MODELLING SPACE-USE AND HABITAT PREFERENCE FROM WILDLIFE TELEMETRY DATA
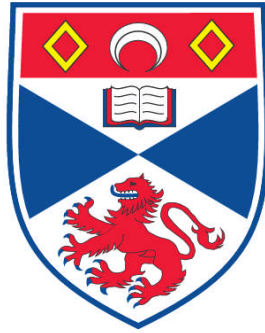
Geert Aarts

A Thesis Submitted for the Degree of PhD
at the
University of St. Andrews

MODELLING SPACE-USE AND HABITAT PREFERENCE FROM WILDLIFE TELEMETRY DATA

Geert Aarts BSc MSc

A thesis submitted to the University of St. Andrews
for the degree of Doctor of Philosophy

School of Biology
Sea Mammal Research Unit

09/02/2007

ABSTRACT

Management and conservation of populations of animals requires information on where they are, why they are there, and where else they could be. These objectives are typically approached by collecting data on the animals' use of space, relating these to prevailing environmental conditions and employing these relations to predict usage at other geographical regions. Technical advances in wildlife telemetry have accomplished manifold increases in the amount and quality of available data, creating the need for a statistical framework that can use them to make population-level inferences for habitat preference and space-use. This has been slow-in-coming because wildlife telemetry data are, by definition, spatio-temporally autocorrelated, unbalanced, presence-only observations of behaviorally complex animals, responding to a multitude of cross-correlated environmental variables.

I review the evolution of techniques for the analysis of space-use and habitat preference, from simple hypothesis tests to modern modeling techniques and outline the essential features of a framework that emerges naturally from these foundations. Within this framework, I discuss eight challenges, inherent in the spatial analysis of telemetry data and, for each, I propose solutions that can work in tandem. Specifically, I propose a logistic, mixed-effects approach that uses generalized additive transformations of the environmental covariates and is fitted to a response data-set comprising the telemetry and simulated observations, under a case-control design.

I apply this framework to non-trivial case-studies using data from satellite-tagged grey seals (*Halichoerus grypus*) foraging off the east and west coast of Scotland, and northern gannets (*Morus Bassanus*) from Bass Rock. I find that sea bottom depth and sediment type explain little of the variation in gannet usage, but grey seals from different regions strongly prefer coarse sediment types, the ideal burrowing habitat of sandeels, their preferred prey. The results also suggest that prey aggregation within the water column might be as important as horizontal heterogeneity. More importantly, I conclude that, despite the complex behavior of the study species, flexible empirical models can capture the environmental relationships that shape population distributions.

DECLARATIONS

I, Geert Aarts, hereby certify that this thesis, which is approximately 41,256 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

date: ....   signature of candidate: ....

I was admitted as a research student in June 2003 and as a candidate for the degree of PhD in June 2004; the higher study for which this is a record was carried out in the University of St Andrews between 2003 and 2007.

date: ....   signature of candidate: ....

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

date: ....   signature of supervisor: ....

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

date: ....   signature of candidate: ....

PREFACE & ACKNOWLEDGEMENTS

Three years ago I came to Scotland and anticipated to broaden my scientific interest and experience by means of self-exploration and 'trial and error'. Instead, I found that most useful insight was directly given to me by both colleagues and friends.

I'm very grateful to one person in particular, Jason Matthiopoulos. Without him I would not be writing the preface of my thesis at this particular moment. As a supervisor as well as friend he always had time to talk about science or any aspect of life. He literately guided me through my PhD and was always positive about my work, even if I wasn't, which happened regularly. He shaped me both as a scientist and as a person. I very much hope we can continue working together until the day we are old, smoke pipes, still listen to bad music and talk about the good old days.

I would also like to thank Prof. Ian Boyd for all the financial arrangements he made, such as making sure I (as a non-UK citizen) could do this PhD. During my time at SMRU I noticed how competitive science in general and marine mammal science in particular is. Things are always changing and to continue functioning one needs to be adaptive and fully aware of all funding opportunities. This is a more than a full time job and I respect Ian Boyd for taking this on with full dedication. As a PhD student, I felt I could always rely on him for financial and political support. This enabled me to fully concentrate on my scientific research.

I would also like to thank Prof. Mike Fedak for sharing his interest in ecology and marine mammals with me. Although I have a strong interest in developing statistical methods, my motivation for doing science is mostly

driven by the desire to understand the behaviour and life-history of animals. During my fieldwork experience, I always personalized a seal and wondered what it had experienced in life, which places it did visit or was about to visit and what it thought about us humans. When sitting excessive amounts of time in an office behind a computer, this fieldwork experience and the interest in the species can easily fade. Fedak's presence and our discussion always kept me strongly interested in biology and made me even more appreciative for fieldwork. I also very much enjoyed discussions I had with Dave Thompson, Bernie McConnell and Jonathan Gordon.

I would also like to thank Simon Wood, my external examiner for providing some very useful suggestions which found their way in both this thesis and a manuscript submitted to ecological monographs.

Clint Blight was always there to help me when I (among many others of SMRU) had problems with computers, network connections, extracting and transforming spatial data, printers (even when printing this thesis), and many other things. Other technical and statistical support was kindly offered by Phil Le Feuvre, Phil Lovell, Rory Beaton and Mike Lonergan.

Concerning my fieldwork, I'm particularly grateful to Simon Moss, and also to Eric Armstrong, Andrew Brierley, the crew from the Calanus and all those that helped out during the seal catching events and surveys on the west coast of Scotland. This work could also not have been done without the appropriate environmental data provided by Collin Graham, Ben Illingsworth, Steve Gontarek, Phil Gillibrand, Christian Wilson, Ben Maughan, Matthew Summers and Nick Weaver.

During my PhD I met many people in St. Andrews (and SMRU in particular) who made my stay in Scotland very enjoyable and some of them became very good friends. These are (in more or less random order) Valia and Spyros Matthiopoulos (and neighbour Peter), Clint Blight, Fernando Tempera, Thomas Goetz, Louise Cunningham, Sonja Heinrich, Clare Embling, Simon Ruddell and Lisa, Charles Paxton, Iain McWilliam, Susan Gallon, Monique McKenzie and Carl Donovan, Tiago Marques, Monica Silva, Sophie Smout, Martin Biuw, Lars Boehme, Aline Arriola, Bill Nellyer, Mark Hulme, Callan Duck, Phil Hammond, Toby Patterson, Ruth Sharples, Julian Dale, Nicola Quick, Alice Mackay, Kelly Macleod, Arliss Winship and Susan,

1.

GENERAL INTRODUCTION

2.

MODELLING HABITAT PREFERENCE AND SPACE -USE OF ANIMALS IN THE WILD

3.

MODELLING SPACE-USE AND HABITAT PREFERENCE IN GREY SEALS (HALICHOERUS GRYPUS) OFF THE EAST COAST OF SCOTLAND

4.

MODELLING SPACE-USE AND HABITAT PREFERENCE IN NORTHERN GANNETS (MORUS BASSANUS)

5.

MODELLING SPACE-USE AND HABITAT PREFERENCE IN GREY SEALS (HALICHOERUS GRYPUS) OFF THE WEST COAST OF SCOTLAND

6.

GENERAL DISCUSSION

APPENDICES

1.   GENERAL INTRODUCTION

1.1 Studies of space use and habitat preference

Animals meet their requirements for survival, growth and reproduction by exploiting available resources within the restrictions imposed by their physiology and environment. Management and conservation of animal populations require a scientific understanding of this process attained only by viewing it in its spatial context (Levin 1992). That is why an increasing number of papers in applied ecology are concerned with where a particular organism is (Kernohan et al. 1998, Blundell et al. 2001, Matthiopoulos 2003a, Matthiopoulos et al. 2004), why it is there (Johnson 1980, Manly et al. 1993, Arthur 1996, Boyce and McDonald 1999) and where else it is likely to be (Buckland and Elston 1993, Guisan & Zimmermann 2000, Guisan et al. 2002, Scott et al. 2002, Wiens 2002). These objectives are typically approached by collecting data on the animals' use of space, relating these observations to prevailing environmental conditions and employing these relations to predict usage at other points in space.

Traditionally, space-use data were obtained from transect surveys (Buckland et al. 1993) which record animals in the vicinity of a set of sampling locations. For example, in aerial or ship-based surveys the sampling locations are arranged along line transects while in trapping grids the sampling locations are point transects. Since this approach focuses on individual points in space, sightings of animals can be low, particularly for rare species living in inaccessible environments. Alternatively, wildlife tracking techniques focus on the individual animal. Radio- and satellite-telemetry have made full use of recent advances in communication and

information technology (White and Garrot 1990, Priede & Swift 1992, Fedak et al. 2002), increased the volume of data on animal usage, the spatio-temporal range of observation and, consequently, the number of field studies on space-use and habitat preference.

The concept of habitat preference (also known as resource selection) attempts to quantify the inherent needs of animals, as expressed in the environment in which they were observed (Johnson 1980, Manly et al. 1993). Studies on preference originally used hypothesis testing to compare between sets of contrasting environmental conditions. More recently, environmental conditions have been incorporated as covariates in spatial models of usage (Boyce and McDonald 1999, Buckland and Elston 1993). In addition to searching for the relevant environmental variables, these studies also quantify the animals' response to them. Because of their perceived potential for prediction, these statistical developments are fuelling the explosive increase of quantitative analyses in applied, spatial ecology (Guisan and Zimmermann 2000, Latimer et al. 2006, Pearce and Boyce 2006, Redfern et al. 2006).

However, these analyses are faced with inherent problems such as the cross-correlation between environmental variables, spatial autocorrelation in animal distribution, variable detectability of animals in different environments, various imbalances in sampling effort, unequal accessibility of different points in space and the animals' complex responses to their environment. Furthermore, the increasing power of telemetry methods presents additional demands for covariate data at the appropriate temporal and spatial scale and the focus of telemetry studies on a few individuals presents new challenges for population-level inference. Although these problems are shared by most studies of space-use and habitat preference there have been few comprehensive attempts to identify them, review existing solutions and draw on recent advances in statistical modeling for new ones. Hence, although most of these problems have, individually, received attention in the statistical literature, I am not aware of any analyses of wildlife telemetry data that deal with all of these problems simultaneously.

1.2 Definitions

In this thesis, I draw heavily from three areas of research: general spatial modelling, the study of habitat preference and the analysis of movement of individuals. Each of these areas has developed its own terminology that, in this study, would lead to redundancy (more than one term for the same concept) and confounding (more than one meaning to the same term) (Hall et al. 1993). To aid clarity and consistency, at least within this thesis, I precede the main body of this work with a brief section of definitions.

Geographical space comprises the three dimensions of latitude, longitude and altitude/depth, usually projected onto a Cartesian system of coordinates. Without loss of generality, I restrict my attention to the two dimensions of longitude and latitude. The *spatial distribution* of a species is the density of animals (or of their usage) over geographical space. According to Turchin (1998), this typifies the Eulerian, or coordinate-based approach, followed by most studies in this field, as opposed to the Lagrangian, or individual-based approach, taken by many behavioral studies.

Environmental space comprises multiple dimensions each of which represents a biotic or abiotic *environmental variable*. The environmental variables that correlate with the density of a species in space are called *covariates* to that species' spatial distribution. In some studies, longitude and latitude (or flexible functions of longitude and latitude) are used as candidate covariates instead of unknown environmental variables (e.g. Borchers et al. 1997). Also, altitude and depth which are generally not used for mapping usage and preference, are often used as candidate covariates (e.g. Sjoberg and Ball 2000, Wright et al. 2000). A *resource* is an environmental variable that can be depleted by an organism. An *environmental condition* is a particular value of an environmental variable. An *environment* is defined as a combination of conditions – a single point in environmental space.

The term habitat is more ambiguous. Traditionally (Whittaker & Root 1973) and in common use, it signified the entire region in geographical or environmental space within which a particular organism lives. So, for example, wildlife documentaries will refer to a rolling grassland scattered

with shrubs and isolated trees, that can be found between a tropical rainforest and a desert as "lion habitat". This species-specific definition of habitat is not very useful for modeling usage or preference because it does not account for gradations in density or usage. Alternatively, *habitat* can be defined as any collection of environments. Hence, grassland scattered with shrubs and isolated trees, that can be found between a tropical rainforest and a desert could be called "savannah". This enables us to identify it as the grouping of environmental conditions used by lions, but also by African elephants, wild dogs and nigriceps ants. This species-independent definition of habitat means that different organisms will use a particular habitat to variable degrees and different habitats will be used by different groups of organisms.

I prefer the second definition of habitat because it allows comparisons between species. In the simplest case, a habitat is a cuboid in environmental space defined by ranges along each of the environmental variables. Such a classification scheme groups similar environments under the same habitat and is the same for all species. The resolution of the classification (i.e. the inverse of the volume enclosed by the cuboid) gives rise to two trivial cases: Under the coarsest classification scheme possible, all environments belong to the same habitat. Conversely, in the finest classification scheme possible, each environment is a unique habitat and every point in geographical space is a unique habitat.

Intuitively, preference can be defined as the process that determines the animals' response to different habitats. However, for the purposes of statistical analysis, a usable definition for a response variable must be quantitative and measurable in the field. Deriving such a measure for habitat preference is not trivial because the term refers to a complicated, unobservable, process driven by physiological, behavioral and energetic constraints and demands. As such, habitat preference can only be measured by proxy, via observations of the individuals' use of different habitats, wherever they occur.

Early work using hypothesis testing, pointed out that preference is revealed by unequal usage of two habitats offered to the animals in equal amounts (Johnson 1980). This was later generalized by defining the lack of

preference as the use of each habitat in proportion to its overall availability in geographical space (Manly et al. 1993). This definition, implicitly and, perhaps, inappropriately (Matthiopoulos 2003b and problem 2 below) identifies overall availability of habitats with availability to the study animals.

Attempts to quantify preference as a function of environmental covariates have led to the development of Resource Selection Functions (RSFs – reviewed in Boyce and McDonald (1999)). These relate a species' spatial distribution (presence or density) to environmental variables that may or may not be depleted by the animals (so, Environmental Selection Functions might have been a more intuitive name).

With all of the above in mind, I adopt the following working definition: Given a habitat classification scheme (including the two trivial schemes outlined above), *habitat preference* is the ratio of the use of a habitat over its availability, conditional on the availability of all habitats to the study animals.

Further to enhance clarity, I use the same notational conventions and symbols throughout the thesis. I have collected these in Appendix 1.A.

