelation in the initial GLM, and the SEVM extracted only 1 eigenvector (Fig. 6B), in which autocorrelation in the Muscat study area predominated. Habitat variables identified as being important in the SEVM-GLMs are distance from shore and slope. Both models' predictions give relatively similar patterns, with the greatest relative abundance of Bryde's whales off Muscat and, to a lesser extent, off the Dhofar coast (Fig. 4B). Mapping residuals (Fig. 5B) suggested that the SEVM-GLM predicts the relative abundance of Bryde's whales somewhat better than does the GLM, although there is a small area in the Gulf of Masirah where neither model predicted their relative abundance well.

Table 3. Coefficients from spatial eigenvector mapping of quasi-Poisson generalized linear models of cetaceans surveyed off Oman coastal waters, 2000 to 2003. For species information see Table 2. *p < 0.05, **p < 0.01, ***p < 0.001

	Estimate	SE	t	p r(>|t|)
Megaptera				
Intercept	-7.10×10^{-1}	2.49×10^{-1}	−2.853	0.005**
DepthMax	4.11×10^{-4}	1.91×10^{-4}	2.150	0.033*
DistShore	-9.02×10^{-6}	1.48×10^{-5}	−0.611	0.542
SlopeMin	-3.29×10^{-6}	1.24×10^{-6}	−2.644	0.009**
fitted(meg.ME.quasi)vec1	−16.6	6.06	−2.747	0.007**
fitted(meg.ME.quasi)vec5	8.84	1.58	5.605	<0.001***
fitted(meg.ME.quasi)vec4	8.67	2.06	4.205	<0.001***
Balaenoptera				
Intercept	−3.17	5.50×10^{-1}	−5.767	<0.001***
DepthMax	-9.69×10^{-4}	8.83×10^{-4}	−1.097	0.274
DistShore	6.03×10^{-5}	2.42×10^{-5}	2.489	0.015*
SlopeMin	3.62×10^{-6}	1.71×10^{-6}	2.117	0.035*
fitted(bal.ME.quasi)	−8.44	2.55	−3.310	0.001**

Table 4. Tests comparing fits of generalized linear models (GLMs) and spatial eigenvector mapping of quasi-Poisson GLMs of cetaceans surveyed off Oman coastal waters, 2000 to 2003. For species information see Table 2. Dev.: deviance; ***p < 0.001

	Residual df	Residual dev.	Δdf	ΔDeviance	p (>|χ|)
Megaptera					
GLM	242	163.450			
SEVM-GLM	239	100.810	3	62.637	<0.001***
Balaenoptera					
GLM	242	48.284			
SEVM-GLM	241	40.535	1	7.749	<0.001***

DISCUSSION

Context

The mark–recapture estimate of abundance for the Arabian Sea population of humpback whales (82; 95% CI: 60–111; Minton et al. in press) is from data that are now almost a decade old. Relying on a time series of mark–recapture estimates of a cetacean population of around 100 animals to determine trends in abundance is futile (e.g. Thompson et al. 2000, Parra et al. 2006b). Analysis of scarring on the caudal peduncle region of photographically identified humpback whales in Oman in 2003 indicated that 30 to 40% of all whales examined were likely to have been involved in entanglements with fishing gear (Minton et al. in press). Despite this apparently high level of interaction with fisheries, there are no estimates of fisheries-related mor-

