



Modelling sperm whale habitat preference: a novel approach combining transect and follow data

Enrico Pirotta^{1,*}, Jason Matthiopoulos^{1,2}, Monique MacKenzie²,
Lindsey Scott-Hayward², Luke Rendell¹

¹Sea Mammal Research Unit, Scottish Oceans Institute, School of Biology, University of St. Andrews, St. Andrews, Fife KY16 8LB, UK

²Centre for Research into Ecological and Environmental Modelling, University of St. Andrews, St. Andrews, Fife KY16 9LZ, UK

ABSTRACT: Sperm whale *Physeter macrocephalus* habitat preferences are still poorly understood in the Mediterranean, despite the population being classified as 'Endangered' by the IUCN. Techniques to make the best use of multiple data sources are important in improving this situation. This work provides a detailed evaluation of sperm whale distribution and habitat use around the Balearic Islands using a novel analytical framework that combines transect and follow data while accounting for any autocorrelation present. During dedicated research cruises (2003 to 2008), sperm whales were located by listening at regular intervals along a search track and subsequently followed acoustically. Sperm whales were encountered 56 times and followed for periods ranging from a few hours to 3 d. Logistic Generalized Additive Models were used to model the probability of whale presence across the study area as a function of environmental variables, and Generalized Estimating Equations were used to account for autocorrelation. The results suggest that sperm whales do not use the region uniformly and that topography plays a key role in shaping their distribution. Moreover, solitary individuals were found to use the habitat differently from groups. This segregation appeared to be driven by water temperature and might reflect different needs or intraspecific competition. By shedding light on sperm whale habitat preference in such a critical area, this study represents an important step towards the implementation of conservation measures for this population.

KEY WORDS: Habitat modelling · Cetacean · Distribution · Balearic Islands · GAM · GEE · Multi-scale

—Resale or republication not permitted without written consent of the publisher—

INTRODUCTION

A permanent population of sperm whales *Physeter macrocephalus* Linnaeus, 1758 inhabits the Mediterranean Sea, where they apparently constitute a genetically distinct stock from that of the neighbouring North Atlantic (Drouot et al. 2004, Engelhaupt 2004). Even though no overall abundance estimate is available for this population, the comparison of recent sperm whale survey data to historical records suggests that, locally, encounter rates are declining; this has led to its classification as 'Endangered' under the International Union for the Conservation of Nature (IUCN) criteria (Reeves & Notarbartolo di Sciara 2006, Notarbartolo di Sciara & Birkun 2010). The presence of the

sperm whale has been historically reported in the waters around the Balearic Islands in the western Mediterranean (e.g. Table 1 in Reese 2005), and recent surveys have shown relatively high encounter rates around the archipelago (Gannier et al. 2002).

Sperm whales across the world have a distinctive social system, segregating into long-term social units containing adult females and their immature offspring, and typically solitary maturing and mature males (Whitehead 2003). This picture seems to also broadly apply to the Mediterranean (Drouot et al. 2004a,b). Together with the waters off Crete (Frantzis et al. 2003), the Tyrrhenian Sea (e.g. Drouot et al. 2004a) and, recently, the Ligurian Sea (occurrences reviewed in Notarbartolo di Sciara & Birkun 2010), the Balearic

*Email: pirotta.enrico@gmail.com

archipelago is one of the few areas in the Mediterranean Sea in which both social units and mature males are observed consistently, suggesting a possibly important role as breeding ground (Gannier et al. 2002, Drouot-Dulau & Gannier 2007). Nevertheless, no long-term study has been conducted to date to specifically estimate the species distribution in this region, and thus little information exists on its space use and habitat preferences around the archipelago.

Habitat preference modelling techniques represent a useful tool to quantify the relationships between a species and its environment (Guisan & Zimmermann 2000, Redfern et al. 2006, Matthiopoulos & Aarts 2010). Knowing where the animals are, what environmental characteristics influence their choice of habitat and how this choice changes with time is crucial to understanding the species' ecology, identifying the areas of critical importance, assessing the overlap with human activities and, ultimately, guiding appropriate conservation efforts (Redfern et al. 2006). Food availability is probably the main determinant of space use by marine mammals (e.g. Benoit-Bird & Au 2003, Hastie et al. 2004, Frederiksen et al. 2006, Friedlaender et al. 2006). Other potentially important factors affecting habitat choice include behavioural state, the presence of calves, interspecific relationships (Cañadas & Hammond 2008), predation risk (e.g. Heithaus & Dill 2002), competition (e.g. Shane 1995) and reproductive needs (e.g. Ersts & Rosenbaum 2003). Prey abundance and other factors are often hard to measure directly (Guisan & Zimmermann 2000, Jaquet & Gendron 2002) and other easier to obtain environmental variables are thus used as proxies, even if they are not always directly and causally related with animal presence (Redfern et al. 2006). Several physiographic (e.g. depth, slope, aspect), oceanographic (e.g. sea surface temperature) and biological variables (e.g. chlorophyll *a* surface concentration) have been successfully employed to describe cetacean habitat preference indirectly (e.g. Cañadas et al. 2002, Davis et al. 2002, Hamazaki 2002, Yen et al. 2004, Cañadas et al. 2005, Ferguson et al. 2006, Panigada et al. 2008, Praca & Gannier 2008). There are however disadvantages of using such proxy measures. Typically, they only explain a small proportion of the observed variability in animal occurrence, and their use limits the ability to extrapolate to other areas because they replace unknown combinations of direct predictors; the same proxy measure might be caused by a different combination of direct predictors in different geographical contexts, resulting in a different relationship with animal occurrence (Guisan & Zimmermann 2000).

Sperm whale habitat use has been investigated both within the Mediterranean (Cañadas et al. 2002, Gan-

nier & Praca 2007, Azzellino et al. 2008, Praca et al. 2009) and worldwide (Waring et al. 2001, Davis et al. 2002, Jaquet & Gendron 2002, Rendell et al. 2004, Embling 2008, Skov et al. 2008; studies before 1996 have been reviewed by Jaquet 1996). Results have varied with some studies being able to draw direct relationships between sperm whale presence and primary productivity, sea surface temperature or some aspects of topography (such as slope or depth), and others finding relatively weak links at certain spatial and temporal scales.

As for many cetacean species, modelling sperm whale habitat use is complicated by the dynamic nature of marine ecosystems and by the whales' mobility, complex life history, and inaccessibility (they are often offshore and underwater; Redfern et al. 2006). In addition, other methodological issues become particularly challenging when assessing the habitat preference of this species. Firstly, distribution information often needs to be recorded in tandem with acoustic or photo-identification data, meaning that detected animals are followed for extended time periods of up to several days. Position data carry information relevant to habitat use so they should ideally be used in the analysis, but advanced statistical techniques are required to correct for their inherent autocorrelation (Scott-Hayward 2006, Embling 2008). Secondly, the relationship between sperm whale occurrence and environmental covariates is dependent on the spatio-temporal scale at which it is evaluated (Jaquet 1996). A substantial temporal or spatial lag (several months and hundreds of kilometres), corresponding to the time needed for primary productivity to work through the trophic web, might occur before top predator distribution responds to variation in a particular proxy measure (Jaquet 1996, Littaye et al. 2004, Croll et al. 2005). Such lags are likely to confound results, so a multi-scale approach is necessary (e.g. Jaquet 1996, Praca et al. 2009). Finally, social groups may have different habitat preferences compared to solitary individuals (e.g. mature males in the case of sperm whales) as a result of their different ecological and biological needs. This difference has been recorded in other terrestrial (Jakimchuk et al. 1987, Litvaitis 1990) and marine mammal species (Stewart & DeLong 1995, Ersts & Rosenbaum 2003), and its evaluation can be particularly interesting in those areas that are identified both as breeding and feeding grounds.

In this study we develop a novel analytical framework that can make use of combined acoustic and tracking data and account for autocorrelation in such data. We use a multi-scale approach to investigate the pattern of sperm whale distribution around the Balearic archipelago, and the potential differences in habitat selection by single individuals versus groups.

MATERIALS AND METHODS

Study area. The study area ran from 38 to 41° N and from 0.5 to 5° E, enclosing the continental shelf, continental slope and offshore waters around the Balearic archipelago (Fig. 1a). The Balearic promontory is characterised by distinct topographic features (Acosta et al. 2002). The narrow shelf to the north of Mallorca and Menorca islands breaks into a steep slope, incised by several gullies, while the southern part presents a more gentle depth gradient with 2 distinct canyon systems (Acosta et al. 2003). At the south-western limit of this shelf, a south-west to north-east linear scarp (the Emile Baudot Scarp) is characterised by a steep slope and numerous small canyon systems. Around the western islands, Ibiza and Formentera, the shelf is characterised by a variable gradient that peaks on the steep western side. In terms of water circulation, the region is an important transition area between the Balearic-Provençal basin and the Algerian basin (Pinot et al. 2002). The interaction between colder, more saline Mediterranean waters flowing from the north and warmer, fresher Atlantic waters creates the Balearic Front over the northern slope of the islands (Garcia et al. 1994). Regional dynamics show a marked interannual and seasonal variability (Pinot et al. 2002).