1.3 PhD objectives

The overall thrust of this PhD is to understand which environmental variables influence the spatial distribution of top-marine predators and to employ such relationships to make spatial predictions of usage. I use the grey seal (Halichoerus grypus) and the gannet (Morus bassanus) as two case studies. Due to the lack of detailed quantitative information on causal environmental variables and the aspects of behaviour and life history that shape the decision of the individual I've chosen to construct an empirical model at the level of the population rather than a mechanistic, individual-based simulation   (see also §6.3.1). Currently, no statistical model-framework exists that can use telemetry data to answer these biological questions. As a consequence, development of a model-framework forms a

prerequisite for the biological objectives and a valid research aim in itself. The objectives of this PhD in the order in which they are addressed are:

1   ...to provide a comprehensive review of existing statistical methods historically used to analyse habitat preference and space use

2   ...to provide an overview of analytical challenges faced by studies on habitat preference that uses wildlife telemetry data, to review solutions and suggest appropriate alternatives where needed and combine those into a single unified statistical framework

3   ...to validate the proposed model framework using a total of three case studies which allow for both inter-species and inter-region comparisons

4   ...to gain biological insights about the relationship between the environment and the distribution of top marine predators

5   ...to place the work of this PhD in a wider context and to recommend future directions of habitat preference and space use studies

1.4 Thesis structure

In chapter 2, I review statistical methods that have traditionally been used to analyse habitat preference and I discuss their limitations. I next show how the basic framework can be extended in stages to overcome its limitations. I then discuss eight problems that have an adverse effect on the framework's ability to estimate, infer and predict habitat preference and space use. These problems are 1) missing environmental data, 2) unequal accessibility of points in space, 3) unbalanced sampling effort across individuals, 4) multi-colinearity of environmental covariates, 5) spatial and 6) temporal autocorrelation in wildlife telemetry observations, 7) unequal detectability of animals in different habitats and 8) the animals' complex response to the their environment. I review existing, and, where necessary, propose alternative solutions to these problems.

In chapter 3, I apply the proposed statistical model framework to a telemetry data set from grey seals foraging off the east coast of Scotland between 1991

and 2001. Static environmental covariates (similar to those used in a qualitative study by McConnell et al. (1999)), such as sediment type and depth were used to investigate the grey seals' habitat preference. The model was fitted to data from the Farne Isle, but was also used to make spatial predictions of usage for Abertay, a different, albeit neighbouring haul-out site.

In chapter 4, I analyse data from adult Gannets making foraging trips from Bass Rock during the breeding season of 1998. As well as static environmental covariates, I have used dynamic variables such as satellite-derived weekly composites of sea surface temperature and chlorophyll concentrations as candidate covariates in the model.

In chapter 5, I investigate space use and habitat preference of grey seals foraging off the west coast of Scotland using data collected in 2002 and 2003. One key characteristic of this study region is that a large number of environmental variables are available, which have entered the model as candidate covariates. I have used this model, its parameter estimates and August counts of grey seals on land made throughout the region, to make predictions of absolute at-sea density.

Finally, in chapter 6, the general discussion, I have provided a summary of my model framework and illustrated its advantages and limitations. I take the opportunity to discuss the biological results of the three case studies and make comparisons across species (chapter 3 and 4) and regions (chapter 3 and 5). Finally, I discuss the wider context of my work and make suggestions for future research.

1.5 References

Arthur, S. M., Manly, B.F.J., McDonald, L.L. & Garner, G.W. 1996. Assessing habitat selection when availability changes. Ecology 77:215-227.

Blundell, G. M., J. A. K. Maier, and E. M. Debevec. 2001. Linear home ranges: Effects of smoothing, sample size, and autocorrelation on kernel estimates. Ecological Monographs 71:469-489.

Borchers, D. L., Buckland, S. T., Clarcke, E. D., and S. Ahmadi. 1997. Improving the daily egg production method using generalized additive models: Estimation of spawning stock biomass of mackerel (Scomber scombrus) and horse mackerel (Trachurus trachurus) in the NE Atlantic Ocean. Canadian journal of fisheries and aquatic sciences 54 (12): 2727-2742.

Boyce, M. S., and L. L. McDonald. 1999. Relating populations to habitats using resource selection functions. TREE 14:268-272.

Buckland, S.T., Anderson, D.R., Burnham, K.P. and J.L. Laake. 1993. Distance Sampling: Estimating Abundance of Biological Populations. Chapman and Hall, London.

Buckland, S. T. E., and D.A. Elston. 1993. Empirical models for the distribution of wildlife. Journal of Applied Ecology 30:478-495.

Fedak, M., Lovell, P., McConnell, B., and C. Hunter. 2002. Overcoming the constraints of long range radio telemetry from animals: Getting more useful data from smaller packages. Integrative and comparative biology 42 (1): 3-10.

Guisan, A., T. C. Edwards, and T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling 157:89-100.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135:147-186.

Hall, L. S., P. R. Krausman, and M. L. Morrison. 1997. The habitat concept and a plea for standard terminology. Wildlife Society Bulletin 25:173-182.

Johnson, D. H. 1980. The comparison of usage and availability measurements for evaluating resource preference. Ecology 6:65-71.

Kernohan, B. J., J. J. Millspaugh, J. A. Jenks, and D. E. Naugle. 1998. Use of an adaptive kernel home-range estimator in a GIS environment to calculate habitat use. Journal of Environmental Management 53:83-89.

Latimer, A. M., S. S. Wu, A. E. Gelfand, and J. A. Silander. 2006. Building statistical models to analyze species distributions. Ecological Applications 16:33-50.

Levin, S. A. 1992. The problem of pattern and scale in ecology. Ecology 73:1943-1967.

Manly, B. F. J., L. L. McDonald, and D. L. Thomas. 1993. Resource selection by animals: statistical design and analysis for field studies. Chapman and Hall, London

Matthiopoulos, J. 2003a. Model-supervised kernel smoothing for the estimation of spatial usage. Oikos 102:367-377.

Matthiopoulos, J. 2003b. The use of space by animals as a function of accessibility and preference. Ecological Modelling 159:239-268.

Matthiopoulos, J., B. McConnell, C. Duck, and M. Fedak. 2004. Using satellite telemetry and aerial counts to estimate space use by grey seals around the British Isles. Journal of Applied Ecology 41:476-491.

McConnell, B. J., M. A. Fedak, P. Lovell, and P. S. Hammond. 1999. Movements and foraging areas of grey seals in the North Sea. Journal of Applied Ecology 36:573-590.

Priede, I. G., and S. M. Swift. 1992. Wildlife telemetry; remote monitoring and tracking of animals. Ellis Horwood, New York.

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

Scott, J. M., Heglund, P. J., Morrison, M. L. Haufler, J. B., Raphael, M. G., Wall, W. A., and F. B. Samson. 2002. Predicting species occurrences; issues of accuracy and scale. Island Press, Washington.

Sjoberg, M., and J. P. Ball. 2000. Grey seal, Halichoerus grypus, habitat selection around haulout sites in the Baltic Sea: bathymetry or central-

place foraging? Canadian Journal of Zoology-Revue Canadienne De Zoologie 78:1661-1667.

Turchin, P. 1998. Quantitative analysis of movement; measuring and modeling population redistribution in animals and plants. Sinauer, Sunderland, Massachusetts.

White, G. C., and R. A. Garrott. 1990. Analysis of wilflife radio-tracking data. Academic Press, London.

Whittaker, R. H. L. S. A., and R. B. Root. 1973. Niche, habitat and ecotype. American Naturalist 107:321-338.

Wiens, J. A. 2002. Predicting species occurrences: progress, problems, and prospects.

Pages 739-749 in Scott, J.M., Heglund, J.M., Morrison, M.L., et al., editors.

Predicting Species Occurrences . Island Press, Washington DC.

Wright, P. J., H. Jensen, and I. Tuck. 2000. The influence of sediment type on the distribution of the lesser sandeel, Ammodytes marinus. Journal of Sea Research 44:243-256.

Appendix 1.A

VARIABLES AND PARAMETERS

---

u_i Number of telemetry observations or the number of used spatial-grid cells occurring in the ith habitat

a_i total area taken up by the ith habitat

o_i number of unoccupied spatial-grid cells in the ith habitat

c area of a cell in the spatial grid

w_i selectivity or preference index for the ith habitat

p_i relative availability of habitat i or probability of selecting a cell belonging to the ith habitat from all cells in the study area

r average number of empty cells selected for each used cell, or the number of absence points generated, on average, per telemetry location

h expected proportion of used cells in the case-control sample

n_i the total number of cells in the case-control data set that belong to the i'th habitat

$\mathbf{s}$ point in space corresponding to coordinates (x,y)

k_a the user-defined quantity of expected number of absences generated in space

k_u expected size of telemetry data set

$\upsilon(\mathbf{s})$ underlying distribution of space-use, such that $\int_{All\,\mathbf{x}} \upsilon(\mathbf{x})d\mathbf{x} = 1$

$\alpha(\mathbf{s})$ assumed distribution of accessibility, such that $\int_{All\,\mathbf{x}} \alpha(\mathbf{x})d\mathbf{x} = 1$

$\mathbf{A}$ the study region

A total area of study region $\mathbf{A}$ (km^2)

x explanatory variable

z random effect covariate

η linear predictor

$g(\cdot)$ Link function

β Fixed-effect model parameter

b Random-effect variable

ψ Variance-covariance matrix of the random effect

v Random-effects error term

$s(x)$ Smooth of an explanatory variable x

$h(\cdot)$ Individual basis function of a b-spline

i,j,l,m,k Subscripts used to index habitats, covariates, the case-control data point, individual and spline basis function, respectively.

$\hat{}$ observed value of a variable (e.g. $\hat{u}_i$)

$\tilde{}$ predicted value of a variable (e.g. $\tilde{u}_i$)

2. MODELLING HABITAT PREFERENCE

AND

SPACE-USE OF ANIMALS IN THE WILD

Experimental studies measure how a "dependent" variable responds to one or a few explanatory variables, while the values of all other, possibly influential, variables are kept constant. Data collection studies conducted in the wild have only limited control over environmental conditions. Analysis of such data must therefore model the effect of all variables simultaneously. This gives rise to many challenges, the exact nature of which depends on the objectives of the study, the method of data collection and the biology of the study species. The objective of this thesis is to find out how the spatial distribution of animals relates to the distribution of environmental variables. In this chapter, I develop a unified framework for modelling space-use and environmental preference (§2.1), and make appropriate adaptations to deal with the particular challenges associated with achieving this objective using telemetry data (§2.2).

Traditionally, this is done by classifying habitats on the basis of relevant environmental variables and testing for differences in use between those habitats while taking into account the unequal availability of habitats (§2.1.1). Alternatively, linear relationships between usage and those environmental variables that characterize habitats, can be estimated by means of linear regression models (for Gaussian response variables) and Generalized Linear Models (GLMs) (§2.1.2 and §2.1.3, respectively) with

categorical covariates. Such models have the additional advantage that they can make spatial predictions of usage. In §2.1.4, I show that by changing the study design (to a case-control) and the response variable (to binomial), reliance on an arbitrary habitat classification can be avoided. I fit a GLM with continuous covariates to such data. This type of model forms the foundation on which further extensions are built to deal with the difficulties that complicate the assessment of the true effect of environmental variables on the observed spatial distribution of animals. These result from the facts that raw environmental data don't usually cover all of space and time (§2.2.1), not all points in space are equally accessible to all animals (§2.2.2), sampling effort is not balanced across individuals or types of animals (§2.2.3) strong correlation between environmental variables makes it complicated to unravel their effect on the response (§2.2.4), spatial (§2.2.5) and temporal (§2.2.6) autocorrelation in animal movement causes data to be non-independent, particular habitats can directly or indirectly impede telemetry observation (§2.2.7) and the nonlinearity of the relationships between animal distribution and environmental variables makes the use of linear models inappropriate (§2.2.8)

2.1 The Statistical Analysis of Habitat Preference

2.1.1 Hypothesis testing

Given sufficient time for movement, the expected spatial distribution of a population of unconstrained random walkers within a region of space is approximately uniform. Although the random walk is rarely a realistic description of animal movement it is, nevertheless, a convenient null model against which hypotheses about individual movement can be tested (Turchin 1998, Morales et al. 2004). Consequently, the uniform distribution is a convenient null model for testing hypotheses about space-use and habitat preference (Manly et al. 1993). By definition, a species of random walkers show no habitat preference. Their uniform use of space means that they are expected to be observed in each habitat in proportion to its availability, the total area occupied by the habitat in geographical space. Deviations from

direct proportionality between usage and availability are interpreted as the tell-tale sign of habitat preference. The statistical significance of these deviations can be investigated using Chi-squared tests (Alldredge and Ratti 1986, Alldredge and Ratti 1992). Alternatively, the null model of proportional use can be employed to derive selectivity indexes (w_i) for the ith habitat such as the ratio between the expected number (u_i) of wildlife telemetry observations occurring in the ith habitat over the total area (a_i) taken up by that habitat (Manly et al. 1993),

$$w_i = \frac{u_i}{a_i} \qquad (a_i > 0) \qquad (2.1)$$

The ratio of observed usage over availability ($\hat{u}_i/a_i$) can be treated as an estimate of w_i and analyzed by means of classic parametric techniques such as ANalysis Of VAriance (ANOVA) (Aebischer et al. 1993, Fox 1997) .

This approach suffers from three drawbacks. Firstly, while the hypothesis tests may provide useful insights, ANOVA is based on a simplistic model (i.e. nominal covariates only) and it is therefore rarely used and almost never useful for prediction. Secondly, it assumes that w_i is normally distributed with constant variance across different habitats. These assumptions rarely hold. Thirdly, it relies on an a-priori classification of habitats which may be arbitrary with respect to the study-animals. These problems can be incrementally dealt with by building on the basic ANOVA framework.

2.1.2 Linear Regression with discrete covariates

The ANOVA framework can firstly be extended into linear regression with discrete covariates (Agresti 1996, Fox 1997). This model can incorporate both ordinal and nominal covariates and can be used to generate predictions about the response variable. The ratio $\hat{u}_i/a_i$ of observed usage per unit area for the ith habitat is modeled as a normal variable with mean w_i and

constant variance σ^2. The mean w_i is modeled as a linear combination of a total of n environmental conditions x_{ij} ($j = 1, \ldots, n$) prevailing in the ith habitat

$$
\begin{aligned}
\frac{\hat{u}_i}{a_i} &\sim \mathrm{N}(w_i, \sigma^2) && \text{for } a_i > 0 \\
w_i &= g^{-1}(\eta_i) = \eta_i \\
\eta_i &= \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_j x_{i,j}
\end{aligned}
\tag{2.2}
$$

The term η_i is known as the linear predictor. In general, the function $g(\cdot)$ is called the link function. In the case of linear regression the link, and therefore its inverse $g^{-1}(\cdot)$, is the identity function. Finally, x denotes a discrete or discretized environmental variable, x_{ij} denotes the value taken by the jth environmental variable in the ith habitat under the habitat classification scheme employed and β_j is the coefficient of the jth environmental variable. In discretized environmental variables, the values x_{ij} are conventionally (but not always, see Agresti 1996) the midpoints of the interval of discretization. Subject to the normality assumption in eq. 2.2 the coefficients β_j are equivalently estimated either by least squares or maximum likelihood. The fact that the coefficients β_j are subscripted by environmental variable and not by habitat, hints at the reason why eq. 2.2 is a predictive model: it attempts to describe a trend across environmental space and therefore provide estimates for the response variable in unobserved habitats.