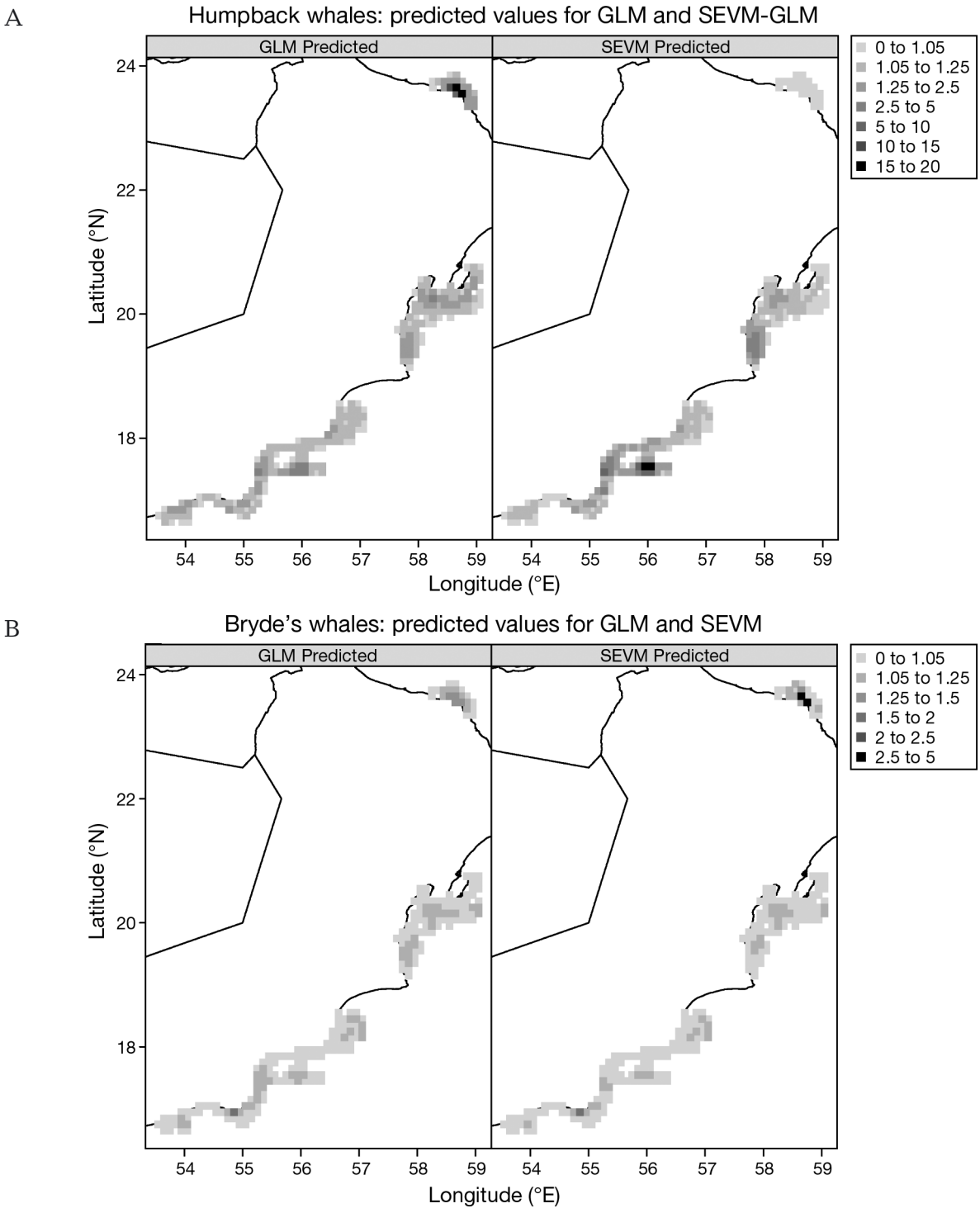


Fig. 4. *Megaptera novaeangliae* and *Balaenoptera* sp. Predicted numbers of whale groups for each grid square from the quasi-Poisson generalized linear models (GLM Predicted), and spatial eigenvector mapping of quasi-Poisson generalized linear models (SEVM Predicted) of cetaceans surveyed off Oman coastal waters, 2000 to 2003. (A) Humpback whales, (B) Bryde's whales

tality for this population. It is thus unrealistic to expect that the now-traditional approach outlined in the Introduction (estimate abundance, estimate anthropogenic mortality and model sustainability of anthropogenic mortality), can be implemented in a timely manner for this population.