Data collection. Dedicated summer research cruises were operated for 6 consecutive seasons (Table 1). Motor-sailing yachts (11–12 m length) were used as research platforms. Boat tracks were recorded through the data logging software Logger 2000 by the International Fund for Animal Welfare (IFAW) that was connected to a global positioning system unit (Garmin GPS12). Transects were not systematically designed, but the resulting route extensively covered the shelf-break region around Mallorca and Menorca as well as some areas south of Ibiza (Fig. 1a). A hydrophone was deployed to detect sperm whales; in 2003, a single hydrophone (Sensor Technology of Canada; frequency response 0.1–22 kHz) was used, while from 2004 onward a dual-element hydrophone (Benthos AQ4; frequency response 0.1–22 kHz) towed at 100 m was used because it allowed monitoring the sea without stopping the boat. Regular acoustic monitoring every 30 min in waters deeper than 200 m checked for the presence of vocalising animals in the area; this condition was defined as 'On-effort'. Anytime the hydrophone was not deployed or no systematic listening was conducted, the research status was considered to be 'Off-effort'. Constant visual scanning of the sea was also carried out during daylight hours to support the acoustic search for sperm whales. Because sperm whales are highly vocal, both when foraging and socialising, they were mainly detected and followed acoustically. When whales were heard, they were

tracked and, during daylight hours, approached to make visual contact. An encounter was defined as the time spent in continuous acoustic contact with the whales, from first hearing them to a loss of acoustic contact exceeding 1 h. A distinction was made between encounters with single individuals, those that performed a stereotyped diving pattern and never associated with other individuals, and groups, i.e. more individuals (often including young animals) that were showing clear signs of direct interaction (e.g. social vocalisations or surface behaviours) or 'moving together in a coordinated fashion over periods of at least hours' (Whitehead 2003). The whales were left when all the desired data (photo-identification pictures, acoustic recordings and sloughed skin for genetic analysis) had been collected, unless contact with them was lost or other conditions necessitated the boat's departure (e.g. sea state, fuel shortage). After the end of an encounter, the searching effort generally continued along the previously planned route.

Unit of analysis. Points corresponding to GPS fixes recorded approximately every 20 min were used as the unit of analysis. 'Off-effort' fixes (i.e. those not with the whales and not searching for them) were excluded. Each point was classified as a 'presence' if the researchers were in acoustic contact with the whales or an 'absence' if no whales were heard. The points were then grouped into 2 types of blocks; a 'follow' was defined as the series of consecutive presence points that formed each acoustic encounter with the whales, and a 'searching transect' was defined as the series of consecutive absence points that constituted each portion of the boat track travelled 'On-effort', i.e. between 2 follows or 'Off-effort' intervals. These blocks were used in the analysis to account for the autocorrelation between the residuals within blocks, while independence was assumed between blocks (see 'Data analysis').

Environmental covariates. The variables used to predict sperm whale presence comprised depth, slope gradient, slope aspect, chlorophyll *a* (chl *a*) surface concentration, sea surface temperature (SST), sea surface height (SSH) deviation, and surface wind direction (Table 2).

Depth was expressed as a negative value in metres and taken from the ETOPO2v2 dataset (resolution 2 arc minutes; NOAA 2006). Slope gradient (hereafter 'slope') was defined as the maximum rate of change in depth in a given grid cell and expressed as percent slope. These data were computed from the ETOPO2v2 data using GIS software (Manifold 8.0) at 3 different spatial scales averaged over progressively larger grids at resolutions of 2 × 2 n miles (1×), 10 × 10 n miles (5×) and 20 × 20 n miles (10×). Slope aspect (hereafter 'aspect') was defined as the compass orientation of the

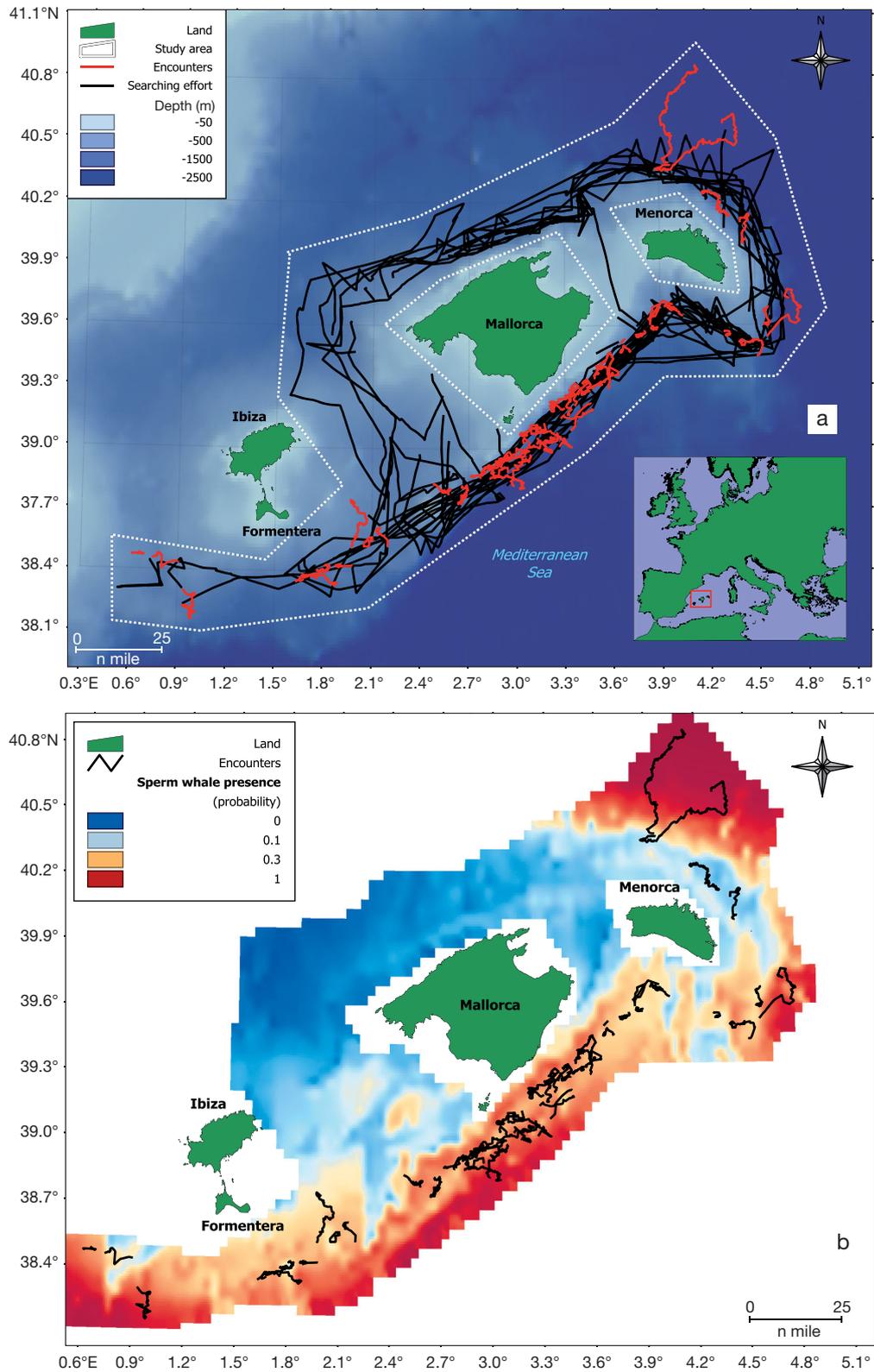


Fig. 1. (a) Study area in the Balearic Islands showing the distribution of effort and whale encounters. Prediction maps for (b) the entire data set; (c) the group subset; and (d) the singleton subset. Projection: UTM 31; Datum: WGS84; Coastline data source: Global Self-consistent Hierarchical High-resolution Shorelines, available at www.ngdc.noaa.gov/mgg/shorelines/gshhs.html; bathymetry source: 2-minute Gridded Global Relief Data ETOPO2v2 from NOAA