Although this approach is capable of generating predictions, it still relies on an arbitrary classification and assumes normal errors with constant variance. These limitations can lead to poor overall predictions and misleading inferences about the significance of particular covariates. In the next section, I discuss how this approach can be further expanded by relaxing the normality assumptions associated with the model's response variable.

2.1.3. Generalized Linear Models with discrete covariates

The number of telemetry locations observed in a particular habitat is a count. Further, if, for the purposes of data collection, storage, or analysis, geographical space is represented by a grid then the number of grid-cells belonging to a particular habitat will also be a count. Therefore, the response variable in eq. 2.1 will always be non-negative and usually a rational number. Furthermore, this response will have a skewed distribution for low counts. This contradicts the requirement for a real-valued response variable with constant variance, an essential part of the normality assumption made by ANOVA and linear regression alike. The number of telemetry observations in the ith habitat is more naturally modeled as a heterogeneous Poisson process with rate $u_i = a_i w_i$

$$
\begin{aligned}
\hat{u}_i &\sim \text{Poisson}(a_i w_i) \\
w_i &= g^{-1}(\eta_i) = e^{\eta_i} \\
\eta_i &= \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_j x_{i,j}
\end{aligned}
\qquad (2.3)
$$

where η_i is the linear predictor, as in eq. 2.2. The expected number of telemetry observations in the ith habitat is also written as

$$
u_i = e^{\log(a_i) + \eta_i} \qquad (2.4)
$$

Eq. 2.3 is a Generalized Linear Model (GLM) which requires likelihood methodology for parameter estimation (Agresti 1996, Fox 1997). In the GLM terminology, eq. 2.3 and 2.4 describes a log-linear model and the term $\log(a_i)$ is known as the model's offset.

This approach still requires a habitat classification. This is often constructed arbitrarily, prior to modeling and can severely prejudice the results of the analysis: If the classification is too fine, then the number of observations associated with each habitat may be too small to detect any differences between them. Conversely, making the classification too coarse merges habitats that appear similar to a human observer but may be

perceived differently by the animals. The trade-offs associated with the resolution of the habitat classification become more acute in studies with many environmental variables because the data become more thinly spread over a higher-dimensional environmental space.

Furthermore, the Poisson error distribution in eq. 2.3 may be unrealistic either due to a superabundance of zeros (e.g. resulting from a very fine habitat classification) or disproportionately high counts (e.g. resulting from unexplained aggregations of telemetry observations). This can be remedied by using a zero-inflated or otherwise over-dispersed Poisson error distribution (Lambert 1992, Welsh et al. 1996, Fox 1997) or by abandoning classification in geographical and environmental space, as described in the following section.

2.1.4 Generalized Linear Models with continuous covariates

In Appendix 2.A I explain how the discretization of geographical and environmental space can be abandoned in favor of a case-control design. This produces a binomial response variable (u_l) which takes the values 1 for l^{th} case-control data point if it belongs to the telemetry data, and 0 for a control data point. This response can be modeled as a Bernoulli process with probability of success h_l

$$\hat{u}_l \sim \mathrm{B}(1, h_l)$$
$$h_l = g^{-1}(\eta_l) = \frac{e^{\eta_l}}{1 + e^{\eta_l}} \tag{2.5}$$
$$\eta_l = \beta_0 + \beta_1 x_{l,1} + \cdots + \beta_j x_{l,j}$$

There are two points of contention concerning the response variable in eq. 2.5 when it is fitted to case-control data. First, because the number of absences used to fit the model is determined arbitrarily, the case-control approach can quantify the relative importance of different covariates but not the absolute abundance of animals. Second, it has been suggested (Manly et al. 1993, Boyce and McDonald 1999) that in logistic models h_l is

proportional to space-use. Keating and Cherry (2004) have argued that this is not the case. In Appendix 2.A, I show that, for a large number of controls, h_l is, in fact, proportional to preference. An approximate relationship between the response variable and usage can be derived as follows: I assume that the telemetry observations ($u_l = 1$) are generated from an heterogeneous, spatial Poisson process whose rate is proportional to the unknown, spatial probability density function $v(\mathbf{s})$, where $\mathbf{s}$ is a position in geographical space (Fig. 2.1a). Similarly, the control observations ($u_l = 0$) are generated from a user-defined spatial Poisson process with a rate $\alpha(\mathbf{s})$ (Fig. 2.1b). In the simplest case, where all points within a study area are assumed to be equally accessible, $\alpha(\mathbf{s}) = A^{-1}$, where A is the total area of the study region. More complex spatial density functions can be used when such an approximation is inappropriate (e.g. for central-place foragers).

Indefinitely increasing the resolution of both spatial and environmental grid (operations O1 and O2, respectively in Appendix 2.A), means that the expected number of telemetry observations in the ith habitat u_i can be approximated by $k_u v(s)c$, where c is the area of a grid cell in geographic space and the proportionality constant k_v depends on the sampling intensity (number of animals tagged and frequency of telemetry locations) and can be thought of as the expected size of the telemetry data set. Similarly, the expected number of controls in the ith habitat $k_a p_i$ (see Appendix 2.A), can be approximated by $k_a \alpha(s)c$, where k_α is the user-defined quantity of the total, expected number of absences generated in space. Therefore it follows that after O1 and O2 the response variable $h(X_\mathbf{s})$ in eq. 2.5 as defined in A3, tends to

$$h(X_\mathbf{s}) = \frac{k_v v(\mathbf{s})}{k_\alpha \alpha(\mathbf{s}) + k_v v(\mathbf{s})} \tag{2.6}$$

This probability $h(X_\mathbf{s})$ depends on the values for the environmental covariates $X = (x_1, \cdots, x_j, \cdots)$ observed at spatial location $\mathbf{s}$ (Fig. 2.1c). These

environmental conditions vary spatially (Fig. 2.1d) and so does $h(X_s)$ (Fig. 2.1e). Finally, and most importantly, I can use the estimated dependence between h and X, to estimate spatial usage at location **s**. Equation 6. can be re-arranges as

$$v(\mathbf{s}) = \frac{h(X_s)}{1 - h(X_s)} r\alpha(\mathbf{s}) \tag{2.7}$$

Where, $r = k_\alpha / k_v$ is the number of controls selected per telemetry observation.

This is a local estimate of usage. Since it is only based on a random sample of controls, the cumulative of estimated usage over all space will usually deviate from unity and will need to be normalized. Experimentation with test data sets has indicated that this does not bias the estimates of usage. This would still need to be proven theoretically as part of future work. The normalized map of usage can be used to calculate the proportion of animals within a specified region (e.g. a Special Area of Conservation (SAC)), by integrating $v(\mathbf{s})$ over that region. If the total number of animals in the region is also known from other sources, relative usage can be scaled up to create a map of absolute population density.

The case-control design also raises the issue of false absences (Boyce et al. 2002) which is an important problem for small-scale studies using spatial grids. However, the larger spatial scales covered by telemetry data, allow us to consider individual animals as practically dimensionless. This implies that the probability of encountering an animal at the exact coordinates of any arbitrarily chosen point in space is zero.

GLMs with continuous covariates accept non-normal response data, do not rely on arbitrary habitat classifications, and can make spatial predictions. Thus, they address all three of the problems historically associated with hypothesis testing in studies of habitat preference. I therefore use them as the foundation of our model framework.
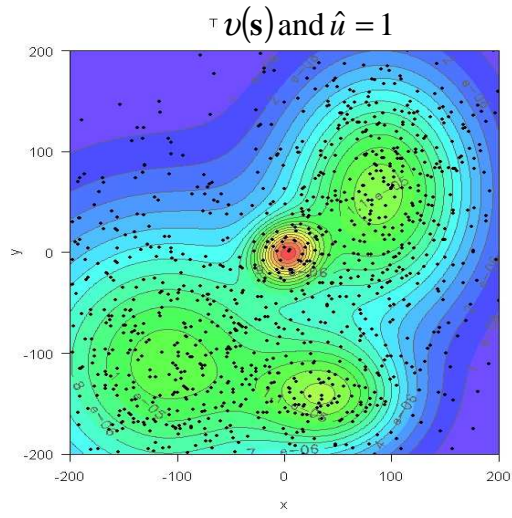
a.
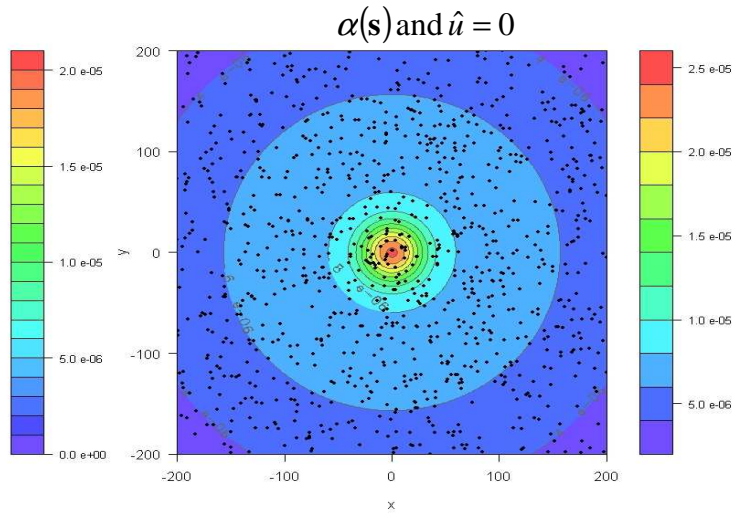
b.
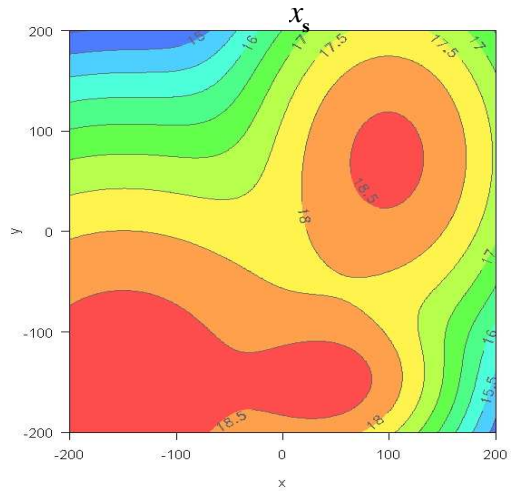
c.
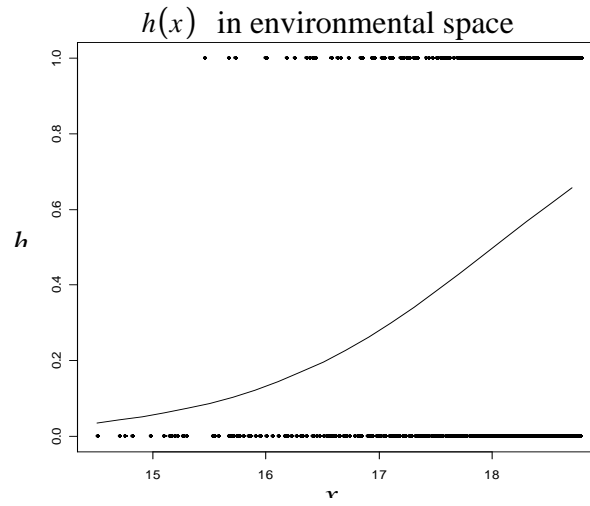
d.
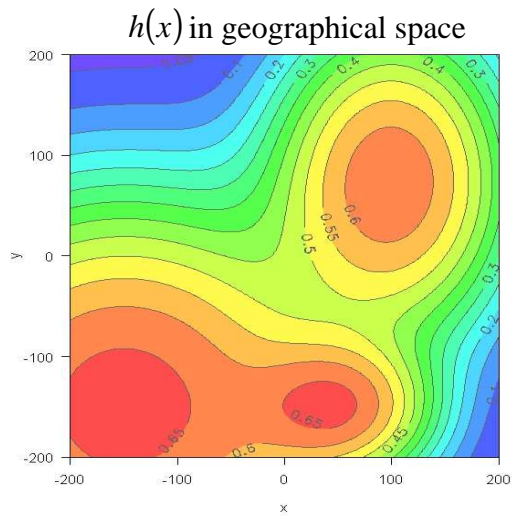
e.
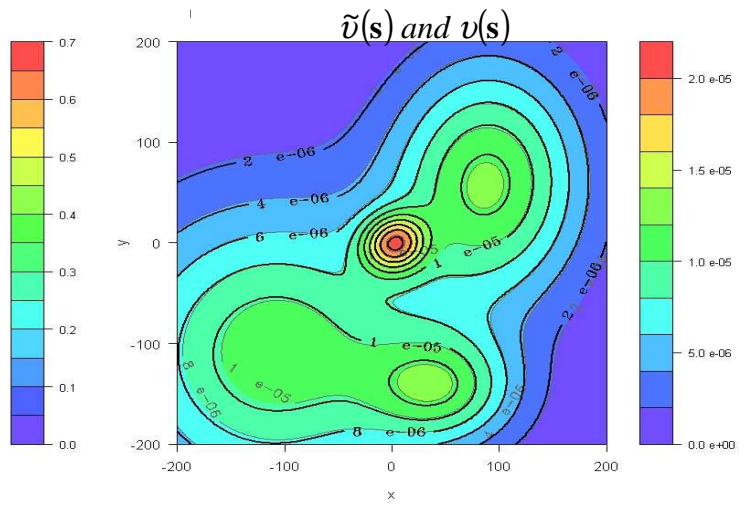
f.

Fig. 2.1 The thesis' fundamental concepts illustrated with the aid of an artificial example. The underlying, unknown surface of space-use $v(\mathbf{s})$, described by the contours in (**a**), is sampled to generate the telemetry observations ($\hat{u} = 1$), shown as black dots. Pseudo-absence data ($\hat{u} = 0$) can be generated according to a spatial rate $\alpha(\mathbf{s})$. In (**b**), I have modeled the accessibility of different points in space from the simulated animals' central place, to generate this rate. The distribution of usage in (**a**) is assumed to result from the animals' response to a single, heterogeneously distributed environmental covariate (**c**). This unknown response is the expectation modeled by the response h in one-dimensional environmental space (**d**). Predictions of the response variable in geographical space (**e**) give a visual representation of preference . The use of simulated data in this example, allows us to perform a spatial comparison between predicted usage $\tilde{v}(\mathbf{s})$ (filled contours) and true, unobserved usage $v(\mathbf{s})$ (bold contour lines) (**f**).