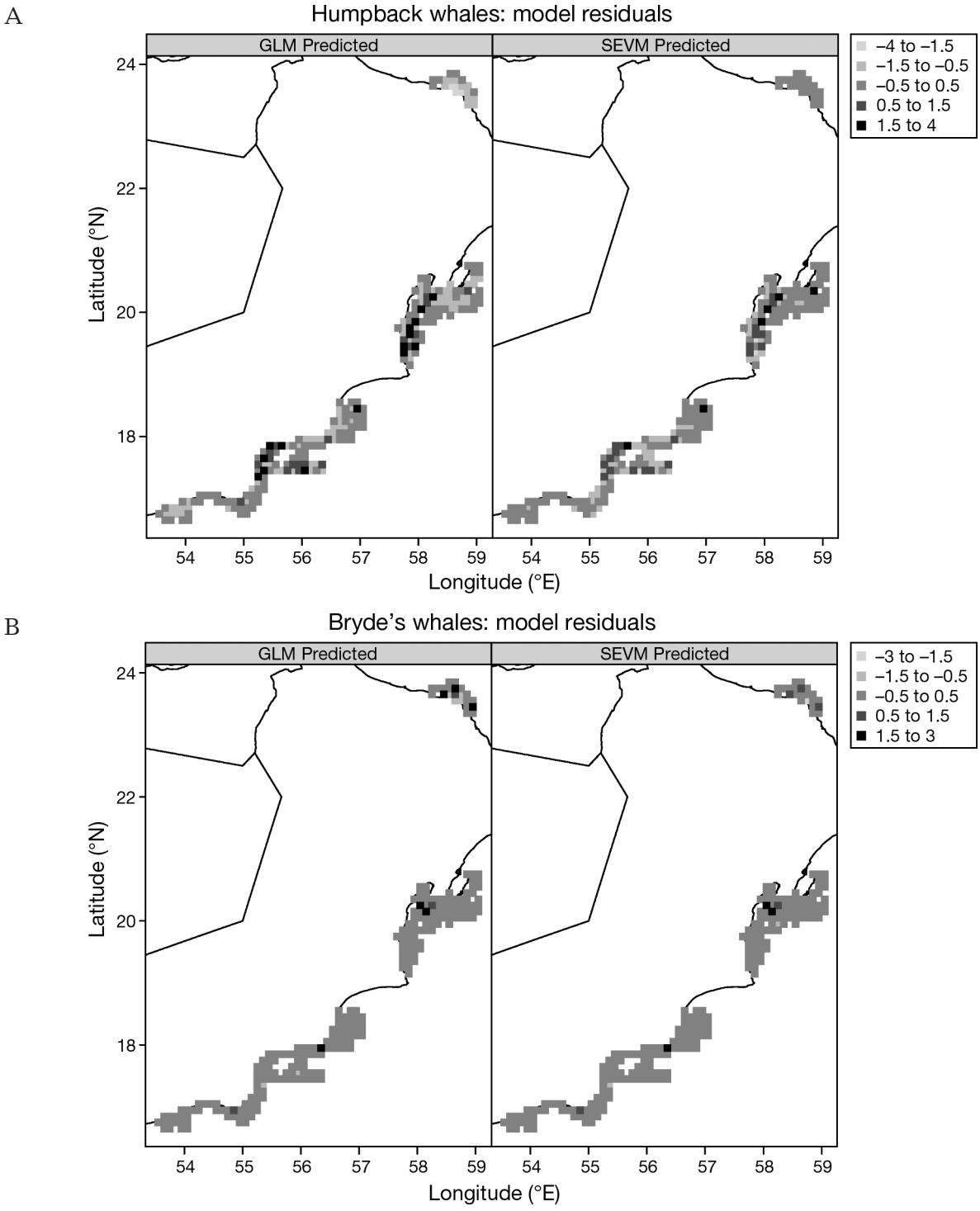
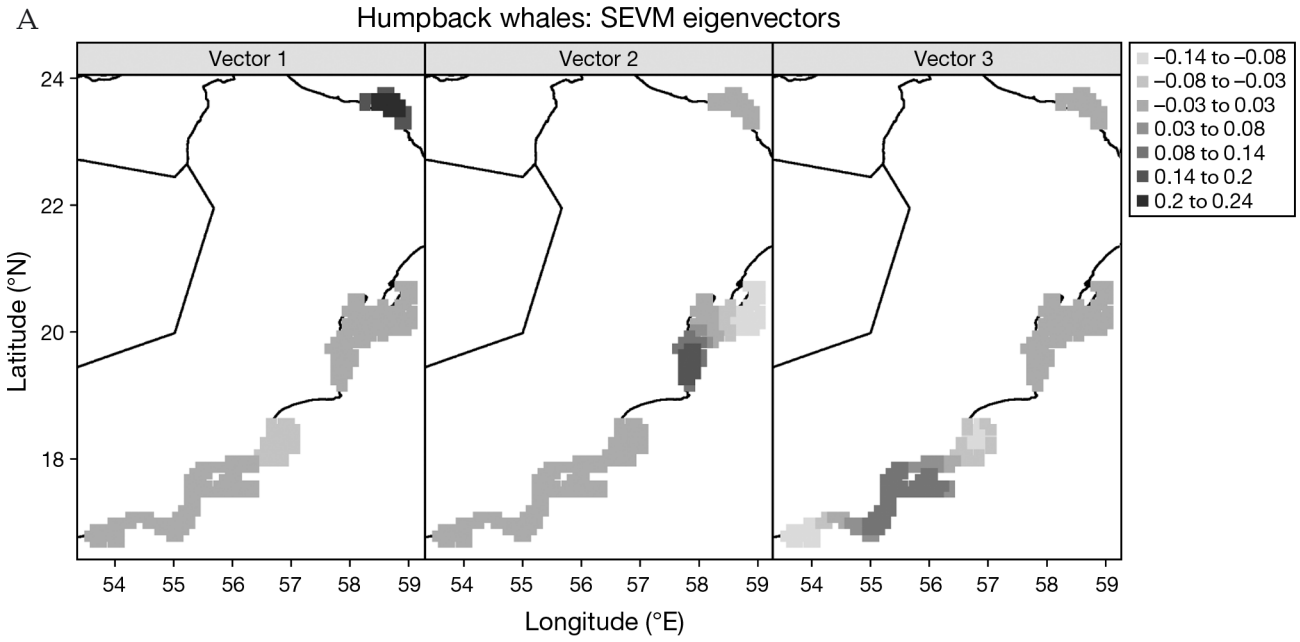