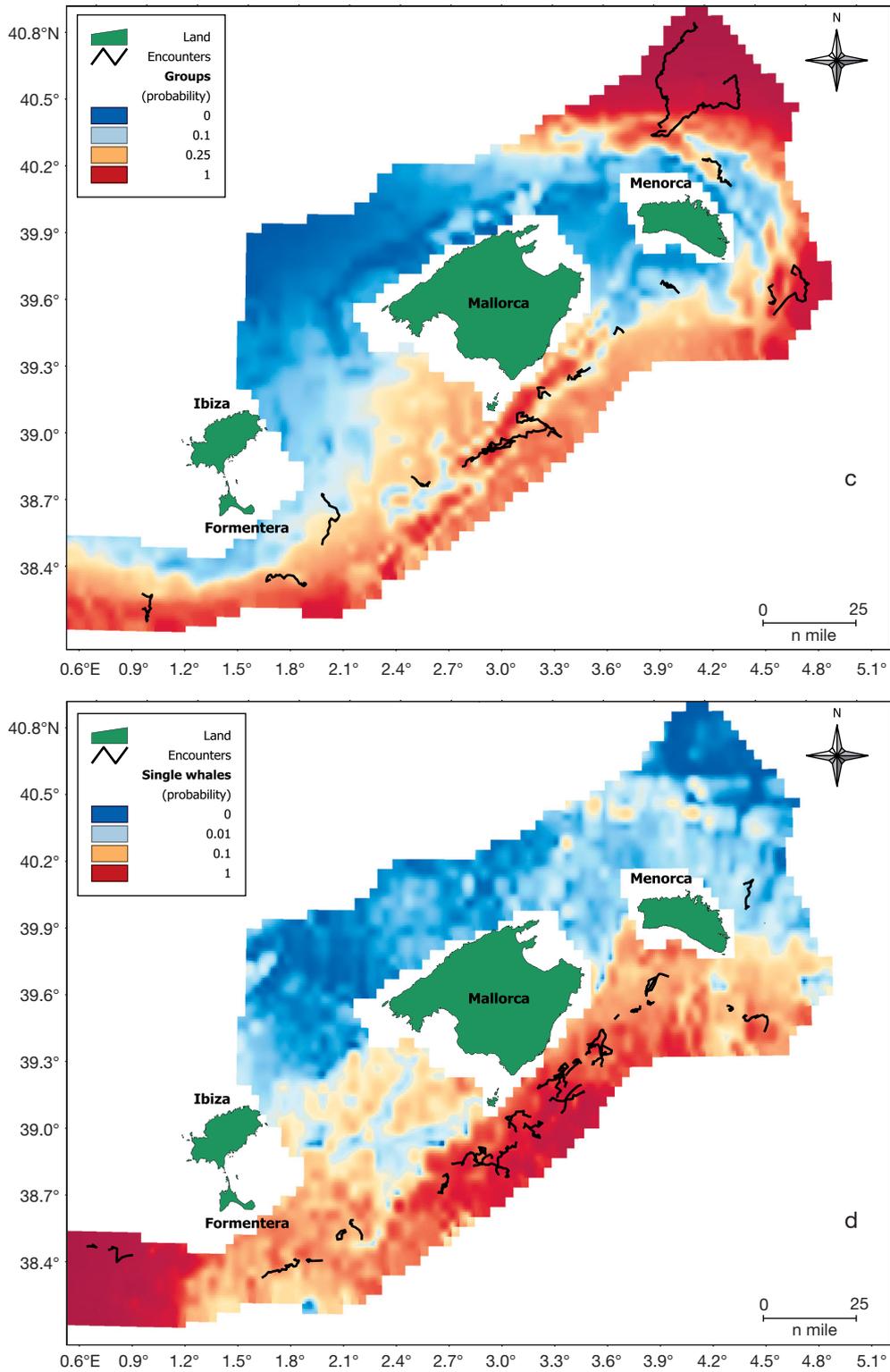


Fig. 1 (continued)

Table 1. Research effort and encounters with sperm whales

Year	Research period	Searching effort (km)	Encounters (km)	No. of encounters	
				Overall	Group
2003	1–28 Aug	2467	141	3	2
2004	10 Jul–6 Aug	2070	310	9	3
2005	9 Jul–5 Aug	1992	392	12	5
2006	14–28 Jul	1702	270	9	2
2007	3–28 Jul	1835	374	11	0
2008	15–27 Jul	1033	233	12	5

Table 2. Environmental covariates considered in the analysis and their scales. SST: sea surface temperature; SSH: sea surface height

Covariate	Unit	Spatial scale(s)	Temporal scale(s)
Depth	m	2 × 2 arcmin	–
Slope	%	2 × 2 arcmin	–
		10 × 10 arcmin	
		20 × 20 arcmin	
Aspect	°	2 × 2 arcmin	–
Chl <i>a</i>	mg m ⁻³	0.05 × 0.05°	1 mo lag (Jun); SD: 1/2, 1/5, 1/10
		0.5 × 0.5°	3 mo lag (Apr); SD: 1/2, 1/5, 1/10
			5 mo lag (Feb); SD: 1/2, 1/5, 1/10
SST	°C	0.05 × 0.05°	Weekly, monthly
SST slope	%	0.05 × 0.05°	Weekly
SST deviation (from median)	°C	0.05 × 0.05°	Monthly
SSH	m	0.25 × 0.25°	Monthly
Surface wind direction	°	0.25 × 0.25°	Monthly

slope, ranging from -180 to $+180^\circ$ with respect to true north, and also computed from the ETOPO2v2 data set on a scale of 2 n miles. Chlorophyll *a* surface concentration (hereafter 'chl *a*') is a proxy of primary productivity and phytoplankton biomass in the surface layer. We used Moderate Resolution Imaging Spectroradiometer (MODIS) data from NASA's Aqua satellite, pre-processed by the NASA's Goddard Space Flight Center (GSFC) using the SeaWiFS Data Analysis System (SeaDAS) software (Fu et al. 1998). The data were available as monthly concentrations (mg m⁻³) at a resolution of 0.05° longitude × 0.05° latitude from the NOAA CoastWatch program (<http://coastwatch.pfeg.noaa.gov/index.html>). Some of the downloaded data were incomplete with missing data generally in the form of stripes in the maps that were filled with the interpolation tools in the Manifold GIS system. Since time is necessary for changes in primary production to filter through the trophic levels to whales, the effect of this covariate was evaluated over different temporal scales. We did not know in advance how long this lag was, so we considered the concentrations of chl *a* around 3 different peak months (February, April and

June). The value of chl *a* for each period was calculated as a weighted average of the monthly chl *a* values around the peak, with the peak month having the largest weight and the months further away from the peak having less weight. To allow for uncertainty in the precise way primary production affects higher trophic levels, we tested 3 candidate weighting functions (all Gaussian kernels), centred on each of the 3 peak months and with standard deviations (SDs) corresponding to 1/2, 1/5 and 1/10 of the time interval between the peak month and the research month (i.e. July, except for 2003, when the survey was conducted in August). The different SDs determined how fast the influence of the monthly chl *a* values declined around each peak. As a result, 9 different chl *a* maps (3 peak maps, each with 3 SDs of the distribution of the weights) were obtained, with grid resolutions of 0.05° (\approx 3 n miles). Each temporal scale was then evaluated on 2 different spatial scales, corresponding to resolutions of 0.05 and 0.5° , the latter achieved by averaging chl *a* concentrations over larger grid cells. We then selected which of these 18 spatio-temporal scales could best predict whale

presence using an information criterion approach detailed below. For SST ($^\circ\text{C}$), we used data collected by NOAA's Advanced Very High Resolution Radiometer (AVHRR) aboard NOAA's Polar Operational Environmental Satellites. We used the Pathfinder Version 5.0 Sea Surface Temperature data set (Kilpatrick et al. 2001) as processed by the University of Miami's Rosenstiel School of Marine and Atmospheric Science and NOAA's National Oceanographic Data Center. The data were downloaded at a resolution of 0.05° longitude × 0.05° latitude from the NOAA CoastWatch program website (see above). Monthly and weekly values of SST were associated with each data point, together with the variability in weekly SST (expressed as slope of the relative surface and calculated in the Manifold GIS system) and the deviation of the SST in each $0.05 \times 0.05^\circ$ cell from the monthly median. This latter was implemented to allow whale presence to respond to relative rather than absolute temperatures, because the median SST varied substantially between years. SSH deviation (hereafter 'SSH') was the difference between the measured height of the sea surface and the expected mean height, calculated by reviewing

historic altimetry data. This is generally influenced on a daily basis by the tidal cycle and, on a longer term, by the overall water circulation. We used data from the AVISO (Archiving, Validation and Interpretation of Satellite Oceanographic data) program, which was available at a resolution of 0.25° longitude \times 0.25° latitude from the NOAA CoastWatch program website (see above). Finally, surface wind direction data, expressed as an angle from 0° to 359° where 0° corresponds to the north, were also available from the NOAA CoastWatch program website (see above), originating from the SeaWinds sensor on NASA's QuikSCAT satellite and processed using NASA-developed algorithms (Freilich 2000). The data were downloaded at 0.25° resolutions in the form of the 2 components of wind velocity: the zonal wind (W-E component) and the meridional wind (S-N component), which were then combined to obtain overall wind direction.

Other variables. In order to assess potential differences in the probability of encountering sperm whales between the different research seasons, year was also included as a factor in the models. Additionally, the latitude and longitude of each location were included; these are generally used to account for unknown predictors and, consequently, they compete with the available ones in trying to explain whale distribution. In this case, their inclusion had the effect of making each point unique. Each point was thus visited only once, so no offset term was required in the model to account for effort. Because of this convenience, latitude and longitude were not subjected to model selection and were retained in the models regardless of their significance.

Data analysis. A Generalized Additive Model (GAM) framework was used to model the relationship between sperm whale presence/absence at each GPS fix and the predictors described above (Hastie & Tibshirani 1990, Wood 2006). This flexible, data-driven approach has already been extensively used in the study of cetacean distributions (e.g. Forney 2000, Hastie et al. 2005, Ferguson et al. 2006). Specifically, a binomial-based GAM with a logit link function was employed to model sperm whale probability of presence.

Whale follows lasted from a few hours up to 3 d. During these periods, the whales were always on the move, sometimes zigzagging and other times moving rectilinearly but generally displacing consistently away from the position where they were initially encountered (Whitehead 2003) and thereby typically traversing a range of environmental conditions. A traditional approach would entail substantial subsetting of the data to reduce serial autocorrelation between successive points; for instance, only the initial, final or middle positions would generally be considered as a presence point in the analysis (e.g. Gannier & Praca

2007). This approach has the disadvantage of discarding information on which habitat whales were selecting during the follow period. Using all the data points to account for the occurrence of the animals over the whole habitat where they were tracked is therefore a preferable approach (Embling 2008). The problem is that GAM inference relies upon independence between model residuals, an assumption that is violated by using all the points within a follow or searching transect because the conditions at each location will be similar to those at the previous location. This spatial autocorrelation leads to the underestimation of the uncertainty associated with model estimates. Data subsampling and the use of a coarser analysis resolution are potential solutions, but they also reduce sample size (Embling 2008). An alternative way to deal with this issue while retaining all the information from each follow or searching transect is to relax the independence assumption and explicitly model the correlation between the residuals using Generalized Estimating Equations (GEEs; Liang & Zeger 1986). Under this approach, data points are grouped in blocks within which residuals are allowed to be correlated, while independence is assumed between separate blocks. A correlation structure may be specified for the errors within each block that can accommodate both spatial and temporal dependence. Several options are available for the definition of the correlation structure, but GEEs are quite robust to misspecifications (Liang & Zeger 1986, Pan & Connett 2002). When unsure about the true underlying structure of the dependence between the residuals, the use of a simple working independence model is advisable (Pan 2001). Therefore, a working independence model was preferred over the specification of a correlation structure in the present study. This approach generally leads to efficient estimates for model coefficients (McDonald 1993, Pan & Connett 2002) and uses robust, modified sandwich variance estimators to produce realistic standard errors. These models combine the 'variance estimate for the specified model with a variance matrix constructed from the data' (Hardin & Hilbe 2003), thus accounting for the observed lack of independence between the residuals within blocks and inflating the standard errors to make robust inference. Sharples (1989, in McDonald 1993) showed by means of a simulation study that this so-called empirical variance estimator performs well when compared to other estimators that assume a specific model for the correlation. This approach has already been successfully applied to the study of cetacean habitat preference by Panigada et al. (2008).