2.2 Modeling habitat preference: Problems and solutions

The quality of a statistical model is determined by its accuracy, precision and parsimony (Buckland et al. 1997, Burnham and Anderson 2002). Specific properties of the environment, the characteristics of the study animals and the ways that both are sampled, can cause one or more of these characteristics to suffer. Loss of accuracy implies biased parameter estimates and predictions, and loss of precision leads to increased uncertainty in parameters and predictions. Loss in parsimony leads to over-parameterized models capable of predicting a particular data set well, but liable to predict new data poorly. Below, I discuss eight problems that can cause such detrimental effects and propose appropriate solutions.

2.2.1 Environmental data rarely coincide with usage data

In trying to relate wildlife telemetry data to environmental variables I implicitly assume that the study animals experience the same conditions described by our environmental data sets and respond to them by being

present or absent from a particular location at a given time. If there are no delays in the animals' response to local environmental conditions, this assumption requires that environmental data are available for those places and instances. However, environmental data are usually collected independently of wildlife telemetry data. Further, unwanted interference in environmental data collection (e.g. cloud cover obscuring remote sensing) and logistical constraints (e.g. limited observation time in transect surveys) mean that spatial coverage is rarely complete. Consequently, it is rarely known exactly what conditions the animals are responding to, at any given point in space and time. This reduces model precision and its effect becomes more acute with highly heterogeneous and dynamic environmental variables (Isaaks and Srivastava 1990).

To address the lack of environmental data, some modern tags also collect data about the animals' environment (Fedak 2004, Cooke et al. 2004) but such valuable technological improvements tell us nothing about conditions at points not visited by the tagged animals. This information is just as important for the analysis of preference.

If there is temporal replication in the data collection for highly dynamic environmental variables (e.g. meteorological variables) it may be possible, to interpolate prevailing conditions at the time the wildlife telemetry data were being collected. However, modeling space-use and preference under changing environmental conditions still presents big challenges (Arthur 1996, Hjermann 2000, and discussions in Boyce et al. 2002). It is possible to avoid these difficulties by using only environmental variables whose spatial distribution remains constant over time (e.g. altitude or sea bottom depth) or, at least, stationary during the temporal scope of the study (e.g. January snow cover, over a decade). This still leaves the problem of incomplete spatial coverage.

If the measurement errors produced by the environmental survey method are negligible and the environmental variable is time-invariant (as is often the case with geophysical variables) then spatial interpolation methods (Ripley 1981, Kafadar and Horn 2002) can be used to estimate the values of the variable in-between the survey locations. The interpolated surface is constrained to pass through the observations at the survey

locations. However, if, the environmental data are affected by measurement error (e.g. data on prey density) or process stochasticity (e.g. meteorological variables), smoothing techniques are a more appropriate estimation method (Ripley 1981, Silverman 1986). Approaches such as kernel-smoothing (Silverman 1986), combined with cross-validation for the selection of the smoothing coefficient, attempt to reproduce the mean and underlying distribution of the stationary process that generated the survey data. Both interpolation and smoothing use the spatial autocorrelation in the environmental data (see §2.2.5 below) and can be seen as the two extremes of the methodological spectrum for density estimation. A third estimation technique, kriging (Isaaks and Srivastava 1990), first models spatial autocorrelation in the form of the variogram (the covariance between the values at sampling locations as a function of their distance) which it then uses for estimation. The intercept of the variogram, also known as the nugget, expresses the amount of stochasticity and measurement error in the data and can either be set by the user or estimated from the data – particularly if replicate measurements exist for the same survey locations. An appealing aspect of kriging is that it can behave as a spatial interpolator as well as a smoother depending on the variogram intercept. A shortcoming that ordinary kriging shares with many smoothing techniques is that it assumes the extent of spatial autocorrelation to be constant throughout the range of the data, i.e. the spatial process is assumed to be stationary. This can smooth out strong local gradients in certain regions of geographical space, as a result of weaker gradients elsewhere.

2.2.2 Points in space are not equally accessible to the animals

The precise definition of availability is important when modeling preference (eq. 2.1). The most obvious measure of availability is the total area taken up by a particular habitat. This either implies that animals have equal access to all points in geographical space (perfect mixing) or that, within their range of sensory perception, they experience a sample of habitats that is representative of the habitat composition of entire space (representative perception).

The assumption of perfect mixing can only ever be approximately true and, in certain cases, it will be so seriously violated as to prejudice the analysis of preference. For example, in the case of central place foragers, accessibility of points in space decays with distance from the central place. At the population level, this is particularly evident in colonial animals (Matthiopoulos et al. 2004). The assumption of representative perception will be violated if the animals move little and the environmental conditions in neighboring sites are more similar than in sites further apart.

Most studies of preference take a pragmatic approach by focusing on arbitrarily-defined regions of geographical space (Manly et al. 1993). In some cases (e.g. lake-dwelling fish), it is easy to define such regions, but generally it is not. This issue was first addressed by Johnson (1980) who identified accessible space with an animal's home range. This suffers from three problems: First, the usage threshold involved in the definition of home ranges is often arbitrary (Aebischer et al. 1993), second, not all points in the home range are equally accessible from its center and third, certain points may be outside the home range because, although they are accessible, they are not preferred.

Other researchers have taken a more mechanistic approach by calculating the accessibility of points in space from the animals' starting position, speed and mode of movement, travel duration and travel medium/obstacles. Such models have been developed for both nomadic (Arthur 1996, Hjermann 2000) and central-place foragers (Matthiopoulos 2003b), and although they vary in complexity (from simple diffusion models to complex individual-based simulations), they can all be parameterized from readily available, independent data. Their output is a spatial surface, which represents the likelihood of observing an animal at a given point, in the absence of habitat preference. It can be treated as a probability density function, from which random points can be sampled to construct the set of absence points in the case-control binomial model of eq. 2.5. In geographical space, the response variable can be interpreted as the probability that an accessible point is visited by the animals. Equivalently, in environmental space, the response variable is interpreted as the probability that a habitat that is available to the study animals is used by that animal. Therefore,

selecting the absence data from an accurate accessibility surface keeps the response variable (eq. 2.5) in agreement with the definition of preference (eq. 2.1).

Certainly, accessibility surfaces will never be perfect and discrepancies between the response variable and true preference will cause the model to over- or under-predict. To absorb these residuals some measure of accessibility (e.g. distance from the central place) can be incorporated into the model as a candidate covariate.

2.2.3 Sampling effort in telemetry studies is rarely balanced across individuals

With the exception of studies focusing on rare or threatened species, where the focus is on the particular animals carrying the telemetry tags, most habitat preference studies aim to make inferences about the behavior of the entire population to which the tagged animals belong. All population-level inferences are subject to sampling uncertainty. In telemetry studies sampling uncertainty is usually large because, due to logistical constraints, the ratio of tagged to un-tagged animals is generally small and because sampling effort between tagged individuals is almost never balanced. Hence, different tags will provide us with different numbers of observations simply because tag-life is a stochastic variable, or because the behavior of individual animals may, for some reason, facilitate or impede information transmission. Also, capture and tagging effort may not be spatially uniform and might not sample animals of different ages and genders representatively. Therefore, estimating habitat preference by pooling telemetry data from all individuals is likely to bias the results towards certain data-rich individuals, types of individuals or regions of geographical space. The alternative, is to recognize the natural hierarchy of sampling units (Fig. 2.2) and to use an error structure that more accurately reflects the variability within and between different levels of this hierarchy (Gillies et al. 2006). Multi-level or Mixed-effects models (Pinheiro and Bates 2000, Fox 2002), can simultaneously model the behavior of the average individual using the population mean (fixed-effect) and the variability in the behavior

across individuals using random effects. Eq. 2.5 can be modified into a mixed-effects model as follows:

$$\hat{u}_l \sim \mathrm{B}(1, h_l)$$
$$h_l = g^{-1}(\eta_l) = \frac{e^{\eta_l}}{1 + e^{\eta_l}}$$
$$\eta_l = b_{0,m} + b_{1,m}x_{l,1} + \cdots + b_{j,m}x_{l,j} + \cdots$$

(2.8)

The coefficients b are, themselves, random variables that can either be specified with a mean and variance, or modeled as functions of class-member characteristics. For example, to capture the individual variation within a group of animals, the class must be defined as the group of individuals. Class-member characteristics appropriate for that class might be an individual's age, sex or mass. Hence, the coefficient $b_{j,m}$ that quantifies how the mth individual responds to the jth environmental variable can be given as a linear combination of individual-specific characteristics ($z_{k,m}$) using coefficients (β_j) that refer to the entire group

$$b_{j,m} = \beta_{j,0} + \beta_{j,1}z_{1,m} + \ldots + \beta_{j,k}z_{k,m} + \ldots + v_j$$

(2.9)

Combining eq. 2.8 and 2.9 reveals that the inclusion of individual-specific characteristics ($z_{k,m}$) enter the linear predictor as interactions with the environmental covariate ($x_{l,j}$)

$$\eta_l = b_{0,m} + b_{1,m}x_{l,1} + \cdots + \left(\beta_{j,0}x_{l,j} + \beta_{j,1}z_{1,m}x_{l,j} + \ldots + \beta_{j,k}z_{k,m}x_{l,j} + \ldots + vx_{l,j} \right) \cdots$$

(2.10)

The coefficients β are also known as the fixed effects. The random effect accounts for the within-class, in our application, between-individual variability. They are denoted as $v_0, v_1, \ldots, v_j, \ldots$ and are commonly assumed to have a joint multivariate normal distribution with mean zero and a

variance-covariance matrix Ψ, representing within-class variability (Pinheiro and Bates 2000). The estimation procedure for mixed-effects model returns values for the fixed effects and estimates of Ψ for the distribution of the random effects.

a) Population

b) Sub-population/Group

c) Individual

d) Trip

e) Telemetry observation
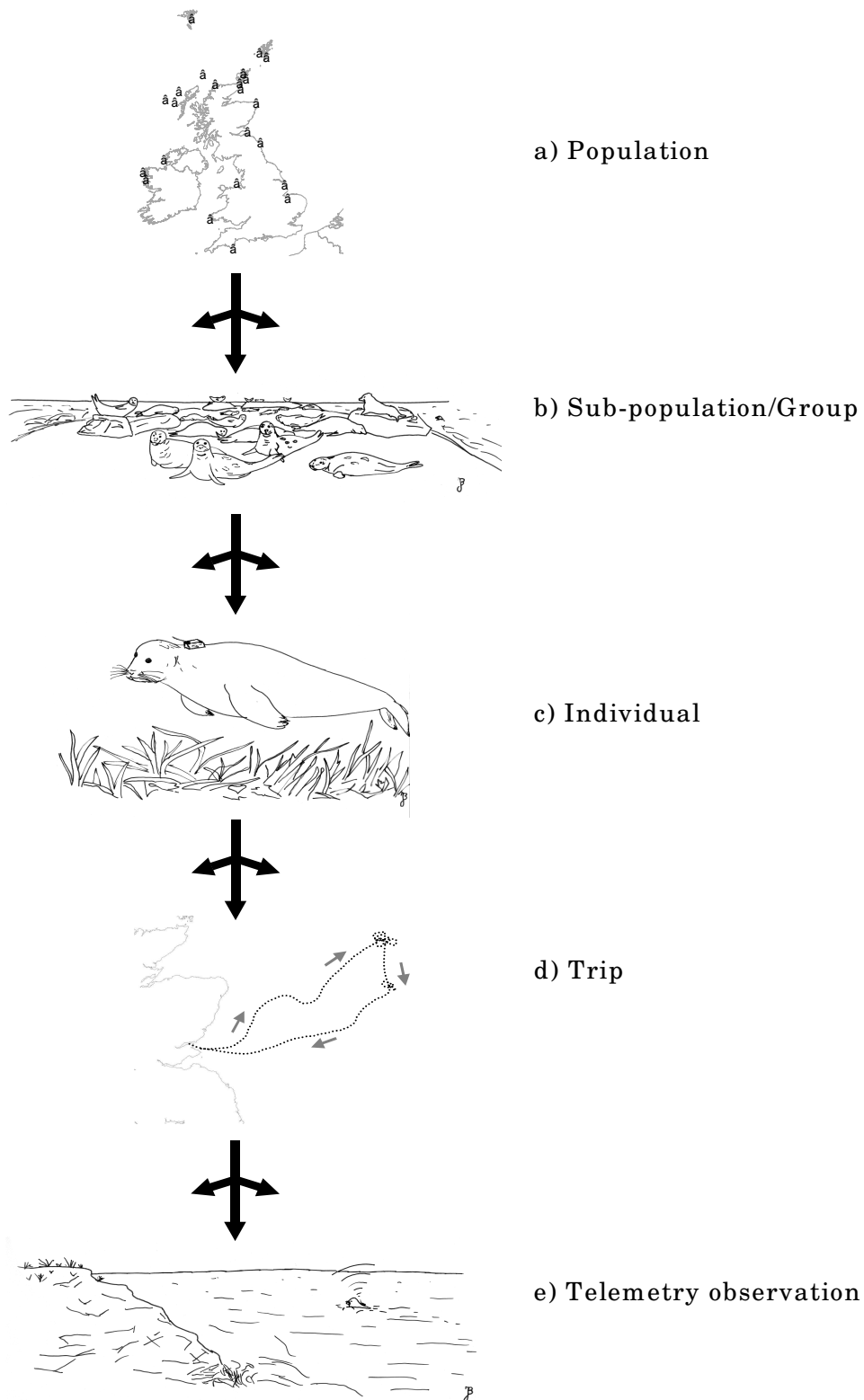
Fig. 2.2 Illustration of the hierarchy of sampling units in a telemetry data set. The population of animals (**a**) can be subdivided into many sub-populations (**b**) each of which contains several individuals (**c**). Every individual makes foraging trips (**d**) to sea during which telemetry observations (**e**) are collected.

In some cases, the responses of individuals to environmental variables are distributed non-normally around the population mean response. Specifically, a few individuals may have extreme responses (implying a non-normal kurtosis for the random effect) or their responses may be asymmetric around the average population response (giving rise to a skewed distribution for the random effect). These two deviations from normality cause imprecision in the estimates of variance for the random effects and biases in the estimates of the fixed effects, respectively. If these deviations are a consistent result of particular characteristics of the individual, then the properties of the random effects distribution will be directly attributable to the relative frequency of different types of animals in the study sample. If these individual characteristics can be identified, they can be included in the model so as to explicitly account for individual variation and yield normally distributed random effects. However, if they are unknown, then a more appropriate random-effects distribution may be required.

2.2.4 Some environmental variables may be correlated

Certain large-scale processes (e.g. meteorological or geological) may influence most of the environmental variables that might be used to explain the spatial distribution of animals. Furthermore, interactions between environmental variables are often just as strong as the links between them and usage. Both of these mechanisms may lead to candidate covariates of usage that are strongly correlated with one another. In the simplest case of colinearity, strong correlation occurs between two variables. In the more general case, known as multi-colinearity, a strong correlation occurs between one variable and a linear combination of other variables (Cramer 1985).