Fig. 5. *Megaptera novaeangliae* and *Balaenoptera* sp. Model residuals for each grid square from the quasi-Poisson generalized linear models (GLM Predicted) and spatial eigenvector mapping of quasi-Poisson GLMs (SEVM Predicted) of cetaceans surveyed off Oman coastal waters, 2000 to 2003. (A) Humpback whales, (B) Bryde's whales

How then can scientific input inform plans to manage anthropogenic activities impacting these whales? Further, are there any results from our study that can inform management more generally? In our experience, these related problems, i.e. small population, no reliable quantification of anthropogenic



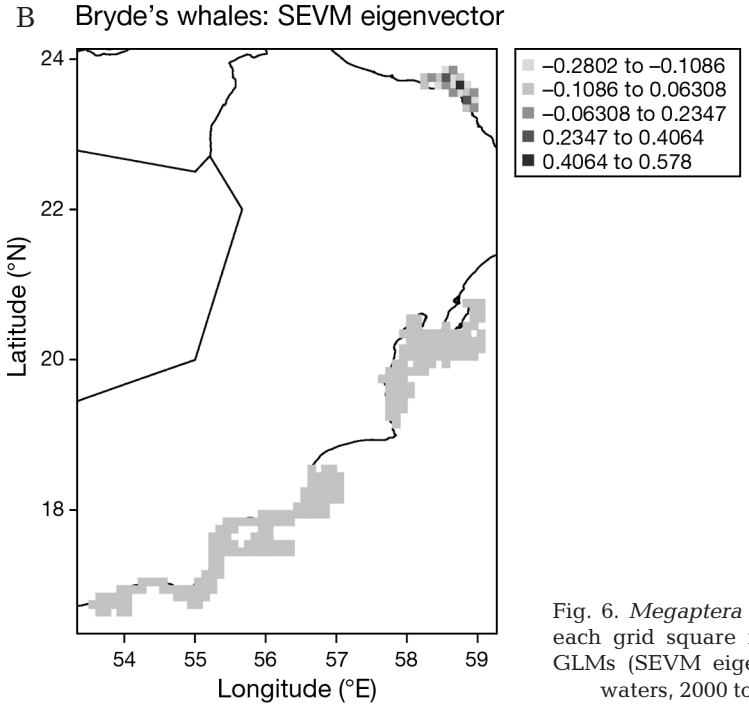


Fig. 6. *Megaptera novaeangliae* and *Balaenoptera* sp. Eigenvalues for each grid square from spatial eigenvector mapping of quasi-Poisson GLMs (SEVM eigenvectors) of cetaceans surveyed off Oman coastal waters, 2000 to 2003. (A) Humpback whales, (B) Bryde's whales

mortality or population trends, and limited resources, are not uncommon in most of the developing world.

Further, throughout the world, there is likely to be a large body of data collected using haphazard sampling methods, e.g. cetacean sighting data with associated effort (note that effort data are essential), but collected on platforms of opportunity, or with additional survey aims that grossly violate the assumptions of line- or strip-transect sampling. Haphazard sampling for photo-identification and genetic sampling is common, as it is for coastal patrols in marine protected areas. How can cetacean biologists make best use of such data to inform conservation planning? Here we show why SEVMs may be the most appropriate modern tool for analyzing this type of spatially autocorrelated data.