The geepack library (Yan et al. 2010) in R (R Development Core Team 2009) was used to fit binomial-based GEE-GLMs with a logit link and a working cor-

relation structure defined by block (i.e. the follows and the searching transects). The splines library (R Development Core Team 2009) then allowed us to build cubic B -splines within the GEE-GLM, thus leading to a GEE-GAM. All the covariates were considered either as linear terms or as 1-dimensional smooth terms (4 degrees of freedom), modelled as cubic B -splines with one internal knot positioned at the average value of each variable. The best subset of variables to retain in the model was identified by means of an approximate form of the quasi-likelihood under the independence model criterion (QIC; Pan 2001), a modified version of the Akaike Information Criterion (AIC) that accounts for the fact that GEEs are based on quasi-likelihood. The approximation is called QIC_u (Hardin & Hilbe 2003), and it can be employed to compare models in a stepwise selection. The QIC_u score is provided by the R library *yags* (Carey 2004). Including all the covariates at the different scales in the same model would have caused instability because of the strong collinearity between them. This necessitated the development of an ad hoc procedure to carry out variable selection. Each covariate was first evaluated at its different spatial and temporal scales in order to select only one to be tested in the full model. This was done by comparing the QIC_u score of a null model (i.e. containing only latitude and longitude) with the score of a series of models, each additionally containing the covariate in question at one of the scales under examination. Because the package *splines* does not allow the selection of the appropriate degree of smoothness, the inclusion of each covariate as a linear term was also evaluated. Once the most appropriate spatial and temporal scale and form (linear or smooth) were identified, a full model was fitted, containing all the covariates selected by the above procedure together with the ones available at a single scale. First, the QIC_u score was used to select the best form (linear or smooth) in which to incorporate these latter single-scale covariates. Then, a manual stepwise selection was carried out, where a series of reduced models was fitted at each step, containing all the terms but one, and the model with the lowest QIC_u was used in the following step. This procedure was continued until each of the covariates, if removed, caused the QIC_u score to increase. Year was evaluated as a factor, and latitude and longitude were not subject to model selection (see 'Other variables'). Repeated Wald's tests (anova. *geeglm* function in the *geepack* library) were carried out on the final model to determine the significance of each covariate (Hardin & Hilbe 2003). Non-significant variables were removed one by one, and the significance of the others was re-tested until all the associated p-values were smaller than 0.05. At this point,

the final model was obtained. Our aim in the study was to accurately understand the factors driving sperm whale distribution, for which the inclusion of unnecessary predictors would be confusing rather than maximising model fit. We thus adopted this intentionally conservative 2-step model selection procedure because some of the predictors selected using the QIC_u returned high p-values under the Wald's tests. The entire procedure was applied to the full dataset and repeated on 2 subsets, one only containing the follows of single individuals (hereafter 'singleton subset') and the other only the groups ('group subset') in order to assess any potential variation in the habitat use by whales showing different grouping behaviours.

The contribution of the explanatory variables in the final model was visualised by means of partial residual plots of the estimated relationship between the response (on the link scale) and each predictor coupled with confidence intervals based on the GEE standard errors. The package *ggplot2* (Wickham 2009) was used for this purpose. We evaluated model performance with confusion matrices, which compare the binary predictions to the observed values and report the true and false presences and the true and false absences, thus summarising the goodness-of-fit of the model (Fielding & Bell 1997). In order to build a confusion matrix, an appropriate cut-off probability value has to be chosen, beyond which a prediction is considered as a presence. Rather than selecting arbitrary cut-offs that could introduce a bias, if inappropriate (Boyce et al. 2002), a Receiver Operating Characteristic (or ROC curve) can be used (library *ROCR* in R; Sing et al. 2009); this curve plots the sensitivity (or true positive rate, i.e. the proportion of correctly classified presences) versus the specificity (or false positive rate, i.e. the proportion of incorrect presence classifications) for a binary response as the cut-off probability is varied (e.g. Zweig & Campbell 1993). The best cut-off probability for the observed data corresponds to the point where the distance between the ROC curve and the 45° diagonal is maximised, which was identified by calculating the perpendicular distance of each point of the ROC plot from the line of slope 1. Additionally, the area under the curve (AUC) can be used as an indication of the model performance (the closer to 1, the better the model; Boyce et al. 2002); this was also computed using the *ROCR* library.

A grid of cells 2×2 n miles was created and the centroids used for prediction. To avoid extrapolations, only points within the margins of the study area were included. GIS tools were used to return the values of the covariates for each point. For the time-varying variables (e.g. SST), average values were computed. When year was kept in the model, the average year

was used (2005) and the values of the time-varying covariates were averaged across this time period. The probability of whale presence in each location was then predicted in R using the final model. All predictions were made on the response scale (i.e. a value of probability between 0 and 1). The predicted values were finally visualised in a map, where a gradation of colours helped to locate the areas of higher predicted probability of encountering the whales.

The annotated R code developed for all the analysis is made available as an electronic supplement at www.int-res.com/articles/suppl/m436p257_supp/.

RESULTS

A total distance of 11 099 km was covered in acoustic search mode across the 6 yr. Sperm whales were encountered 56 times (of which 17 were sightings of groups), for a total of 1720 km of acoustic contact (Table 1; Fig. 1a).

Entire data set

The selected chl *a* covariate was the April peak with SD of the weights equal to 1/10 of the lag on a spatial scale of 20×20 n miles and in a linear form. For SST, weekly values (in a linear form) were retained. Slope was selected on the 20×20 n mile spatial scale, again as a linear term. These covariates were then entered in the full model with those remaining. The final model after variable selection retained depth, slope, aspect, weekly SST and chl *a* as predictors of whale presence. Wald's tests for the significance of these covariates excluded slope, SST and chl *a* so that the final model only included depth ($p = 0.0165$) and aspect ($p = 0.0002$) in addition to latitude and longitude. The realistic modelling of the autocorrelation in our data means that the confidence intervals around the modelled relationships remain wide (Fig. 2). Therefore, the detailed form of the best fit relationships must be interpreted with caution. Nevertheless, our conservative model selection procedure ensures that the retained variables are genuinely important predictors of sperm whale distribution because our model performed well in terms of its fit to the data. The cut-off for the construction of the confusion matrix was chosen at a probability of 0.2516, as indi-

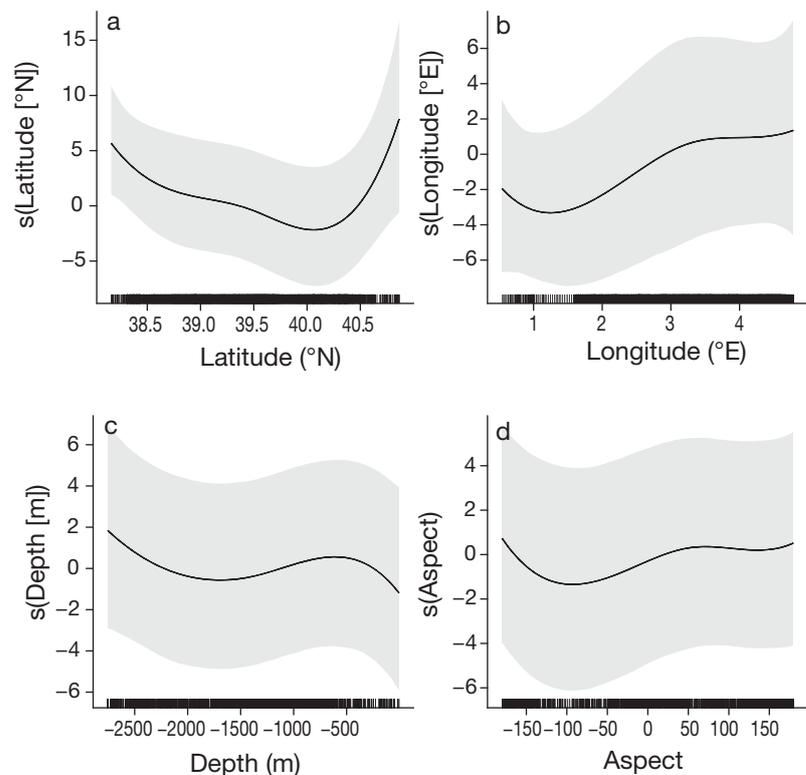


Fig. 2. Sperm whale presence modelled as a smooth function of (a) latitude, (b) longitude, (c) depth and (d) aspect. Shaded areas represent 95% CIs (GEE-based). A rug plot with the actual data values is also shown

cated by the ROC curve. The resulting matrix suggested that the model predicts correctly 72% of the presences and 67% of the absences. The area under the ROC curve (AUC) was equal to 0.77 (SD = 0.007), confirming a good model performance and providing further assurance that the model was not overfitted. The probability of sperm whale occurrence declines with decreasing depth (i.e. approaching the coast) with a second peak around 500 m depth (Fig. 2c). Additionally, sea floor oriented eastward and southward (i.e. aspect > 0 and < -150) seemed to be preferred (Fig. 2d). The prediction map showed that the model correctly identified the area north of Mallorca and Menorca and west of Mallorca as a region where the probability of encountering sperm whales is low (no encounters across the 6 yr), while defined hot-spots are located south, east and north-east of these 2 islands and south of the 2 western islands (Fig. 1b). Despite its overall close fit to the data, the model failed to recognise some regions around Menorca as suitable habitat for the species; whales were encountered closer to the coast both to the south and to the north-east of this island, but low probabilities were predicted here.