Using n environmental variables requires fitting a model in n-dimensional environmental space. However, in the presence of strong correlations between environmental variables there is often insufficient information in the data to support such a model. An intuitive illustration of this is provided by bivariate, linear regression, where a planar response is fitted in two-dimensional environmental space. Strong correlation between the two environmental variables means that the fitted plane balances unstably on data arranged approximately along a line (Fox 1997), which results in unstable parameter estimates. It also leads to parameter estimates that have large standard errors and are even more sensitive to outliers. This corresponds to a loss in both accuracy and precision and is a general consequence of colinearity in multivariable models. Furthermore, since colinearity indicates lack of support for a high-dimensional model, it results in a loss in parsimony. Colinearity is a problem for the estimation of individual parameters and, consequently, for drawing inferences about the relative importance of individual environmental variables on usage. If the objective of the modeling study is focused on getting the best fitting model of a particular dataset, there is an argument for ignoring colinearity. However, if biological interpretation of the results (based on the parameter estimates and associated variances) and predictions elsewhere are of the essence, then colinearity needs to be detected and treated.

The most evident relationships in the explanatory data are usually seen when inspecting pair-wise correlations, but this ignores multi-colinearities. An alternative is to use Variance Inflation Factors (VIFs) (Fox 1997), given in terms of R_j^2, the un-adjusted Pearson correlation coefficient, obtained when the jth explanatory variable is modeled as a linear function of all other explanatory variables.

$$VIF\left(\hat{\beta}_j\right) = \frac{1}{1 - R^2{}_j} \tag{2.11}$$

Generally, VIFs greater than 6 indicate strong multi-colinearity and variance estimates of the affected parameters need to be adjusted.

Traditional treatments of colinearity involve either dropping as many environmental variables as required to get to a lower-dimensional environmental space, or transforming the variables so that they are uncorrelated. Dropping environmental variables can be done automatically by means of model-selection criteria such as the change in deviance, approximate F-tests, or ICs (McCullagh and Nelder 1989, Hastie and Tibshirani 1990, Augustin et al. 1996). Automatic model-selection may be augmented by models based on auxiliary biological knowledge on causal relationships between variables. Treating colinearity by transformation of the candidate covariates can be achieved with techniques such as principal components analysis (Jolliffe 1990). In practice, this also leads to a lower-dimensional environmental space because the last few principal components usually contribute little to the model. A disadvantage of this technique is that relationships between response and the principal components of several environmental variables are difficult to interpret biologically.

More recent approaches, using simple GLMs, have instead sought to treat the consequences rather than the causes of multi-colinearity (Fox 1997). It is possible to use VIFs to correct for the effect of multi-colinearity on the standard errors of model parameters, as follows,

$$\mathrm{var}(\beta_j) = \frac{\sigma_j^2}{(n-1)s_j^2} VIF(\hat{\beta}_j) \qquad (2.12)$$

where σ_j^2 is the estimated variance for the model parameter corresponding to the j th explanatory variable, s_j^2 is the sample variance estimated from the n observations on that variable. This is a very promising approach and I look forward to its further development for use in mixed-effect models.

2.2.5 Species distributions are spatially autocorrelated

Positive spatial autocorrelation is a typical characteristic of animal distributions. It leads to nearby points having more similar values of usage than would be expected by chance. In telemetry data, this is manifested as

clusters of observations in space. In itself, this is not problematic. In fact, this interdependence between points in space is usefully employed by all usage estimation techniques (such as interpolation, smoothing and kriging, discussed in §2.2.1, above) (Blundell et al. 2001, Matthiopoulos 2003a). In habitat preference studies, it is hoped that autocorrelation in usage is a result of autocorrelation in the available covariates and therefore that it will be captured by the model (Diniz et al. 2003). However, sometimes the best-fitting model presents residual autocorrelation: It systematically over/under-estimates usage in entire regions of geographical space.

Residual autocorrelation violates the central assumption of independence in the parameter's standard errors. If spatial autocorrelation is positive this leads to underestimates of the standard errors for the parameters (a loss in model precision) and a more likely inclusion of irrelevant environmental variables (i.e. a loss in model parsimony).

To deal with the problem it is first necessary to detect it, using two well-established statistics, Moran's I and Geary's C (Cliff & Ord, 1973), or by constructing spatial variograms of the model residuals (Isaaks & Srivastata 1990). If there is no significant residual spatial autocorrelation, then no action needs to be taken because, even when usage itself is known to be spatially autocorrelated, this is entirely accounted for by the model's covariates. On the other hand, if residual spatial autocorrelation is detected, it could be due to either extrinsic (e.g. autocorrelated environment) or intrinsic (e.g. conspecific attraction, dispersal limitations) factors (Legendre 1993, Keitt et al. 2002, Overmars et al. 2003).

An intrinsic form of spatial autocorrelation means that the value of the response at a point in space is a direct consequence of its values at neighboring points. If intrinsic causes are suspected, a natural choice is to use autoregressive linear models (also called spatial lag models (Anselin 2002)). For each point in space, the model's linear predictor is augmented with an auto-covariate that is derived as a weighted function of values of the response variable from the neighborhood of that point. The weights that specify the relative contribution of each neighboring point, can be obtained from different functions of distance, such as the exponential, Gaussian or

inverse (see derivation in Cliff and Ord 1973 and application in Keitt et al. 2002 and Lichstein et al. 2002).

Extrinsic factors give rise to residual autocorrelation for one of two reasons: Either an important, autocorrelated covariate has been omitted from the analysis, or the model has been mis-specified (Cliff and Ord 1973). So, when extrinsic causes are suspected, the first step is to introduce new covariates or to ensure that the model is sufficiently flexible by including non-linear terms or interactions between existing covariates. If residual spatial autocorrelation persists and no additional environmental variables are available, this may warrant the use of conditional or simultaneous autoregressive models (CAR and SAR, respectively - Keitt et al. 2002), also known as models with spatially filtered variables. They are similar to autoregressive models, except that the spatial covariate is a function of neighboring model residuals (i.e. the difference between the observed response values and those predicted by the model). An alternative to CAR and SAR, are geostatistical models such as co-kriging (Stein and Corsten 1991). Geostatistical Models account for spatial pattern by modeling the correlation between the errors as a function of distance (Keitt et al. 2002).

However, these only apply if the response variable is normally distributed. (Cliff and Ord 1973, Keitt et al. 2002). Very recently, autoregressive linear models have been extended to autoregressive GLMs, such as the auto-logistic (Augustin 1996) or the auto-Poisson (Huffer & Wu 1998) and geostatistical models have been extended by including auto-correlated random effects within GLMMs (Diggle et al. 1998, Stephenson et al. 2006). However, parameter estimation of such models by traditional likelihood methods is difficult due to the high-dimensional numerical integration it requires. This is especially true for telemetry studies where the number of data points is generally large. Therefore, parameter estimation needs to rely on other methods such as generalized estimating equations (GEE - see Hanley et al. (2003)), penalized quasi-likelihood methods (Dean et al. 2004) or Bayesian estimation using MCMC (Thomas et al. 2004). The choice of spatial model and structure of spatial lag operator will also lead to different spatial predictions and different conclusions about

which environmental variables are important (Griffith 2005, and see Keitt et al. 2002 for normal response models).

The main advantage of successfully modeling spatial autocorrelation is that standard information criteria (e.g. AIC) can, once again, be used for model selection. However, although the use of these spatial models will become increasingly important in future studies, the current lack of guidelines for model structure and estimation software makes them an impractical proposition. A more practical method is to use a conservative model selection protocol involving a higher penalty in the information criteria, or to implement model selection by re-sampling (e.g. bootstrapped *p*-values, cross-validation).

Using information criteria to systematically search the entire set of possible covariate combinations for a good model is a scientifically vulnerable practice, referred to as "data dredging" (Burnham and Anderson 2002). Indeed, a large set of candidate models increases the probability that a model accidentally finds pattern in stochasticity and over-fits the data. It has therefore been argued (Burnham and Anderson 2002) that model selection should be carried out from within a small set of models (<20) that are directly interpretable as biological hypotheses (Burnham & Anderson 2002).

Although this is sound advice for studies where a small number of hypotheses exist, it does not help with exploratory analyses of habitat preference where it may be better to directly address the problem of over-fitting. This can be done by cross-validating the predictions of a model with a subset of the data not used for fitting. The choice of how to subdivide the data for cross-validation depends on the objectives of the study: If the aim is to construct the best descriptive model for a population of animals, the data should first be organized by animal and the animals should then be split randomly into two groups. If the objective is to construct the best predictive model that also applies elsewhere or at a different time, then the available sample of animals should be disaggregated by space or time. It should be noted that cross-validation and the use of models that explicitly model residual spatial autocorrelation are not mutually exclusive. In fact, I believe

that their combination, in the future, will make for a very powerful modeling approach.

2.2.6 Telemetry locations from the same individual are serially correlated

The term "serial autocorrelation" refers to the dependence between two observations made at instants t and $t + \Delta t$ along a time series (Swihart & Slade 1985). The degree of autocorrelation in telemetry data depends on how fast the study animal moves and how often its movement is sampled. If, additionally, the environmental variables are spatially autocorrelated, then the closer two telemetry observations are in time, the more likely they are to occur at similar environmental conditions. This similarity between contemporaneous locations will increase as the scale of spatial autocorrelation increases (Fig. 2.3). Biologically, this may give the impression of preference for these conditions when, in fact, they are due to slow movement and frequent sampling in a spatially autocorrelated environment. Statistically, the presence of positive serial autocorrelation, will lead to underestimation of standard errors for model parameters (loss in precision) and some irrelevant environmental variables being retained during model selection (loss in parsimony).
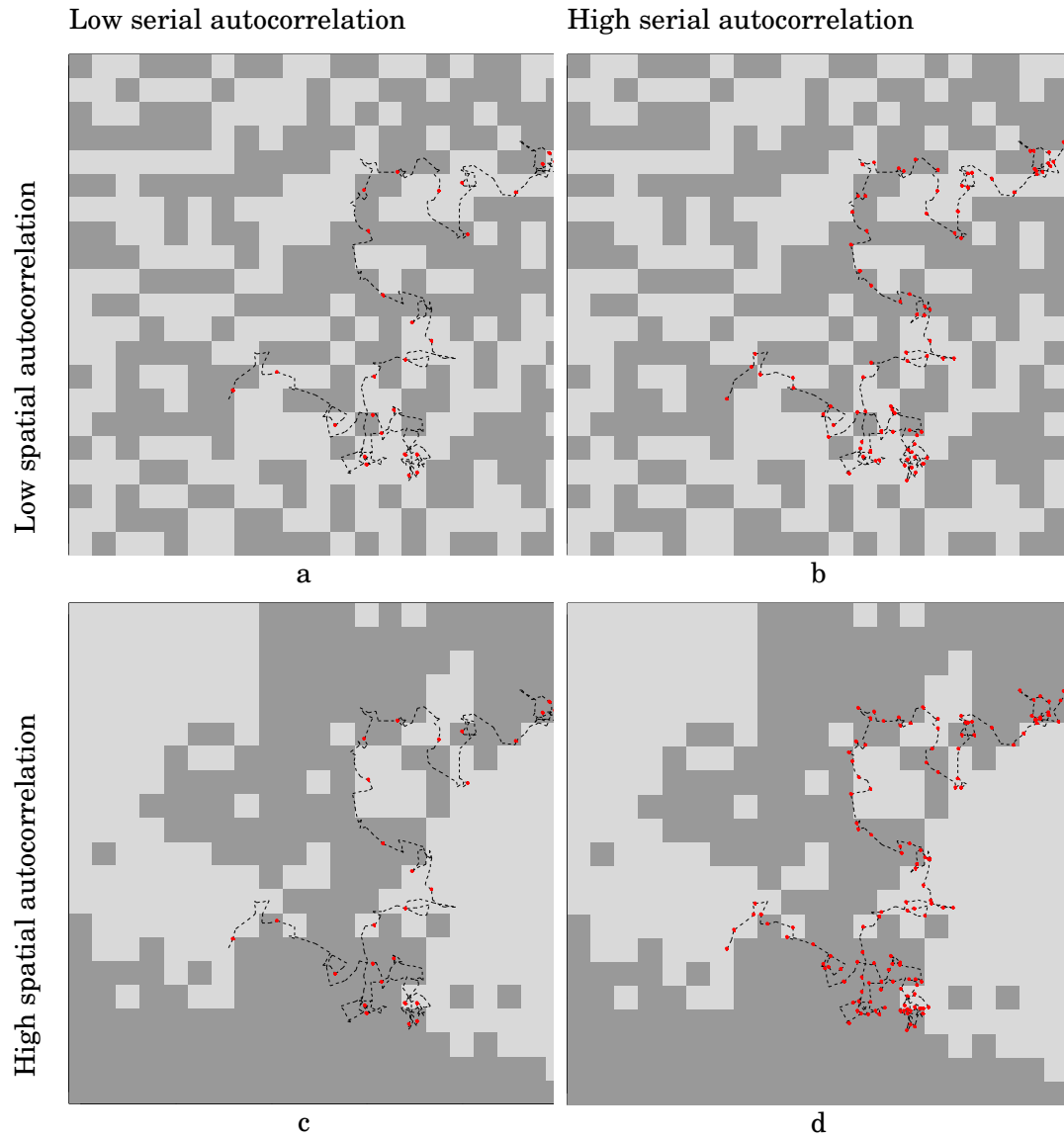
Fig. 2.3 The extent of inter-dependence in the data used for models of habitat preference is determined by how fast the animals move and how frequently the observations are collected (serial correlation) but, also, by the degree of similarity in conditions between neighboring points in space. High serial autocorrelation (right-hand column) means that successive observations will be too close to be considered independent and high spatial autocorrelation (bottom row) implies that the conditions facing the animals will tend to be similar between successive observations. The assumption of independence is most severely violated when usage data obtained by frequent telemetry sampling are regressed against strongly autocorrelated environmental variables (part **d**).

The most direct solution is to remove the spatial dependency among observations either by filtering out the spatial structure or by removing observations based on the so-called "time to independence" (Swihart & Slade 1985). However, this has three drawbacks: First, it treats independence between two points as either present or absent. In reality, there are degrees of independence and this is particularly true for observations of animal movement. Even if inter-dependence between observations decays monotonically with time, no-two points in an animal's path can ever be assumed to be entirely independent. By the same token, irrespective of how close in time two observations are, one can never be exactly predicted from the other. Therefore, censoring the data leads to a data-set that is not completely free of autocorrelation and poorer in information since the highly autocorrelated points that were removed contained some useful information (e.g. Rooney et al. 1998). The second problem is that the time to independence calculated from a telemetry data set depends non-trivially on the overall time of observation and the geometry of the animal's path (Solow 1989). Finally, censoring relies on hypothesis-testing with an arbitrary critical value which makes it sensitive to the underlying distribution of telemetry locations (Solow 1989) and therefore unreliable without previous power analysis (Swihart and Slade 1985). An alternative is to use permutation tests (Legendre et al. 1990) or to correct the degrees of freedom used in estimating the standard errors of the model's parameters (Clifford et al. 1989).