Other modeling approaches

Before discussing the results from our models, we outline why we did not use other approaches to habitat modeling.

Mixed models

We initially attempted to run generalized linear mixed models (GLMMs) with spatial structure to the random effect, as outlined by Dormann et al. (2007). First, we note that a caveat with this technique is that those authors refer to using a 'random' effect with only 1 category as a 'cheat' (Dormann et al. 2007, their supplementary material). We ran quasi-Poisson (and Poisson) GLMMs, but they produced non-positive definite approximate variance–covariance matrices, making it impossible to check the confidence intervals of the 'random' effect. We therefore did not pursue this line of analysis.

Additive models

Despite the popularity of generalized additive models (GAMs) in modeling cetacean habitat use (e.g. Gómez de Segura et al. 2007, Cañadas & Hammond 2008, Redfern et al. 2008), we did not use them, for 3 reasons. First, with relatively few samples for whale species (56 sightings over 4 yr for humpback whales, 15 for Bryde's whales, Table 1), techniques such as GAMs that can handle spatial autocorrelation by seeking nonlinear patterns in the raw data themselves become inherently less useful. Second, we wanted to compare model outputs between species in order to differentiate between autocorrelation due to haphazard survey design, and autocorrelation due to cetacean ecology. As GAMs, by definition, fit nonlinear curves to the data, we judged that they were less likely to allow us to make these distinctions. Finally, the SEVM-GLM approach allows a clear comparison to be made between a model that is unlikely to successfully account for spatial autocorrelation (the GLM) with one that does (the SEVM-GLM). Differences between these model predictions allow us to make some inference on the manner in which spatial autocorrelation influences these predictions, and so provides another form of insight into the driver(s) of the autocorrelation. A GAM-based approach would not allow this.

General niche-environment system factor analysis

Ecological niche factor analysis (ENFA), a form of niche-environment factor analysis (Calenge & Basille 2008), has recently become a popular tool for developing habitat models for cetaceans (e.g. Oviedo & Solís 2008, Praca et al. 2009). We did not use ENFA on our data for 2 principal reasons. First, ENFA is, by definition, based on Hutchinsonian niche hyperspace. We did not attempt to model whale niches, as we knew that we did not have environmental data available that would be appropriate for niche modeling (see Introduction). Secondly, the mathematical formulation for ENFA (Hirzel et al. 2002) does not explicitly account for spatial autocorrelation, especially that due to haphazard sampling in a 'design I' habitat use study (as defined by Thomas & Taylor 1990), such as this one. As the main point of our modeling exercise was to account for this form of spatial autocorrelation, ENFA was inappropriate.

Details from Oman

Our model results confirm and provide a statistically robust underpinning for previous work based on the same survey data (Minton et al. 2011). Our results clearly demonstrate the importance of the Dhofar coast, particularly in the region of the Hallaniyat Islands and Hasik, for the Arabian Sea population of humpback whales over our study period. Examination of the differences in outputs between models (i.e. with and without autocorrelation accounted for), and between species, allow us to identify the most reliable model outputs and thus the information that is of greatest value for management.

For humpback whales, there are substantial differences between the model predictions of the GLM and the SEVM-GLM (Fig. 4A). The GLM predicts the greatest relative abundance of humpbacks to be off Muscat, while the SEVM-GLM predicts most humpbacks off the Dhofar coast. Examination of model residuals (Fig. 5A) demonstrates that the SEVM-GLM prediction is more robust. The GLM prediction appears driven by the substantial search effort off Muscat, and the similarity in habitat characteristics between Muscat and the Dhofar coast (where most humpback sightings were made). Note that model bias is introduced into the GLM by focusing survey effort in an area known to be important for humpbacks that had similar physiographic characteristics (i.e. the Dhofar coast). The SEVM-GLM successfully handles this bias. Both models poorly predicted humpback occurrence in the shallow coastal shelf waters of the Gulf of Masirah (Fig. 4A), although the SEVM-GLM fit is somewhat better.

The way in which model outputs for Bryde's whales differ from humpback whales is of interest. As the effort data are from the same survey series, one would expect the confounding effects of haphazard sampling to be consistent, and as such, model differences would reflect biological/ecological traits of the species rather than sampling artefacts. The maximum cell count for Bryde's whales is approximately an order of magnitude less than that for humpbacks. The pattern of spatial autocorrelation in the data is also different. The distribution of Bryde's whales across the study area is more regular, and they are more prevalent off Muscat than are humpbacks. This suggests that the clumped distribution of humpbacks is real, as is the importance of the Dhofar coast for humpbacks. This has important local, and regional, implications for management.