Analysis by grouping behaviour: groups and single individuals

For the subset of data originating from social groups, the preliminary investigation on the multi-scale covariates retained slope on the 2×2 n mile scale and SST on a weekly scale, both as smooth terms (i.e. non-linear). The February peak of chl *a* with standard deviation of the weights equal to 1/10 of the lag and spatial scale 20×20 n miles was also selected for the full model as a smooth term. The subsequent variable selection, based first on QIC_u scores and then on Wald's tests, produced a final model with latitude, longitude, weekly SST ($p = 0.02$) and slope (2×2 n mile resolution; $p = 8.155 \times 10^{-8}$) as cubic *B*-splines with 4 degrees of freedom (Fig. 3). The considerations about confidence intervals noted with respect to the overall analysis above also apply to these results. While sperm whale groups seemed to prefer colder waters (Fig. 3c), they also tended to occur with lower probability in areas with intermediate slope gradients (Fig. 3d). The ROC curve selected a cut-off probability of 0.2205 for the construction of the confusion matrix; the model correctly predicted 67% of the presences and 89% of the absences. The AUC was 0.85 (SD = 0.009), again showing good model perfor-

mance. The prediction map (Fig. 1c) was broadly comparable to the one obtained from the model fitted on the entire data set; some areas south, east and north-east of the islands were identified as suitable habitat for sperm whale groups, even though the southern hot-spots appeared to be more restricted. Interestingly, the model performed better on the north-eastern side of Menorca where the previous one predicted whale presence poorly.

For the singleton subset (i.e. including only the follows of single individuals), the investigation of multi-scale covariates selected monthly values of SST (as a linear term) and slope on the 20×20 n mile scale (as a smooth term). For chl *a*, in this case, variable selection identified 2 different temporal peaks on the 2 spatial scales, so that both of them were kept in the subsequent analysis; the April peak (SD of the weights: 1/10 of the lag) was retained as a smooth term on the 20×20 n mile scale, while the June peak (SD of the weights: 1/5 of the lag) was selected as a smooth term on the 2×2 n mile scale. The final model after variable selection contained latitude, longitude, year ($p = 0.016$), monthly SST ($p = 0.02$) and aspect ($p = 2.163 \times 10^{-7}$), the latter as a smooth term with 4 degrees of freedom (Fig. 4). Single individuals concentrated in areas with high monthly average surface temperatures (Fig. 4d). There were significant differences between years (Fig. 4c) with higher sighting probability in 2007. Aspect showed a similar pattern to the analysis of the complete dataset, albeit with much more associated uncertainty. Again, confidence intervals were wide around most of the modelled relationships, but the selected model performed well. The cut-off probability selected through the ROC was 0.1162. The confusion matrix showed that the model predicted correctly 91% of the presences and 65% of the absences. As in the 2 previous cases, the AUC was quite high (0.83; SD = 0.007). The predictive map (built with 2005 as the year factor and the relative values of SST) differed from the previous 2 (Fig. 1d); single individuals only used the area off the southern and south-eastern coast of the islands, while low probabilities were predicted for the eastern and north-eastern sides (where only one individual was ever encountered). Although not shown here, 5 additional maps were drawn using the other years (i.e. not 2005) and the associated average

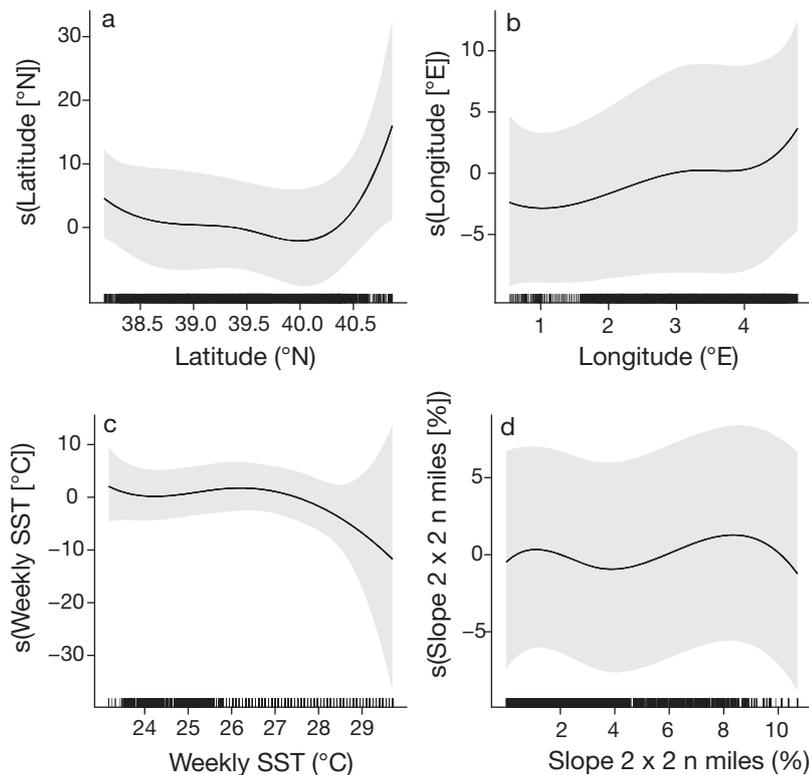


Fig. 3. Presence of sperm whale groups modelled as a smooth function of (a) latitude, (b) longitude, (c) weekly sea surface temperature (SST) and (d) slope at 2×2 n miles. Shaded areas represent 95% CIs (GEE-based). A rug plot with the actual data values is also shown

of the modelled relationships, but the selected model performed well. The cut-off probability selected through the ROC was 0.1162. The confusion matrix showed that the model predicted correctly 91% of the presences and 65% of the absences. As in the 2 previous cases, the AUC was quite high (0.83; SD = 0.007). The predictive map (built with 2005 as the year factor and the relative values of SST) differed from the previous 2 (Fig. 1d); single individuals only used the area off the southern and south-eastern coast of the islands, while low probabilities were predicted for the eastern and north-eastern sides (where only one individual was ever encountered). Although not shown here, 5 additional maps were drawn using the other years (i.e. not 2005) and the associated average