Aebischer et al. (1993), point out that the objective of most ecological studies is to draw inferences about the population, and therefore that biological hypotheses must be tested at the level of the individual animal rather than the telemetry observation. The problem of temporal autocorrelation is therefore thought to be circumvented by using the animal as the sampling unit (Aebischer et al. 1993, Otis & White 1999). For example, a GLM, could be fitted to data from each animal separately and parameter estimates (the β_j's, in eq. 2.5) from all individuals pooled into a sample leading to a mean and associated individual variation for each parameter. These point- and interval-estimates could be used to make population-level inferences about the significance of specific terms in the

model. In this example, significant deviations of the individual-specific parameters (β_j's) from the population mean would occur either because individuals are truly different or because they appear to be different due to the natural stochasticity in the data. The degree of stochasticity depends critically on the number of observations obtained for each animal. As a consequence, the between-individual variability may be over-estimated and this approach may be too conservative.

Mixed-effect models, can distinguish between these two cases by explicitly modeling individual variation and stochasticity as the random effects and the variance of the fixed effects, respectively. Results based on simulated data indicate that, in mixed-effects models, serial correlation still causes under-estimation of the variance of the fixed effects, but leads to increases in the variance of the random effects. This is because deceptively low, within-individual variability caused by serial correlation makes individuals appear less similar with each-other. Using hypothesis-testing to make population level inferences based on the random effects, rather than standard errors of fixed-effects models, is therefore more conservative.

Model selection is increasingly being used over hypothesis-testing to draw inferences about the distribution of a species but, if autocorellation is not modeled in the likelihood, use of ICs for model selection leads to a loss in parsimony. Explicitly modeling the autocorrelation in the data involves specifying a matrix containing all pair-wise correlations between the data points, with each correlation being specified as a function of the time between those points. This is computationally very demanding, because it requires the estimation of a $n \times n$ correlation matrix for a total number of n data points.

As illustrated in Fig. 2.3, serial dependence is aggravated by the existence of strong spatial autocorrelation in the covariates of usage. Therefore, the use of autoregressive or spatial error models for treating the effects of spatial autocorrelation (discussed in §2.2.5) could also help reduce the consequences of serial autocorrelation. This would involve specifying the correlation as a function of spatio-temporal displacement between two telemetry observations but to our knowledge no studies have, as yet, implemented these ideas for GLMMs.

In the absence of the technical and computational capability to model serial autocorrelation for large and complicated datasets, it is imperative to find a practical treatment of its consequences for model selection. I advocate the use of cross-validation, on the basis of arguments similar to those presented for spatial autocorrelation.

2.2.7 Animals are not equally detectible in different habitats.

If telemetry observations are received at a constant rate, their number in each habitat is an unbiased estimate of the proportion of time spent in that habitat. However, the rate of telemetry data acquisition may be affected by i) behavior (e.g. different detectability of individuals when traveling or foraging), ii) the environment (e.g. reduced signal transmission caused by dense forest canopy), and iii) satellite reception (e.g. orbital variability in satellite coverage), (Frair et al. 2004). The environment in which the animal lives can directly (e.g. forest canopy) or indirectly (i.e. change of behavior) affect detectability, biasing estimates of usage, and models of preference.

To treat environmental and reception-related biases, Frair et al. (2004) suggested quantifying the rate of data acquisition as a function of environmental variables and then incorporating this into the habitat preference model. However, independently measuring the probability of detection for inaccessible (e.g. marine) areas is difficult. Accounting for behaviorally-induced changes in detection probability is, generally, not possible.

Alternatively the path of the individual can be reconstructed using either interpolation or smoothing techniques, to obtain a sample of locations at regular time intervals. Although this reduces the bias in parameter estimates, the precision with which the position of the animal can be obtained from a reconstructed path, at any given instant, varies with the number of observations around that instant. Consequently, the response data in data-poor habitats or during cryptic modes of movement will be less precise. On the other hand, path reconstructions based on smoothing can improve overall precision by correcting some of the erroneous outliers in the raw data.

2.2.8 Animals respond non-linearly to their environment

Ecologists are interested in whether animals prefer or avoid certain environments but, also, in the, often non-linear, shape of their response to environmental variables. Although linear models (e.g. GLMs) can include non-linear transformations of covariates, it is often unknown a-priori what these functional relationships should be. Under a suitably flexible modeling framework, the appropriate functional form can be dictated by the data. Generalized Additive Models (GAMs) are extensions of GLMs that use scatter-plot smoothers to determine the appropriate functional form (Hastie and Tibshirani 1990, Wood 2006) between response and explanatory variables. The GAM equivalent of the linear predictor in eq. 2.8 is:

$$\eta_l = b_{0,m} + s_1(x_{l,1}) + \cdots + s_j(x_{l,j}) + \cdots$$
where   (2.13)
$$s_j(x_{l,j}) = \sum_{k=1}^{K} b_{k,j} f_{k,j}(x_{l,j})$$

Smoothers are classified as either parametric (e.g. *b*-splines, natural splines) or non-parametric (e.g. running mean, bin & kernel). Most parametric smoothers apply a set of pre-defined (e.g. cubic polynomial) transformations $f(\cdot)$, known as basis functions or the column of a spline, to an explanatory variable. Each basis function is constructed from the explanatory variable and a set of pre-specified points on the *x*-axis, known as knots, which are often based on quantiles of the explanatory variable (de Boor 1978). Each basis function is specified using a different set of knots and therefore behaves differently at different parts of the range of values taken by the explanatory variable (Fig. 2.4). The set of basis functions, evaluated at the covariate values, can be implemented as a new set of covariates, replacing each of the original environmental variables in eq. 2.8.
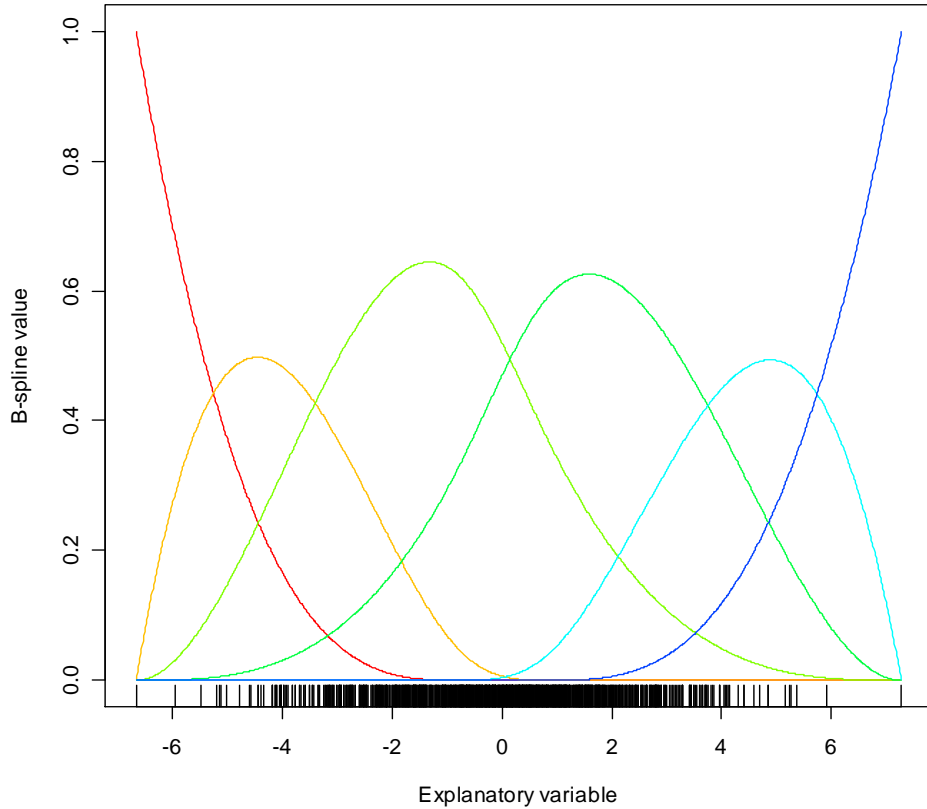
Fig. 2.4 B-spline values for a randomly generated explanatory variable that is normally distributed with a mean 0 and variance 4. Knot positions are at the 33% and 66% quantiles of the explanatory variable at -0.75 and 0.83, respectively.

In a mixed-model approach, the *b*'s are random variables rather than parameters (eq. 2.8 and 2.9). Since each basis function, applies to a particular range of the covariate, the mixed-effects approach models the amount of individual variation at different values of the covariate. This detects whether different individuals are affected by particular covariates but, also, whether the functional form of this relationship differs between individuals.

2.3 References

Aebischer, N. J., P. A. Robertson, and R. E. Kenward. 1993. Compositional Analysis of Habitat Use from Animal Radio- Tracking Data. Ecology **74**:1313-1325.

Agresti, A. 1996. An introduction to categorical data analysis. Wiley, New York.

Alldredge, J. R. R., and J.T. Ratti. 1986. Comparison of some statistical techniques for analysis of resource selection. Journal of Wildlife Management **50**:157-165.

Alldredge, J. R., and J. T. Ratti. 1992. Further Comparison of Some Statistical Techniques for Analysis of Resource Selection. Journal of Wildlife Management **56**:1-9.

Anselin, L. 2002. Under the hood - Issues in the specification and interpretation of spatial regression models. Agricultural Economics **27**:247-267.

Argos (1989). Guide to the Argos System. Argos CLS, Toulouse.

Arthur, S. M., Manly, B.F.J., McDonald, L.L. & Garner, G.W. 1996. Assessing habitat selection when availability changes. Ecology **77**:215-227.

Augustin, N. H., Mugglestone, M.A. & Buckland, S.T. 1996. An autologistic model for the spatial distribution of wildlife. Journal of Applied Ecology **33**:339-347.

Blundell, G. M., J. A. K. Maier, and E. M. Debevec. 2001. Linear home ranges: Effects of smoothing, sample size, and autocorrelation on kernel estimates. Ecological Monographs **71**:469-489.

Boor, C, de. 1978. A practical guide to splines, Springer-Verlag, New York.

Boyce, M. S., and L. L. McDonald. 1999. Relating populations to habitats using resource selection functions. TREE **14**:268-272.

Boyce, M. S., P. R. Vernier, S. E. Nielsen, and F. K. A. Schmiegelow. 2002. Evaluating resource selection functions. Ecological Modelling **157**:281-300.

Buckland, S. T., K. P. Burnham, and N. H. Augustin. 1997. Model selection: An integral part of inference. Biometrics **53**:603-618.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference. A practical Information-Theoretic Approach, 2nd edition edition. Springer-Verlag, New York.

Cliff, A. D., and J. K. Ord. 1973. Spatial autocorrelation. Pion, London.

Clifford, P., Richardson, S., and Hémon, D. 1989. Assessing the significance of the correlation between two spatial processes. Biometrics **45**: 123-134.

Cooke, S. J., S. G. Hinch, M. Wikelski, R. D. Andrews, L. J. Kuchel, T. G. Wolcott, and P. J. Butler. 2004. Biotelemetry: a mechanistic approach to ecology. Trends in Ecology & Evolution **19**:334-343.

Cramer, E. M. 1985. Multicollinearity. Pages 639-643 in S. Kotz, N. L. Johnson, and C. B. Read, editors. Encyclopaedia of statistics. John Wiley and Sons, New York.

Dean, C. B., M. D. Ugarte, and A. F. Militino. 2004. Penalized quasi-likelihood with spatially correlated data. Computational Statistics & Data Analysis **45**:235-248.

Diggle, P. J., J. A. Tawn, and R. A. Moyeed. 1998. Model-based geostatistics. Journal of the Royal Statistical Society Series C-Applied Statistics **47**:299-326.

Diniz, J. A. F., L. M. Bini, and B. A. Hawkins. 2003. Spatial autocorrelation and red herrings in geographical ecology. Global Ecology and Biogeography **12**:53-64.

Fedak, M. 2004. Marine Mammals as platforms for oceanographic sampling: a "win/win" situation for biology and operational oceanography. Mem. Natl Inst. Polar Res **58**:133-147.

Fox, J. 1997. Applied regression analysis, linear models, and related methods. Sage Publications Ltd., London.

Fox, J. 2002. Linear Mixed Models: Appendix to An R and S-PLUS Companion to Applied Regression. Sage Publications Ltd., London.

Frair, J. L., S. E. Nielsen, E. H. Merrill, S. R. Lele, M. S. Boyce, R. H. M. Munro, G. B. Stenhouse, and H. L. Beyer. 2004. Removing GPS collar bias in habitat selection studies. Journal of Applied Ecology **41**:201-212.

Gillies, C.S., Hebblewhite, M., Nielsen, S.E., Krawchuk, M.A, Aldridge, C.L., Frair, J.L., Saher, D.J., Stevens, C.E., and C.L. Jerde. 2006. Application of random effects to the study of resource selection by animals. Journal of Animal Ecology **75**: 887-898.

Griffith, D. A. 2005. A comparison of six analytical disease mapping techniques as applied to West Nile Virus in the coterminous United States. International Journal of Health Geographics **4**:1-14.

Hanley, J. A., A. Negassa, M. D. D. Edwardes, and J. E. Forrester. 2003. Statistical analysis of correlated data using generalized estimating equations: An orientation. American Journal of Epidemiology **157**:364-375.

Hastie, T. J., and R. J. Tibshirani. 1990. Generalized Additive Models. Chapman & Hall, London.

Hilborn, R. 1975. The effect of spatial heterogeneity on the persistence of predator-prey interactions. Theoretical Population Biology **8**: 346-355.

Hjermann, D. O. 2000. Analyzing habitat selection in animals without well-defined home ranges. Ecology **81**:1462-1468.

Huffer, F. W., and H. L. Wu. 1998. Markov chain Monte Carlo for autologistic regression models with application to the distribution of plant species. Biometrics **54**:509-524.

Isaaks, E. H., and R. M. Srivastava. 1990. Applied goestatistics. Oxford university press, Oxford.

Jansen, V.A.A. 1995. Regulation of predator–prey systems through spatial interactions – a possible solution to the paradox of enrichment. Oikos **74**: 384–390.

Johnson, D. H. 1980. The comparison of usage and availability measurements for evaluating resource preference. Ecology 6:65-71.

Jolliffe, I. T. 1990. Principal component analysis: A beginner's guide - I. Introduction and application. Weather **45**:375-382.

Kafadar, K., and P. S. Horn. 2002. Smoothing. Pages 2014-2020 in A. H. El-Shaarawi and W. W. Piegorsch, editors. Encyclopedia of Environmetrics. John Wiley & Sons, Chichester.

Keating, K. A., and S. Cherry. 2004. Use and interpretation of logistic regression in habitat selection studies. Journal of Wildlife Management **68**:774-789.

Keitt, T. H., O. N. Bjornstad, P. M. Dixon, and S. Citron-Pousty. 2002. Accounting for spatial pattern when modeling organism- environment interactions. Ecography **25**:616-625.