The Arabian Sea population of humpback whales is the smallest population of humpback whales known to exist, the only population known not to undertake an extensive seasonal migration, and one of the most endangered baleen whale populations (Minton et al. 2008). The Dhofar coast, in particular in the region of the Hallaniyat Islands and Hasik, was identified previously (Minton et al. in press) as likely to be an important habitat for this population. Our modeling work quantifies the significance of this area for these whales.

Caveats

We considered it inappropriate to attempt to make inference on parts of the Oman coast not covered by the surveys. Although it is theoretically possible for us to project model predictions into other areas, we consider this inadvisable, as our basic design was not to make inference about the distribution of humpback whales along the entire Oman coast. Given the constraints under which field work was undertaken, both logistic and financial, a synoptic survey of the entire coast was impossible. We focused our survey effort on areas which available information suggested were likely the most important areas for humpback whales.

An extension of these caveats is that although we make recommendations on the relative importance for conservation of the Dhofar coast (see below), we cannot state with certainty that other areas will not prove equally important. Nevertheless, the information available is sufficient to note the importance of starting the process of mitigating inadvertent anthropogenic mortality on Arabian Sea humpback whales, and that the science available suggests that the best place to start is off the Dhofar coast. Recent (March 2011) fieldwork off the Dhofar coast, focusing near the village of Hasik (Fig. 1) was planned based in part on the results of our model. This field season resulted in regular, multiple sightings of humpback whales, and observations of feeding and breeding behavior, confirming the area's continued relative importance (A. Willson pers. obs.). Unfortunately, the threat of pirate activity offshore prevented field work around the Hallaniyat Islands.

Implications for humpback whale conservation

The coastal zone of Oman is experiencing rapid transformation as the country moves beyond a wholly petroleum-dependent economy. Oman's population growth rate is among the highest in the world (3.14% per annum), and there is a continuing demographic shift towards coastal areas (Oman Ministry of National Economy 2009). Fishing effort off the coast of Oman and in other parts of the Arabian Sea is increasing dramatically (Oman Ministry of Agriculture and Fisheries 2002, FAO 2007, Oman Ministry of National Economy 2009), and drifting and set gillnets as well as traps are already widely used (Stengel & Al Harthy 2002).

One of the most important findings of this study is that the clustering of humpback whales along part of the Dhofar coast and the Hallaniyat Islands is not a sampling artefact, but a result of the whales' ranging behavior. This suggests that a spatially-explicit management program should be implemented along this section of the Dhofar coast, as a preliminary step to larger-scale marine conservation planning in Oman. There are instances where declaring a marine protected area for cetacean conservation has not led to cessation of threatening processes, particularly gillnetting (e.g. Notarbartolo di Sciara et al. 2008). There are also examples where it has been successful: the implementation of netting restrictions to protect dugongs, and the general process of rezoning in the Great Barrier Reef Marine Park (Dobbs et al. 2008, Grech et al. 2008) provides an example of how to achieve spatially-explicit restrictions on netting. We suggest that this process, suitably modified for Omani cultural norms and local capacity for management, start as soon as possible.

Finally, we suggest that other researchers working with spatially autocorrelated data on cetacean (and other marine wildlife) distribution give serious consideration to using spatial eigenvector models.

There will be instances where sampling regimes are just too haphazard, and there are no options to distinguish between the likely causes of spatial autocorrelation, where these models may prove ineffective. But in those instances, it is possible that any other spatial modeling approach will be equally ineffective. Researchers need to ensure that they do not make inference beyond their data with this tool, as is the case for all spatial or habitat models. Tools to run spatial eigenvector models on all operating systems are available for free download as part of the R statistical language (R Development Core Team 2010), and example code to run eigenvector filtering is provided as an appendix to the paper by Dormann et al. (2007).

Acknowledgements. Logistic support and research permits for surveys were provided by Oman's Ministry of Environment and Climate Affairs, the Oman Natural History Museum, and Oman's Ministry of Agriculture and Fisheries. Fieldwork was supported by: The Ford Environmental Grants, The UK Foreign and Commonwealth Office, Shell Marketing Oman, Petroleum Development Oman, Veritas Geophysical, The Peter Scott Trust for Education and Research in Conservation, and the Marina Bandar al Rowdah. The manuscript was improved by comments from Daniel Palacios and two anonymous reviewers.