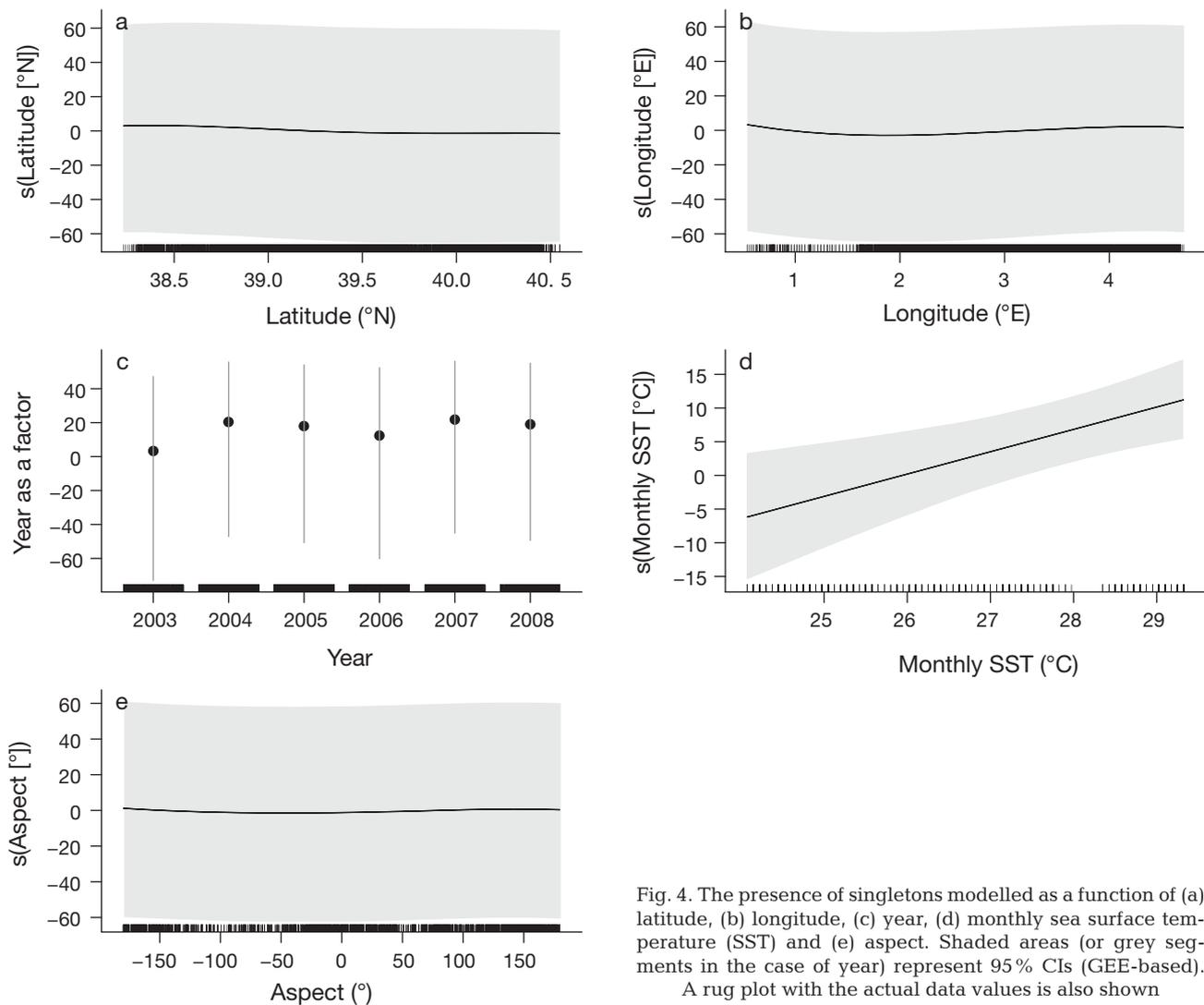


Fig. 4. The presence of singletons modelled as a function of (a) latitude, (b) longitude, (c) year, (d) monthly sea surface temperature (SST) and (e) aspect. Shaded areas (or grey segments in the case of year) represent 95% CIs (GEE-based). A rug plot with the actual data values is also shown

value of SST; despite some small-scale differences, the general pattern was consistent, suggesting that the segregation of single whales in a smaller portion of the study area is not an artefact.

DISCUSSION

Our data show that in the period from 2003 to 2008 sperm whales did not use the study area uniformly but were concentrated in the southern, eastern and north-eastern waters of the Balearic archipelago. No whale was ever encountered in the region north and west of the archipelago, and the final model correctly associated this area with low occurrence probabilities. This preference seems to be quite stable across time; the visual inspection of yearly data shows that the whales were consistently encountered mainly in those areas that have been identified as hot-spots. In addition, no

differences in sighting probability were found between research seasons.

Gannier et al. (2002) and Gannier & Praca (2007) hypothesised a spatially bimodal distribution for this species in the western basin, an idea that was subsequently supported by Azzellino and colleagues (2008) in the *Pelagos* Sanctuary (Ligurian Sea). As a result of their opportunistic feeding strategy, Mediterranean sperm whales were found both over the continental slope, where habitat selection is believed to be driven mainly by bathymetric features, and in the offshore waters (i.e. far from topographic singularities) where the animals appear to respond to the position of thermal fronts (Gannier & Praca 2007). This was explained as the result of the positive influence of both complex topography (steep slopes, seamounts and canyons) and downwelling/upwelling water movements (associated with frontal zones) on high trophic level biomass. Even though the prey targets might change, sperm whales

would therefore be able to exploit profitable food resources in both habitats (Gannier et al. 2002). Similarly to other studies where the research effort was likewise focused on the continental slope and adjacent waters (e.g. Cañadas et al. 2002), topography was thus found to be the driving factor affecting sperm whale distribution in the study area. Although physiographic variables were competing with spatial covariates (latitude and longitude) in model selection and despite the robust standard errors used in the hypothesis tests, the present analysis provides compelling evidence that bottom depth and aspect are significantly influencing the presence of sperm whales in the Balearic region. Specifically, sperm whales seem to prefer waters deeper than 2000 to 2500 m, but they are also present over the 500 to 1000 m contour. Additionally, they concentrate in areas where the bottom aspect is approximately between 0° and 210° , i.e. where the seafloor is oriented north-eastward, eastward or southward. Depth might be associated with the bathymetric zonation of cephalopod assemblages (Quetglas et al. 2000), and a comparable range has been identified by other studies both within the Mediterranean (e.g. Cañadas et al. 2002 in the Alboran Sea) and outside the basin (e.g. Davis et al. 2002 in the Gulf of Mexico or Embling 2008 off the west coast of Scotland). Slope aspect could interact with water circulation to determine the downwelling/upwelling movements that are believed to influence the availability and concentration of sperm whale prey, giving the significant relationship we found. In contrast with other studies, we found no relationship with the steepness of the slope, which has been deemed crucial in aggregating prey (e.g. Praca et al. 2009), even though we tested for its potential effect at 3 different spatial scales. In short, a steep slope alone might be insufficient to support sperm whale presence if it is not oriented correctly, possibly in relation to the directionality of the main water currents. This idea could help to reconcile the contradictory results regarding the role of slope as a predictor of sperm whale presence in the Mediterranean (see, for example, Praca & Gannier 2008 versus Cañadas et al. 2002). The interaction between water flow and complex seafloor characteristics appears to be relevant to the species' distribution in other areas (e.g. Jaquet 1996, Davis et al. 2002, Tynan et al. 2005, Skov et al. 2008). The combination of oceanographic and topographic features is likely to promote vertical and horizontal water movements that enhance primary productivity and thus sustain a richer biomass along the entire trophic web (Tynan et al. 2005). Consequently, these areas become centers of trophic transfer where dense patches of food are predictably available to all top predators (Davis et al. 2002, Yen et al. 2004).

While groups were distributed throughout the preferred areas identified by the overall model, the hotspots in the southern waters were smaller and different variables were retained by model selection. Groups showed a bimodal relationship with slope gradient on a small scale (2 n miles) with a preference for either steep or flat bottom gradients. This pattern may reflect a short-term bimodality in their habitat use resulting from different habitat preferences associated with different activities, such as foraging and socialising. For example, Lusseau & Higham (2004) found that bottlenose dolphins use a different habitat when involved in social activities. We did not have enough data to partition by activity, so further research is required to assess this hypothesis. In addition, sperm whale groups appeared to prefer cooler waters. A negative relationship with temperature has already been documented (Jaquet 1996, Rendell et al. 2004, Embling 2008), and Rendell et al. (2004) speculated that lower temperatures might correspond to a better habitat for sperm whale prey. A more direct effect on the prey (rather than the enhancement of primary productivity) would also justify the small temporal scale (weekly values) that was found to best predict sperm whale presence.

In contrast, single animals were encountered in a limited portion of the area identified by the model using the overall data set, being restricted to 2 main regions south and south-east of the islands. These solitary individuals showed a stereotyped diving pattern for long time spans, a behaviour that is typically associated with foraging males, although the sex of these individuals was not verified. Their probability of presence varied significantly between years, possibly suggesting a variable profitability of the area as a feeding ground. Males are known to move between this region and other parts of the western basin (Drouot-Dulau & Gannier 2007) so that alternative areas might be chosen to forage according to variation in resource availability. In contrast to groups, we found an increased probability of presence for singletons associated with higher monthly SST. This was unexpected, given the literature listed above, but a segregation from social groups might help to explain the discrepancy.

In summary, the habitat exploited by sperm whale groups and solitary individuals appeared to differ. This segregation appears to be mainly described by contrasting responses to sea surface temperature, with groups less frequently found in warm waters where, conversely, the probability of sighting single individuals was higher. The reasons for this separation are unknown, but different ecological requirements may represent one potential explanation. While single animals can focus the choice of their habitat on prey availability, groups might have to make a trade-off between

contrasting needs, such as the search for food and social interactions or the presence of immature individuals, which are known to heavily affect the behaviour of adults (Whitehead 1996). This could also lead individuals in groups to target other prey species with different spatial distributions (Whitehead 2003). Alternatively, groups could outcompete singletons, forcing them to use suboptimal warmer waters (Whitehead et al. 1989). Whitehead (2003) hypothesised that a reduced feeding success of adult males in those places where they overlap with groups could be due to the competition with groups of smaller individuals. Demonstrating that a sexual segregation exists on smaller scales could offer interesting insights into the mechanisms that have led to its development at a large scale. These findings also have important methodological implications, as they point out that a distinction between groups and single animals is necessary for an accurate evaluation of sperm whale space use. While in the southern waters pooling all the encounters in a single data set only resulted in the identification of larger hot-spots, fitting a separate model for groups determined a better goodness-of-fit on the eastern side of Menorca where the overall model performed poorly. Some contradictory results available in the literature could also derive from pooling the observations from sperm whales in different behavioural or social states; for instance, no relationship with surface temperature would have been found in this work if the data were not split into the 2 subsets.