Lambert, D. 1992. Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. Technometrics **34**:1-14.

Legendre, P. 1993. Spatial Autocorrelation - Trouble or New Paradigm. Ecology **74**:1659-1673.

Legendre, P., Oden, N.L., Sokal, R.R., Vaudor, A., and Kim, J. 1990. Approximate analysis of variance of spatially autocrrelated regional data. Journal of Classification **7**: 53-75

Lichstein, J. W., T. R. Simons, S. A. Shriner, and K. E. Franzreb. 2002. Spatial autocorrelation and autoregressive models in ecology. Ecological Monographs **72**:445-463.

Manly, B. F. J., L. L. McDonald, and D. L. Thomas. 1993. Resource selection by animals: statistical design and analysis for field studies. Chapman and Hall, London

Matthiopoulos, J. 2003a. Model-supervised kernel smoothing for the estimation of spatial usage. Oikos **102**:367-377.

Matthiopoulos, J. 2003b. The use of space by animals as a function of accessibility and preference. Ecological Modelling **159**:239-268.

Matthiopoulos, J., B. McConnell, C. Duck, and M. Fedak. 2004. Using satellite telemetry and aerial counts to estimate space use by grey seals around the British Isles. Journal of Applied Ecology **41**:476-491.

McCullagh, P., and J. A. Nelder. 1989. Generalized Linear Models (2nd edition). Chapman & Hall, London.

Morales, J.M., Haydon, D.T., Frair, J., Holsinger, K.E., and J.M. Fryxell. 2004. Extracting more out of relocation data: building movement models as mixtures of random walks. Ecology **85**: 2436-2445.

Otis, D. L. W., and G.C. White 1999. Autocorrelation of location estimates and the analysis of radiotracking data. Journal of Wildlife Management **63**:1039-1044.

Overmars, K. P., G. H. J. de Koning, and A. Veldkamp. 2003. Spatial autocorrelation in multi-scale land use models. Ecological Modelling **164**:257-270.

Pinheiro, J. C., and D. M. Bates. 2000. Mixed-effects models in S and S-plus. Springer-Verlag, New York.

Ripley, B. D. 1981. Spatial statistics. John Wiley & Sons, New York.

Rooney, S. M., A. Wolfe, and T. J. Hayden. 1998. Autocorrelated data in telemetry studies: time to independence and the problem of behavioural effects. Mammal Review **28**:89-98.

Silverman, B. W. 1986. Density estimation for statistics and data analysis. Chapman & Hall, London.

Solow, A. R. 1989. A Randomization Test For Independence of Animal Locations. Ecology **70**:1546-1549.

Stein, A., and L. C. A. Corsten. 1991. Universal Kriging and Cokriging as a Regression Procedure. Biometrics **47**:575-587.

Stephenson, C. M., M. L. MacKenzie, C. Edwards, and J. M. J. Travis. in press. Modelling establishment probabilities of an exotic plant, Rhododendron ponticum, invading a heterogeneous, woodland landscape using logistic regression with spatial autocorrelation. Ecological Modelling **193** (3-4) : 747-758.

Swihart, R. K. S., and N.A. Slade 1985. Testing for independence of observations in animal movements. Ecology **66**:1176-1184.

Thomas, A., N. Best, D. Lunn, R. Arnold, and D. Spiegelhalter. 2004. GeoBUGS User Manual. Rolf nevanlinna Institute, Helsinki.

Turchin, P. 1998. Quantitative analysis of movement; measuring and modeling population redistribution in animals and plants. Sinauer, Sunderland, Massachusetts.

Welsh, A. H., R. B. Cunningham, C. F. Donnelly, and D. B. Lindenmayer. 1996. Modelling the abundance of rare species: Statistical models for counts with extra zeros. Ecological Modelling **88**:297-308.

Wood, S. N. 2006. Generalized Additive Models: an introduction with R. In press.

Appendix 2.A

To derive a model that does not require discrete geographical space or covariates, I envisage two distinct, limiting operations: Operation 1 (O1) involves indefinitely increasing the resolution of the spatial grid. Eventually, this will lead to most spatial-grid cells being empty of observations and occupied cells containing, at most, one observation. Operation 2 (O2) corresponds to an arbitrary increase in the resolution of the grid in environmental space. This will eventually lead to each cell in space being a unique habitat.

I first increase the resolution of the spatial grid (O1) which means that the number of telemetry observations occurring in the ith habitat tends to become the same as the number (u_i) of cells of that habitat that contain an observation. Using eq. 2.1, I can rewrite preference as

$$w_i = \frac{u_i}{a_i} = \frac{u_i}{\left(o_i + u_i\right)c} \tag{A1}$$

where o_i is the number of unoccupied spatial-grid cells of the ith habitat and c is the area of a cell in the spatial grid.

Operation O1 leads to an ever-increasing number of cells to a level beyond practical use. To overcome this problem, I use a case-control approach (Prentice and Pyke 1979, Stephenson et al. 2006) that was originally used for the analysis of rare diseases and is designed to deal with data sets containing presence and absence data (the cases and controls, respectively). To obtain the controls, an arbitrary number of absences are retrospectively selected from the same population as the cases. In the context of telemetry studies, this implies the random selection of a number of points in space. Because O1 leads to an increasing number of unoccupied cells and a finite number of used cells, the probability of selecting a used cell tends to zero. Furthermore, the asymptotic theory on case-control studies (Prentice and Pyke 1979)

ensures that, as long as the sample of controls is a) sufficiently large compared to the number of telemetry locations and b) representative of all accessible points in geographical space, the estimates of all coefficients (except the intercept) will not be sensitive to sample size.

I denote by k_a the total number of spatial-grid cells selected as controls and by p_i the probability of selecting a cell of the ith habitat. I assume that this probability is given by the relative availability of that habitat

$$p_i = \frac{a_i}{\sum_{all\ j} a_j} \tag{A2}$$

Then, $k_a p_i$ gives the expected number of cells of habitat i contained in the control. The expected proportion of used cells of the ith habitat in the case-control sample is

$$h_i = \frac{u_i}{k_a p_i + u_i} \tag{A3}$$

Comparing the values of these proportions for any two given habitats as k_a gets very large yields

$$\lim_{k_a \to \infty} \frac{h_1}{h_2} = \lim_{k_a \to \infty} \frac{(u_1 u_2)/k_a + u_1 p_2}{(u_1 u_2)/k_a + u_2 p_1} = \frac{u_1/a_1}{u_2/a_2} = \frac{w_1}{w_2} \tag{A4},$$

indicating that, under the case-control paradigm, the quantity h_i defined in eq. A3 can be treated as proportional to preference,

$$\lim_{k_a \to \infty} h_i \propto w_i \tag{A5}$$

I can model the observed number of presences in the case-control sample of cells from the i th habitat as a realization from a Binomial process with probability h_i and number of trials n_i (the total number of cells in the case-control data set that belong to the ith habitat).

$$\hat{u}_i \sim \mathrm{B}(n_i, h_i) \tag{A6}$$

Now, consider performing O2. As a result, habitat i in environmental space is either present or absent in geographical space,

$$n_i \in \{0,1\} \tag{A7}$$

The implication of $n_i = 0$ is that this habitat was not available to the animal and is therefore not considered in further analysis. Thus, the process in eq. A6 becomes a Bernoulli

$$
\begin{aligned}
&\hat{u}_l \sim \mathrm{B}(1, h_l) \\
&h_l = g^{-1}(\eta_l) = \frac{e^{\eta_l}}{1 + e^{\eta_l}} \\
&\eta_l = \beta_0 + \beta_1 x_{l,1} + \cdots + \beta_j x_{l,j}
\end{aligned} \tag{A8}
$$

Originally the subscript i referred to a specific habitat available to the animal. To avoid confusion, I replace i by the subscript l referring to a specific habitat or data point in the case-control sample.

_____

3. MODELLING SPACE-USE AND HABITAT PREFERENCE IN GREY SEALS (HALICHOERUS GRYPUS) OFF THE EAST COAST OF SCOTLAND

_____

3.1 Grey seal natural history

An estimated 130,000 grey seals inhabit the coasts and seas around the British Isles (SMRU 2005). They spend long periods of time on land during the breeding (October to November) and moulting (January to March) seasons (Bonner 1981). During the remainder of the year, individuals frequently aggregate on coastal sites, known as haul-outs. When setting out to forage at sea, grey seals perform predominantly (88% of times) return trips each lasting, on average, 2.33 days. However, they are not completely site-faithful, occasionally performing transition trips to other haul-outs which can be hundreds of kilometres away (McConnell et al. 1999).

Grey seals are generalist predators, feeding on more than 20 prey species (Hammond et al. 1994a, Hammond et al. 1994b, Thompson et al. 1996). There is considerable individual, spatial and temporal variation in their diet, which is partly believed to be due to spatio-temporal variation in the abundance of different prey. Nevertheless, sandeels (Ammodytes marinus), a small cryptic species that spend part of their time buried in coarse sediment

(Wright et al. 2000), are a major component of grey seal diet (Thompson et al. 1991, Thompson et al. 1995, McConnell et al. 1999).

3.2 Methods

3.2.1 Response variable.

In the period 1991 to 2001 a total of 58 grey seals were caught at the Farne Isles (55°38' N, 1°37' W), Abertay (56º24' N 3º05' W) and Isle of May (56˚19' N, 2˚56' W) haulout sites (Fig. 6a). Each animal was anaesthetized, fitted with a Satellite Relay Data Logger (SRDLs) and released (McConnell et al. 1999).

During their lifetime, the SRDLs sent UHF signals to two polar-orbiting satellites with an Argos satellite system. The location of the animal was determined using the frequency Doppler shift of the signal (Argos 1989). These estimates are vulnerable to bias and imprecision when they are based on a low number of successive uplinks. In particular, because the distribution of the Argos observation error for poor-quality locations is thick-tailed (Vincent et al. 2002) the data contained a small number of highly erroneous location fixes which were removed by applying the filtering algorithm described by (McConnell et al. 1992), using a maximum speed parameter of 5 m/s. To treat observation error in the remaining data I used a smoothing algorithm developed (M. Lonergan unpublished) within the MGCV (Wood 2001) package in R. Briefly, this uses a Generalized Additive Model to produce a smooth path in space as a parametric function of time. In this, Location Quality (LQ) is accounted for by weighting the influence of different locations by the inverse of their associated error variance (as measured experimentally by (Vincent et al. 2002)). I used the output of this algorithm to interpolate the animals' positions at 3-hourly intervals. Obtaining temporally regular estimates of the animal's position also helped to reduce bias due to variable observability of the animals (see §2.2.7).

Habitat preference depends on the type of activity performed by the animals. Different habitats may be suitable for different activities such as resting, breeding, migrating or foraging. I was primarily interested in foraging behaviour and therefore restricted my attention to return trips from

the same haul-out and removed all locations that were at, or close to, the haulout-sites. For the purposes of this work a return-trip was defined as a foray outside the 5km radius around the haul-out site. For the analysis I only included individuals for which I had in-excess of 7 days of return-trip data. I used the data from the Farne Isles for fitting and model-selection and those from Abertay as a validation data set.

The error-corrected and temporally interpolated data set of satellite locations represented a sample of points in geographical space visited by the tagged grey seals (presence data). However, they were only part of the data-set for the response variable. Under the case-control paradigm, to provide the contrast necessary for modelling preference, these had to be complemented with a set of points not visited by the study animals (absence data). The Bernoulli response variable in eqs 2.14 took the value 1 wherever there was an animal present and 0 where there was not.

I selected the absence points from the surface of accessibility calculated by Matthiopoulos (2003) using information from the entire population of UK grey seals and described by the simple relationship

$$\alpha = \left(\frac{d}{5}\right)^{-1.98} \qquad 3.1$$

where d is the distance (in km) between every point in space and the haulout. Rather than Euclidean distance, I used an algorithm developed in-house (J. Matthiopoulos, unpublished) to calculate at-sea distance, defined as the shortest distance needed to travel between any point at sea and the haulout without crossing land.

The mathematical results pertaining to case-control studies (Prentice and Pyke 1979) are asymptotic, meaning, in this context, that habitat availability is only approximately represented in the sample of response data, the approximation improving with an increasing number of absence data. The number of absence data necessary to obtain a sufficiently good approximation will vary from study to study. By experimenting with different proportions of presence and absence data I concluded that, for the

data, model parameters remained effectively unchanged when the absence data were at least twice as many as the presence data.

3.2.2 Explanatory variables

For each set of spatial coordinates included in the response data I collated a set of values for the explanatory variables corresponding to local environmental conditions. I selected environmental variables on the basis of possible biological relevance and the availability of data. As is often the case with studies of habitat preference, the variables that are seen as the most relevant to the study-species are rather data-poor. In studies of higher marine predators, such as the grey seal, this predicament is particularly evident in the sparsity of prey data. Since the primary off-shore activity of seals is foraging, it is certain that their off-shore distribution is influenced by the distribution of their prey. This is likely to be a complex relationship because grey seals feed on a large number of prey species, each having its own patterns of spatial and temporal variation. I found little readily usable data on these patterns. Instead, I opted for three static variables: sediment type, sea depth and distance from the haulout. Although this decision was primarily motivated by current data availability, it was further justified by the requirements for a parsimonious, predictive and, yet, biologically relevant model: A model using static environmental variables is more parsimonious because it only requires that the distribution of foragers is the result of time-invariant environmental cues, rather than an optimised/adaptive foraging strategy (Ollason 1980).

Furthermore, developing a predictive model on grey seal distribution that relied on the continued measurement of species that are just as variable would defeat the purpose of the modelling exercise. So, even if distributional data were currently available for all or some prey species, they would be of little use for future model predictions. On the other hand, if prey distributions are predictable, either by the seals or by human observers, it will partly be from cues provided by geophysical variables such as the ones I have used

Indeed, grey seals are suspected to prefer foraging in certain sediment types. This is either a direct consequence of the seals' bio-turbating search tactics (Bowen et al. 2002) or an indirect consequence of the habitat preference of sediment-burrowing prey such as sandeels (Wright et al. 2000, Holland et al. 2005). Sediment type was derived from British Geological Survey (BGS) data obtained from core samples, spaced, on average, at 5km intervals throughout the study area. For every core sample, the data were given as a percentage-by-weight of gravel (defined as particles greater than 2.0mm in diameter), sand (particles 0.0625-2.0mm in diameter) and mud (particles smaller than 0.0625mm in diameter). I used a random sub-sample of cores to calculate the semi-variogram (Isaaks and Srivastava 1990) characterising the spatial autocorrelation of each of the three sediment components. The semi-variograms were then used to generate kriged estimates of each sediment component throughout the study region, at a resolution of 1km2. I kriged the three sediment components independently despite the fact that they are measured as complementary percentages. To check that the three kriged maps were approximately complementary, I checked the distribution of the sum of estimated percentages over all grid cells in the study region. This had an average value of 101% with standard deviation of 5%. These estimates were then normalised to 100% for each cell in the study area. In the analysis, I only used two (i.e. the percentage mud and gravel) out of three sediment components to avoid severe problems of multi-collinearity (see §2.2.5).