LITERATURE CITED

Baddeley A, Turner R (2005) Spatstat: an R package for analyzing spatial point patterns. J Stat Softw 12:1–42

Basille M, Calenge C, Marboutin É, Andersen R, Galliard JM (2008) Assessing habitat selection using multivariate statistics: some refinements of the ecological-niche factor analysis. Ecol Model 211:233–240

Bivand R, Altman M, Anselin L, Assunção R and others (2010) spdep: Spatial dependence: weighting schemes, statistics and models. R package version 0.5-16. Available at http://CRAN.R-project.org/package=spdep

Bivand RS, Pebesma EJ, Gómez-Rubio V (2008) Applied spatial data analysis with R. Springer, New York, NY

Braulik GT, Ranjbar S, Owfi F, Aminrad T, Dakhteh SMH, Kamrani E, Mohsenizadeh F (2010) Marine mammal records from Iran. J Cetacean Res Manag 11:49–64

Burkill PH (1999) ARABESQUE: an overview. Deep-Sea Res II 46:529–547

Calenge C, Basille M (2008) A general framework for the statistical exploration of the ecological niche. J Theor Biol 252:674–685

Cañadas A, Hammond PS (2008) Abundance and habitat preferences of the short-beaked common dolphin *Delphinus delphis* in the southwestern Mediterranean: implications for conservation. Endang Species Res 4: 309–331

Crowder LB, Osherenko G, Young OR, Airame S and others (2006) Resolving mismatches in U.S. ocean governance. Science 313:617–618

Dobbs K, Fernandes L, Slegers S, Jago B and others (2008) Incorporating dugong habitats into the marine protected area design for the Great Barrier Reef Marine Park, Queensland, Australia. Ocean Coast Manag 51:368–375

Dormann CF (2009) Response to Comment on 'Methods to account for spatial autocorrelation in the analysis of species distributional data: a review'. Ecography 32:379–381

Dormann CF, McPherson JM, Araújo MB, Bivand R and others (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30:609–628

FAO (Food and Agriculture Organisation of the United Nations) (2007) The state of world fisheries and aquaculture 2006. FAO, Rome

Fernandes L, Day J, Lewis A, Slegers S and others (2005) Establishing representative no-take areas in the Great Barrier Reef: large scale implementation of theory on marine protected areas. Conserv Biol 19:1733–1744

Gómez de Segura A, Hammond PS, Cañadas A, Raga JA (2007) Comparing cetacean abundance estimates derived from spatial models and design-based line transect methods. Mar Ecol Prog Ser 329:289–299

Grech A, Marsh H (2008) Rapid assessment of risks to a mobile marine mammal in an ecosystem-scale marine protected area. Conserv Biol 22:711–720

Grech A, Marsh H, Coles R (2008) A spatial assessment of the risk to a mobile marine mammal from bycatch. Aquatic Conserv 18:1127–1139

Hamazaki T (2002) Spatiotemporal prediction models of cetacean habitats in the mid-western North Atlantic Ocean (from Cape Hatteras, North Carolina, U.S.A. to Nova Scotia, Canada). Mar Mamm Sci 18:920–939

Hirzel AH, Hausser J, Chessel D, Perrin N (2002) Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? Ecology 83:2027–2036

Hutchinson GE (1957) Concluding remarks. Cold Spring Harbor Symp Quant Biol 22:415–427

Johnston DW, Westgate AJ, Read AJ (2005) Effects of fine-scale oceanographic features on the distribution and movements of harbour porpoises *Phocoena phocoena* in the Bay of Fundy. Mar Ecol Prog Ser 295:279–293

Lewin-Koh NJ, Bivand R, Pebesma EJ, Archer A and others (2010) maptools: Tools for reading and handling spatial objects. R package version 0.7-34. Available at http://CRAN.R-project.org/package=maptools

Macleod K, Fairbairns R, Gill A, Fairbairns B, Gordon J, Blair-Myers C, Parsons ECM (2004) Seasonal distribution of minke whales *Balaenoptera acutorostrata* in relation to physiography and prey off the Isle of Mull, Scotland. Mar Ecol Prog Ser 277:263–274

Margules CR, Pressey RL (2000) Systematic conservation planning. Nature 405:243–253

Mikhalev YA (2000) Whaling in the Arabian Sea by the whaling fleets Slava and Sovetskaya Ukraina. In: Yablokov AV, Zemsky VA, Tormosov DD (eds) Soviet whaling data (1949–1979). Centre for Russian Environmental Policy, Moscow, p 141–181