The present study contributes to the development of a modelling framework for the analysis of sperm whale habitat use in 3 specific ways. (1) The non-linear relationships between whale probability of presence and most environmental predictors suggests that GLMs may be insufficient to capture the species' habitat preference correctly; the importance of environmental variables could have been missed or misinterpreted if underlying shapes of their influence were assumed (Hastie et al. 2005). The selection we performed on how each variable should enter the model settled on non-linear forms for all except SST for singleton animals, which supports this view. (2) Although the details of the smooth terms employed here could not be optimised as suggested by Wood (2006), the use of one-dimensional *B*-splines allowed autocorrelation to be accounted for via the GEEs (as in Panigada et al. 2008). Cetacean observations are often spatially and temporally autocorrelated. In particular, sperm whale acoustic and tracking data are unavoidably so, because the same individual or group is generally followed over consequent sampling points (Embling 2008). Correcting for this dependence is thus critical to correctly estimate uncertainty (Redfern et al. 2006). These data are expensive and time-consuming to col-

lect and so discarding intermediate data points to obtain independence is undesirable because it reduces the already limited effective sample size (Redfern et al. 2006). Our study confirms that GEEs are an appropriate tool to deal with this issue. (3) The effect of some of the covariates was evaluated at different spatial and temporal scales in order to account for a potential delay in sperm whale responses to oceanographic and physical processes, such as the increase in chl *a* surface concentration, SST or and the effects of seafloor steepness on water circulation. The choice of an arbitrary scale at which to investigate the significance of these factors can lead to confusing and contradictory results, so a multi-scale approach is required (Jaquet 1996). For example, we found that the presence of sperm whale groups was influenced by SST on a weekly scale, while single individuals appeared to respond to monthly averages of the same variable. We used a multi-scale approach based on a series of weighted averages to model lags of approximately 5, 3 and 1 mo between the peak in chl *a* surface concentration and whale occurrence (4 mo is the time lag generally recognised to exist between phytoplankton and cephalopods; Jaquet 1996). Despite the evaluation of multiple scales and the correlation that previous studies have found outside the Mediterranean (Jaquet 1996), chl *a* did not show any significant relationship with the species in the area. Possibly, this factor affects sperm whale presence at much larger spatial scales, which might include the entire region of the Balearics. Indeed, Jaquet & Whitehead (1996) recorded a correlation in the Pacific at a scale of more than 320 n miles (i.e. about twice the size of our entire study area).

On the whole, these analytical techniques gave good predictions of sperm whale presence around the Balearic Islands both when the entire data set was considered (70% of the points were correctly classified) and when it was split into the 2 subsets (78% correct classifications were obtained for both the group and the singleton subsets). Nonetheless, our approach almost certainly oversimplifies the complex relationships between whale presence and indirect environmental variables, not least by the exclusion of unknown relevant covariates. Direct information on the availability and movements of sperm whale prey would thus be beneficial for a better understanding of the relationship of this species with its environment (Jaquet & Gendron 2002, Friedlaender et al. 2006).

The identification of sperm whale key areas is the first step in developing specific conservation measures for the Balearic archipelago. For instance, habitat modelling results could help define the boundaries of candidate Marine Protected Areas (MPAs) by providing a better description of the species distribution compared to other simple measures of occurrence (e.g. encounter

rates; Cañadas et al. 2005). The establishment of a network of MPAs encompassing the identified hot-spots in this important feeding and breeding ground would probably have beneficial effects on the entire 'Endangered' stock of Mediterranean sperm whales (Reeves & Notarbartolo di Sciara 2006, Notarbartolo di Sciara & Birkun 2010). Evaluating the stability of the species habitat preference in time and space and validating it with independent data is however necessary before the effective implementation of any measure. Particularly, further research is needed in the regions at the north-eastern and at the south-western ends of the study area, which were identified as important hot-spots of whale presence. The searching effort was limited in both of these areas and additional evidence is required to exclude any edge effects (e.g. resulting from their extreme values of depth or latitude and longitude). The high probability of presence predicted here must therefore be interpreted with caution. Moreover, no extrapolation outside the survey region is advisable until the mechanisms underlying the observed use of space are rigorously tested (Ferguson et al. 2006, Panigada et al. 2008). Similar habitats or processes may be characterised elsewhere by different indirect variables (Hamazaki 2002), and the effect of the identified predictors could also differ outside the sampled area (Panigada et al. 2008).

Future research should aim to improve the analytical framework presented here. For instance, testing more sophisticated correlation structures for the GEEs would be interesting; a user-defined correlation matrix that better reflects the true structure of the autocorrelation could increase the efficiency of model estimates, although the issue is controversial (Pan & Connett 2002). Moreover, the interactions among the available covariates (e.g. between seafloor characteristics and the processes affecting water circulation, such as the main winds or currents) should be included in the models to better describe the complex underlying relationships between environmental proxies and whale presence. The role of other ecologically relevant processes in shaping habitat use also remains to be tested (e.g. the interspecific competition with other squid-eating cetaceans, as in Waring et al. 2001).

In parallel to the refinement of the analytical techniques, an extensive research effort is required to provide new systematic data for the construction of a global model of sperm whale distribution covering both offshore and continental slope regions of the Mediterranean and other ocean basins. Basin-wide data would also offer reliable indications on the densities, movements, site fidelity and habitat segregation of single males and social groups. We have shown how new analytical approaches can offer robust insights into the factors driving the distribution of this species

in the critical region of the Balearic archipelago, using the kind of data that are commonly collected during surveys at sea. This information is increasingly needed to improve our limited understanding of the species ecology in the Mediterranean Sea, assess the current status of this population and inform effective conservation and management efforts.

Acknowledgements. This work uses the data collected during the Balearics Sperm Whale Project, a research program started in 2003 by L.R. in collaboration with A. Cañadas (Alnitak, Spain) and C. Mundy (One World Wildlife, UK) to investigate the biology and ecology of sperm whales around the Balearic Islands. We are grateful to One World Wildlife and the Whale and Dolphin Conservation Society for financial support. Our fieldwork was invaluable supported by the Balearic Government Office of Fisheries Management, especially by J. M. Brotons of that office. We thank all the volunteer assistants who took part in data collection, C. Booth for advice on GEEs, A. Cañadas for providing information on the environmental covariates and C. Blight for help with GIS analysis.

LITERATURE CITED

- Acosta J, Canals M, López-Martínez J, Muñoz A and others (2003) The Balearic Promontory geomorphology (western Mediterranean): morphostructure and active processes. *Geomorphology* 49:177–204
- Azzellino A, Gaspari S, Airolidi S, Nani B (2008) Habitat use and preferences of cetaceans along the continental slope and the adjacent pelagic waters in the western Ligurian Sea. *Deep-Sea Res I* 55:296–323
- Benoit-Bird KJ, Au WWL (2003) Prey dynamics affect foraging by a pelagic predator (*Stenella longirostris*) over a range of spatial and temporal scales. *Behav Ecol Sociobiol* 53:364–373
- Boyce MS, Vernier PR, Nielsen SE, Schmiegelow FKA (2002) Evaluating resource selection functions. *Ecol Model* 157: 281–300
- 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
- Cañadas A, Sagarminaga R, García-Tiscar S (2002) Cetacean distribution related with depth and slope in the Mediterranean waters off southern Spain. *Deep-Sea Res I* 49: 2053–2073
- Cañadas A, Sagarminaga R, De Stephanis R, Urquiola E, Hammond PS (2005) Habitat preference modelling as a conservation tool: proposals for marine protected areas for cetaceans in southern Spanish waters. *Aquat Conserv* 15:495–521
- Carey VJ (2004) yags: yet another GEE solver. R package, version 4.0-2. Available at: r-forge.r-project.org/R/?group_id=329
- Croll DA, Marinovic B, Benson S, Chavez FP, Black N, Ter-nullo R, Tershy BR (2005) From wind to whales: trophic links in a coastal upwelling system. *Mar Ecol Prog Ser* 289:117–130
- Davis RW, Ortega-Ortiz JG, Ribic CA, Evans WE and others (2002) Cetacean habitat in the northern oceanic Gulf of Mexico. *Deep-Sea Res I* 49:121–142
- Drouot V, Bérubé M, Gannier A, Goold JC, Reid RJ, Palsbøll PJ (2004) A note on genetic isolation of Mediterranean

- sperm whales (*Physeter macrocephalus*) suggested by mitochondrial DNA. *J Cetacean Res Manag* 6:29–32
- Drouot V, Gannier A, Goold J (2004a) Summer social distribution of sperm whales (*Physeter macrocephalus*) in the Mediterranean Sea. *J Mar Biol Assoc UK* 84:675–680
- Drouot V, Gannier A, Goold J (2004b) Diving and feeding behaviour of sperm whales (*Physeter macrocephalus*) in the northwestern Mediterranean Sea. *Aquat Mamm* 30: 419–426
- Drouot-Dulau V, Gannier A (2007) Movements of sperm whale in the western Mediterranean Sea: preliminary photo-identification results. *J Mar Biol Assoc UK* 87: 195–200
- Embling CB (2008) Predictive models of cetacean distributions off the west coast of Scotland. PhD dissertation, University of St Andrews
- Engelhaupt DT (2004) Phylogeography, kinship and molecular ecology of sperm whales (*Physeter macrocephalus*). PhD dissertation, Durham University
- Ersts PJ, Rosenbaum HC (2003) Habitat preference reflects social organization of humpback whales (*Megaptera novaeangliae*) on a wintering ground. *J Zool* 260:337–345
- Ferguson MC, Barlow J, Reilly SB, Gerrodette T (2006) Predicting Cuvier's (*Ziphius cavirostris*) and *Mesoplodon* beaked whale population density from habitat characteristics in the eastern tropical Pacific Ocean. *J Cetacean Res Manag* 7:287–299
- Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv* 24:38–49
- Forney KA (2000) Environmental models of cetacean abundance: reducing uncertainty in population trends. *Conserv Biol* 14:1271–1286
- Frantzis A, Alexiadou P, Paximadis G, Politi E, Gannier A, Corsini-Foka M (2003) Current knowledge of the cetacean fauna of the Greek Seas. *J Cetacean Res Manag* 5: 219–232
- Frederiksen M, Edwards M, Richardson AJ, Halliday NC, Wanless S (2006) From plankton to top predators: bottom-up control of a marine food web across four trophic levels. *J Anim Ecol* 75:1259–1268
- Freilich M (2000) SeaWinds algorithm theoretical basis document, NASA ATBD-SWS-01, Jet Propulsion Laboratory, California Institute of Technology
- Friedlaender AS, Halpin PN, Qian SS, Lawson GL, Wiebe PH, Thiele D, Read AJ (2006) Whale distribution in relation to prey abundance and oceanographic processes in shelf waters of the Western Antarctic Peninsula. *Mar Ecol Prog Ser* 317:297–310
- Fu G, Baith KS, McClain CR (1998) SeaDAS: the SeaWiFS data analysis system. Proc 4th Pacific Ocean Remote Sensing Conf, Qingdao, China, July 28–31, 1998, p 73–79
- Gannier A, Praca E (2007) SST fronts and the summer sperm whale distribution in the north-west Mediterranean Sea. *J Mar Biol Assoc U K* 87:187–193
- Gannier A, Drouot V, Goold JC (2002) Distribution and relative abundance of sperm whales in the Mediterranean Sea. *Mar Ecol Prog Ser* 243:281–293
- García E, Tintoré J, Pinot JM, Font J, Manriquez M (1994) Surface circulation and dynamics of the Balearic Sea. In: La Violette PE (ed) Seasonal and interannual variability of the western Mediterranean Sea. *Coast Estuar Stud* 46: 73–91
- Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecol Model* 135:147–186
- Hamazaki T (2002) Spatiotemporal prediction models of cetacean habitats in the mid-western North Atlantic Ocean (from Cape Hatteras, North Carolina, USA to Nova Scotia, Canada). *Mar Mamm Sci* 18:920–939
- Hardin JW, Hilbe JM (2003) Generalized estimating equations. Chapman & Hall/CRC Press, London
- Hastie GD, Wilson B, Wilson LJ, Parsons KM, Thompson PM (2004) Functional mechanisms underlying cetacean distribution patterns: hotspots for bottlenose dolphins are linked to foraging. *Mar Biol* 144:397–403
- Hastie GD, Swift RJ, Slesser G, Thompson PM, Turrell WR (2005) Environmental models for predicting oceanic dolphin habitat in the Northeast Atlantic. *ICES J Mar Sci* 62:760–770
- Hastie TJ, Tibshirani RJ (1990) Generalised additive models. Monographs on statistics and applied probabilities, Vol 46. Chapman & Hall, London
- Heithaus MR, Dill LM (2002) Food availability and tiger shark predation risk influence bottlenose dolphin habitat use. *Ecology* 83:480–491
- Jakimchuk RD, Ferguson SH, Sopuck LG (1987) Differential habitat use and sexual segregation in the Central Arctic caribou herd. *Can J Zool* 65:534–541
- Jaquet N (1996) How spatial and temporal scales influence understanding of sperm whale distribution: a review. *Mammal Rev* 26:51–65
- Jaquet N, Gendron D (2002) Distribution and relative abundance of sperm whales in relation to key environmental features, squid landings and the distribution of other cetacean species in the Gulf of California, Mexico. *Mar Biol* 141:591–601
- Jaquet N, Whitehead H (1996) Scale-dependent correlation of sperm whale distribution with environmental features and productivity in the South Pacific. *Mar Ecol Prog Ser* 135: 1–9
- Kilpatrick KA, Podestá GP, Evans R (2001) Overview of the NOAA/NASA advanced very high resolution radiometer Pathfinder algorithm for sea surface temperature and associated matchup database. *J Geophys Res C* 106: 9179–9197
- Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. *Biometrika* 73:13–22
- Littaye A, Gannier A, Laran S, Wilson JPF (2004) The relationship between summer aggregation of fin whales and satellite-derived environmental conditions in the northwestern Mediterranean Sea. *Remote Sens Environ* 90:44–52
- Litvaitis JA (1990) Differential habitat use by sexes of snowshoe hares (*Lepus americanus*). *J Mammal* 71:520–523
- Lusseau D, Higham JES (2004) Managing the impacts of dolphin-based tourism through the definition of critical habitats: the case of bottlenose dolphins (*Tursiops* spp.) in Doubtful Sound, New Zealand. *Tour Manage* 25:657–667
- Matthiopoulos J, Aarts G (2010) The spatial analysis of marine mammal abundance. In: Boyd IL, Bowen WD, Iverson SJ (eds) Marine mammal ecology and conservation: a handbook of techniques. Oxford University Press, Oxford, p 68–97
- McDonald BW (1993) Estimating logistic regression parameters for bivariate binary data. *J R Stat Soc B* 55:391–397
- NOAA (National Oceanic and Atmospheric Administration) (2006) 2-minute Gridded Global Relief Data (ETOPO2v2), National Geophysical Data Center www.ngdc.noaa.gov/mgg/fliers/06mgg01.html
- Notarbartolo di Sciara G, Birkun A Jr (2010) Conserving whales, dolphins and porpoises in the Mediterranean and Black Seas: an ACCOBAMS status report, 2010. ACCOBAMS, Monaco
- Pan W (2001) Akaike's information criterion in generalized estimating equations. *Biometrics* 57:120–125

- Pan W, Connett JE (2002) Selecting the working correlation structure in generalized estimating equations with application to the lung health study. *Statist Sinica* 12:475–490
- Panigada S, Zanardelli M, MacKenzie M, Donovan C, Mélin F, Hammond PH (2008) Modelling habitat preferences for fin whales and striped dolphins in the Pelagos Sanctuary (Western Mediterranean Sea) with physiographic and remote sensing variables. *Remote Sens Environ* 112: 3400–3412
- Pinot JM, López-Jurado JL, Riera M (2002) The CANALES experiment (1996–1998). Interannual, seasonal and meso-scale variability of the circulation in the Balearic Channels. *Prog Oceanogr* 55:335–370
- Praca E, Gannier A (2008) Ecological niches of three teuthophageous odontocetes in the northwestern Mediterranean Sea. *Ocean Sci* 4:49–59
- 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
- Quetglas A, Carbonell A, Sánchez P (2000) Demersal continental shelf and upper slope cephalopod assemblages from the Balearic Sea (north-western Mediterranean). Biological aspects of some deep-sea species. *Est Coast Shelf Sci* 50:739–749
- R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: www.R-project.org
- Redfern JV, Ferguson MC, Becker EA, Hyrenbach KD and others (2006) Techniques for cetacean–habitat modeling. *Mar Ecol Prog Ser* 310:271–295
- Reese D (2005) Whale bones and shell purple-dye at Motya (Western Sicily, Italy). *Oxf J Archaeol* 24:107–114
- Reeves R, Notarbartolo di Sciarra G (compilers and editors) (2006) The status and distribution of cetaceans in the Black Sea and Mediterranean Sea. IUCN Centre for Mediterranean Cooperation, Malaga, Spain
- Rendell LE, Whitehead H, Escribano R (2004) Sperm whale habitat use and foraging success off northern Chile: evidence of ecological links between coastal and pelagic systems. *Mar Ecol Prog Ser* 275:289–295
- Scott-Hayward LAS (2006) Modelling the spatial distribution of sperm whales, *Physeter macrocephalus*, in the Gulf of Mexico. MS dissertation, University of St. Andrews
- Shane SH (1995) Relationship between pilot whales and Risso's dolphins at Santa Catalina Island, California, USA. *Mar Ecol Prog Ser* 123:5–11
- Sharples KJ (1989) Regression analysis of correlated binary data. PhD dissertation, University of Washington, Seattle
- Sing T, Sander O, Beerenwinkel N, Lengauer T (2009) ROCR: visualizing classifier performance in R. R package, version 1.0-4. Available at: rocr.bioinf.mpi-sb.mpg.de/
- Skov H, Gunnlaugsson T, Budgell WP, Horne J and others (2008) Small-scale spatial variability of sperm and sei whales in relation to oceanographic and topographic features along the Mid-Atlantic Ridge. *Deep-Sea Res II* 55: 254–268
- Stewart BS, DeLong RL (1995) Double migrations of the northern elephant seal, *Mirounga angustirostris*. *J Mammal* 76:196–205
- Tynan CT, Ainley DG, Barth JA, Cowles TJ, Pierce SD, Spear LB (2005) Cetacean distributions relative to ocean processes in the northern California Current System. *Deep-Sea Res II* 52:145–167
- Waring GT, Hamazaki T, Sheehan D, Wood G, Baker S (2001) Characterization of beaked whale (*Ziphiidae*) and sperm whale (*Physeter macrocephalus*) summer habitat in shelf-edge and deeper waters off the northeast US. *Mar Mamm Sci* 17:703–717
- Whitehead H (1996) Babysitting, dive synchrony, and indications of alloparental care in sperm whales. *Behav Ecol Sociobiol* 38:237–244
- Whitehead H (2003) Sperm whales: social evolution in the ocean. University of Chicago Press, London
- Whitehead H, Papastavrou V, Smith SC (1989) Feeding success of sperm whales and sea-surface temperature off the Galápagos Islands. *Mar Ecol Prog Ser* 53:201–203
- Wickham H (2009) ggplot2: an implementation of the grammar of graphics. R package, version 0.8.5. Available at: had.co.nz/ggplot2/
- Wood SN (2006) Generalized additive models, an introduction with R. Chapman & Hall/CRC, London
- Yan J, Højsgaard S and Halekoh U (2010) geepack: generalized estimating equations package. R package, version 1.0-17. Available at: cran.r-project.org/web/packages/geepack/index.html
- Yen PPW, Sydeman WJ, Hyrenbach KD (2004) Marine bird and cetacean associations with bathymetric habitats and shallow-water topographies: implications for trophic transfer and conservation. *J Mar Syst* 50:79–99
- Zweig MH, Campbell G (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 39:561–577

Editorial responsibility: Matthias Seaman,
Oldendorf/Luhe, Germany

Submitted: February 25, 2011; Accepted: May 31, 2011
Proofs received from author(s): August 23, 2011