Minton G, Collins TJQ, Pomilla C, Findlay KP, Rosenbaum HC, Baldwin R, Brownell RL Jr (2008) *Megaptera novaeangliae*, Arabian Sea subpopulation. IUCN Red List of Threatened Species. Available at www.iucnredlist.org/details/132835

Minton G, Cerchio S, Collins T, Ersts P and others (2010) A note on the comparison of humpback whale tail fluke catalogues from the Sultanate of Oman with Madagascar and the East African Mainland. J Cetacean Res Manag 11:65–68

Minton G, Collins TJQ, Findlay KP, Baldwin R (2011) Cetacean distribution in the coastal waters of the Sultanate of Oman. J Cetacean Res Manag 11:301–313

Minton G, Collins T, Findlay K, Ersts P, Rosenbaum H, Berggren P, Baldwin R (in press) Seasonal distribution, abundance, habitat use and population identity of humpback whales in Oman. J Cetacean Res Manag (Spec Issue Southern Hemisphere Humpback Whales)

Notarbartolo di Sciara G, Agardy T, Hyrenback D, Scovazzi T, Van Klaveren P (2008) The Pelagos Sanctuary for Mediterranean marine mammals. Aquatic Conserv 18: 367–391

Oman Ministry of Agriculture and Fisheries (2002) Fisheries statistical year book 2001. Ministry of Agriculture and Fisheries, Muscat

Oman Ministry of National Economy (2009) 2009 statistical yearbook. Ministry of National Economy, Muscat

Oviedo L, Solís M (2008) Underwater topography determines critical breeding habitat for humpback whales near Osa Peninsula, Costa Rica: implications for marine protected areas. Rev Biol Trop 56:591–602

Parra GJ, Schick R, Corkeron PJ (2006a) Spatial distribution and environmental correlates of Australian snubfin and Indo-Pacific humpback dolphins. Ecography 29:396–406

Parra GJ, Corkeron PJ, Marsh H (2006b) Population sizes, site fidelity and residence patterns of Australian snubfin and Indo-Pacific humpback dolphins: implications for conservation. Biol Conserv 129:167–180

Praca E, Gannier A, Das K, Laran S (2009) Modelling the habitat suitability of cetaceans: example of the sperm whale in the northwestern Mediterranean Sea. Deep-Sea Res I 56:648–657

R Development Core Team (2010) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available at www.R-project.org

Redfern JV, Barlow J, Ballance LT, Gerrodette T, Becker EA (2008) Absence of scale dependence in dolphin–habitat models for the eastern tropical Pacific Ocean. Mar Ecol Prog Ser 363:1–14

Rosenbaum HC, Pomilla C, Mendez M, Leslie MS and others (2009) Population structure of humpback whales from their breeding grounds in the South Atlantic and Indian Oceans. PLoS ONE 4:e7318

Stafford KM, Citta JJ, Moore SE, Daher MA, George JE (2009) Environmental correlates of blue and fin whale call detections in the North Pacific Ocean from 1997 to 2002. Mar Ecol Prog Ser 395:37–53

Stengel H, Al Harthy A (2002) The traditional fishery of the Sultanate of Oman (fishing gear and methods). Ministry of Agriculture and Fisheries, Directorate General of Fisheries Resources, Marine Science and Fisheries Center, Muscat

Taylor BL, Wade PR, DeMaster DP, Barlow J (2000) Incorporating uncertainty into management models for marine mammals. Conserv Biol 14:1243–1252

Thomas D, Taylor E (1990) Study designs and tests for comparing resource use and availability. J Wildl Manag 54: 322–330

Thompson PM, Wilson B, Grellier K, Hammond PS (2000) Combining power analysis and population viability analysis to compare traditional and precautionary approaches to conservation of coastal cetaceans. Conserv Biol 14:1253–1263

Venables WN, Ripley BD (2002) Modern applied statistics with S, 4th edn. Springer, New York, NY

Wade PR (1998) Calculating limits to the allowable human-caused mortality of pinnipeds and cetaceans. Mar Mamm Sci 14:1–37

Wray P, Martin KR (1983) Historical whaling records from the western Indian Ocean. Rep Int Whal Comm 5 (Spec Issue):213–241

Editorial responsibility: Daniel Palacios, Pacific Grove, California, USA

Submitted: January 16, 2011; Accepted: July 8, 2011
Proofs received from author(s): October 6, 2011