I also included sea depth as a potential covariate. Although grey seals probably take fish from the entire water column, they spend a large proportion of their dives foraging close to the seabed. It is conceivable that seals have a preference for a particular range of depths. This could be due to physiological restrictions in maximum dive depth, the need to make efficient use of maximum dive duration, or possible depth-segregation between various prey. The two primary sources of bathymetry are the British Geological Survey (BGS - 200,000km of geophysical line-transects collected from 1966 to 1987) and United Kingdom Hydrographic Office Admiralty Charts and Survey data. These were combined and collated by the BGS into 1:250,000 geological maps and digitised into Digbath250. This digitised data

is represented in the form of bathymetric contour lines every 10 meters between 0 and 100m, every 20m between 200m and 400m and every 100m at depths exceeding 400m. I transformed these maps using an equal-distance projection (National Grid of Great-Britain). I placed points at 1km intervals along the bathymetric contour lines and interpolated linearly between these points to obtain depth estimates for every point of the regular, 1 km grid (see also Fig. 3.4a).

Finally, to account for residual issues of accessibility, I used distance from the seals' haulout as the final environmental variable. This was obtained over a grid of 1km resolution (See also Fig. 3.3a).

3.2.3 Model structure

The basic structure of the model follows from using a binomial response variable (h) with a logit link

$$\hat{u}_l \sim \mathrm{B}(1, h_l)$$
$$h_l = g^{-1}(\eta_l) = \frac{e^{\eta_l}}{1 + e^{\eta_l}}$$

3.2

To increase the model's flexibility (problem 8), the linear predictor (η_l) was structured as a GAM with a maximum of four explanatory variables

$$\eta_l = b_0 + s_1(\mathrm{distance}_l) + s_2(\mathrm{depth}_l) + s_3(\mathrm{mud}_l) + s_4(\mathrm{gravel}_l)$$
$$= b_0 + \sum_{k=1}^{6} b_{1,k} f_{1,k}(\mathrm{distance}_l) + \cdots$$

3.3

where each spline $s(\cdot)$ is a flexible function of each explanatory variable, and $b_{j,k}$ is the random effect parameter for the kth out of 6 basis function of the spine for the jth explanatory variable. I used b-splines because they afford local control in modelling the response at different domains of the explanatory variable and produce robust result in data-poor regions of environmental space (Ramsay 2004). For each explanatory variable, I used a

composite spline with 6 df, two internal knots at the 33% and 66% quantiles of the frequency distribution of observed values for the environmental variable at the animal locations. Although it is possible by means of cross-validation (Wood 2001, 2006) to automatically select the most appropriate number of splines and knot positions, computational restrictions forced me to pre-specify these. However, as a minimum preventative measure against over-fitting, I compared the performance of the GAM with the output of a GLM.

To account for unbalanced sampling effort (2.2.3) I implemented eq. 3.3 as a mixed-effects model. Each model parameter $b_{j,k}$ was treated as a normal variable, containing a fixed effect $\beta_{j,k}$ (applicable to the entire population of animals), and a random effect error term $v_{j,k} \sim N(0, \sigma_{j,k})$, representing the variability in the response of different animals to the kth spline of the jth explanatory variable. This model is similar to a Generalized Additive Mixed Model (GAMM - Zhang and Davidian 2004, Wood 2006).

Part of this variability may be due to characteristics of the animals such as age, sex, weight or body length. Incorporating these in the model is important for two reasons: First, it helps account for biases in sampling effort across different types of animals, particularly so for the purposes of population-level predictions (see §3.2.6). Second, it is a practical way of modelling deviations from normality using random effect (§2.2.3).

In the most saturated form of the model, each model parameter, was expressed as a function of individual characteristics

$$b_{j,k,m} = \beta_{j,k,0} + \beta_{j,k,1} sex_m + \beta_{j,k,2} length_m + v_{j,k} \qquad\qquad 3.4$$

where m refers to an individual animal and the individual characteristic "length" is measured from nose to tail and $v_{j,k}$ denotes the random effects. Other candidate individual characteristics were mass and age of the animal. I chose to exclude mass because it is strongly correlated with length and because it is harder than length to observe remotely (e.g. via aerial survey). Age was excluded because it wasn't recorded for 5 individuals and its

inclusion would therefore necessitate a reduction in the number of animals used in the analysis.

3.2.4 Parameter estimation

The parameters that needed to be estimated were the model's coefficients $\beta_{j,l}$ and the elements σ_j^2 of the random effects. The estimation methods that are most often used for GAMMs are penalized quasi-likelihood (PQL) or maximum-likelihood methods (ML). PQL is generally fast, produces nearly unbiased parameter estimates, but doesn't produce a likelihood estimate that can be used for model selection. ML techniques, on the other hand, produce exact marginal likelihood estimates, but are computationally intensive because, for every candidate set of parameter values, the likelihood-maximization algorithm needs to perform a numerical approximation of the likelihood over all possible realizations of the model's random effects term. In general, the likelihood function is written (Pinheiro and Bates 2000).

$$L(\beta,\psi/\hat{u}) = \prod_{m=1}^{M} \int p(\hat{u}_m/b_m,\beta,\psi)p(b_m/\psi)db_m \qquad 3.5$$

where b_m is a vector of random effects, each applying to a kth basis function for the spline of the jth covariate (eq. 3.3) of the mth of a total of M individuals. The integrand in eq. 3.5 consists of the response and the random effect components. The random effects component ($p(b_m/\psi)$), is a multivariate normal probability density function with means 0 and variance-covariance matrix ψ. The random effect variances (diagonal elements of ψ) represent the individual variation in the animals' response to environmental variables. The covariances (off-diagonal elements of ψ) quantify within-individual correlations in the response to different environmental covariates. If q is the number of random effects included in the model, then the variance-covariance matrix ψ is a $q \times q$ matrix. For flexible models such as GAMMs, q is generally large and estimating ψ is numerically difficult. To

reduce complexity, I restricted estimation to the diagonal terms (variances) and set the off-diagonal elements of ψ to zero.

In this study, the response component of the likelihood function is a binomial probability density function

$$p\left(\hat{u}_m / b_m, \beta, \psi\right) = \prod_{l=1}^{n} h_l^{\hat{u}_{m,l}} \left(1 - h_l\right)^{\left(1 - \hat{u}_{m,l}\right)} \qquad 3.6$$

where $\hat{u}_{m,l}$ is lth observation from the mth individual and h_l is the predicted response as defined in eqs. 3.2-3.4.

The ML estimates are those values of the parameters β and ψ, that maximize eq. 3.5. I used the Automatic-Differentiating Model-Builder (ADMB) and its Random Effects module (Otter Research Ltd 2004, Skaug & Fournier 2003) to minimize the quantity $-L\left(\beta, \psi / \hat{u}\right)$ (see Appendix A). ADMB-RE first approximates the likelihood function using the Laplace approximation and Importance sampling. It then uses automatic differentiation to obtain exact derivatives of this likelihood approximation, which is maximized with a quasi-Newton method with line search (Skaug 2002, Fletcher 1987).

3.2.5 Model selection and model validation

Model-selection examines if the improvement in the quality of fit, gained by adding model variables, justifies the associated increase in model complexity. This trade-off between parsimony and goodness-of-fit is adjudicated by metrics known as model-selection criteria. The choice of model-selection criterion is not straightforward, not least because parsimony is difficult to quantify, and depends on the study's objectives.

All information criteria (IC) such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) (Burnham and Anderson 2002) achieve parsimony by penalizing the likelihood of the model by the number of parameters it contains. However, the effectiveness of IC relies on the correctness of the assumptions underpinning the likelihood function. In

telemetry studies, violation of the independence assumption can lead to over-fitted models. To avoid this problem, I used IC only as a rough guide and relied on cluster-level cross-validation for final model selection. I initially fitted GAMMs to a subset of the data (19 out of 29 individuals) from the Farne Isles. I started with an intercept-only model and used forward model-selection (with AIC) to arrive at a model containing all four environmental variables. This reduced the number of models to be investigated by cross-validation from a possible 41 to 5. From these five models I selected the one that best predicted the data from the remaining 10 animals. This procedure still left room for over-fitting because the number of knots used for each environmental variable was pre-specified. I therefore replaced the spline-based models with linear terms, if these simpler models attained a higher likelihood under the validation data set.

The model obtained via this selection process was then extended with individual characteristics (i.e. sex and body-length). Again, forward model selection based on AIC was used to arrive at a set of candidate models, each of which was validated against the test data set to yield a final model. To assess the predictive performance of this model, I estimated its likelihood under data from 13 individuals from a neighbouring sub-population, in Abertay. I compared this value with, similarly calculated, likelihoods for 5 other candidate models. Goodness-of-fit and predictive performance was also visually assessed by generating spatial predictions for both the Farne and Abertay.

3.2.6 Predicting spatial usage

The major objective of this study is to estimate the spatial usage of the entire sub population using data from only a sample of individuals. Eq. 3.2 can be used to calculate the expected usage $\upsilon(\mathbf{s})$ of every point in space $\mathbf{s}$, with known values for the environmental variables and known accessibility $\alpha(\mathbf{s})$.

This requires predicting h_l which, for random-effects models, is not straightforward: Assuming, for simplicity, that individual characteristics are

not included in the model, the random effects b_j are normally distributed with mean $\bar{b}_j = \beta_j$ and variance σ_j^2 (Fig. 3.1a). The response variable h_l is modelled as a non-linear function (i.e. the inverse of the logit) of the linear predictor which contains these random effects. This means that the distribution of the random effects viewed on the scale of the response is transformed (Fig. 3.1b) and, therefore, the fixed effects part β_j is no-longer the mean of this distribution. Consequently, the predictions of usage generated from the model using the fixed effects alone, are not the same as the average of prediction generated using multiple realizations of the random effects.

a.   b.
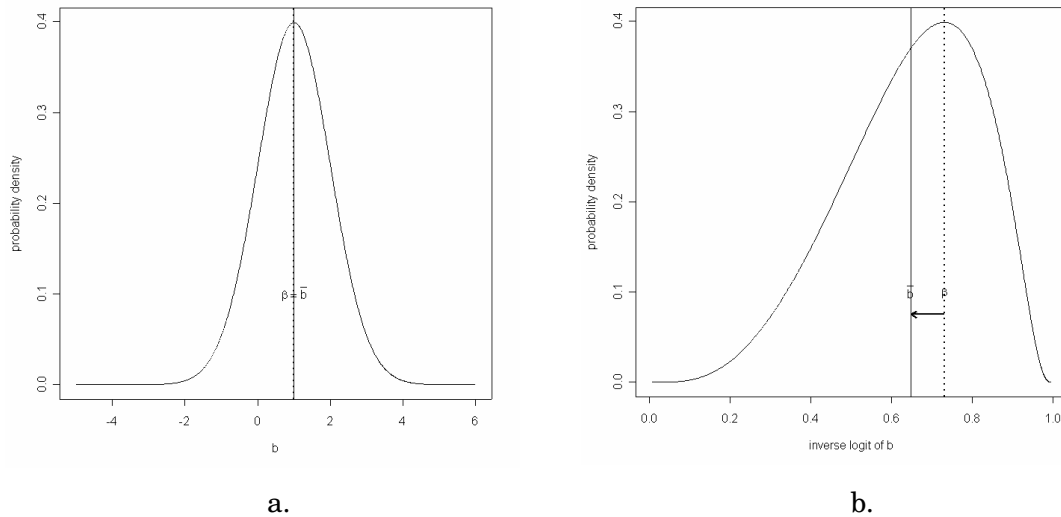
Fig. 3.1 The normal probability density function of the random effects (with mean and variance equal to 1) on the scale of the linear predictor. The mean of the random effects' distribution is equal to the fixed-effect component (i.e. $\beta = 1$) of the random effect (**a**). The probability density function of the random effect on the scale of the response. The mean of this random effects distribution is not equal to the fixed effect component of the random effect on the scale of the response (i.e. inverse logit of $\beta = 1$ - **b**)

In practice, the best way to calculate h_l, is as the average of a random sample of predictions generated from the estimates for the fixed effects β_j, incremented by a value drawn from a normal distribution with mean zero and variance σ_j^2.

For models that include individual characteristics, such as sex and body length, the values for these can be drawn from within the pool of observed values associated with the sample of tagged individuals. However, biases in catching effort will still lead to biased predictions in usage. A better alternative is to use samples of individual characteristics from other sources. For example, the age- or sex-structure could be obtained from demographic models and the distribution of length within the population could be derived from aerial surveys of haulout sites if it can be assumed that the animals at a haulout are representative of the population.

I predicted usage $\upsilon(\mathbf{S})$ on a 1 km resolution grid. Every such map obtained for a single realization of the model's random effects can be thought of as the space-use of one individual from the population. However, such surfaces did not exactly add up to 1 (and needed to be normalized) primarily for two reasons: First, the environmental conditions at some points in space were outside the range of the data used to fit the model. Predictions for these points in space are equivalent to extrapolation in environmental space and can lead to spurious, local over-estimates of usage. In this study, the distribution of values for the sediment and depth variables was heavily skewed to the right. I therefore excluded predictions outside the ranges of these environmental variables. Second, only a sample of points was included in the analysis and therefore, by chance, relatively more (or less) preferred environmental conditions might be present within the study area compared to the data.

For conservation purposes, not the relative, but the absolute population density at sea is often of main interest. If the population size N is known as well, multiplying $\upsilon(\mathbf{x})$ by this scalar will normally yield an absolute measure of population density for each grid cell at sea. For the Farne Islands the total population size is estimated to be 2950 (CI: 2742-3159 -

Matthiopoulos et al. 2004). The estimated population density at sea would be correct if predictions were made for all points in space. This is not the case, since the area within 5 km of the haul-out site, points in space for which no environmental data is available (e.g. points on or very close to the shore) and all points in space that are outside the prediction interval, are excluded. The proportion of time spend in the excluded areas by the population can be approximated by the fraction p of animal locations that fall inside these regions. Next, usage in a region c (e.g. a 1 km² cell) in space can be estimated as follows

$$u_c = N(1-p)\int\limits_{\mathbf{s}\in c} \upsilon(\mathbf{s})d\mathbf{x} \qquad\qquad 3.7$$

One note of caution is that such estimates of usage might be underestimated in some regions. Due to the error in the animal location estimates, locations at sea but close to land have a higher probability of falling on land and therefore being excluded from the analysis.

3.2.7 Predicting preference

Grey seals spend a considerable portion of their time offshore. Those areas in which they spend relatively more time than expected by chance (taking into account accessibility §2.2.2), are said to be preferred. Spatial plotting of preference is useful, because it indicates the position and extent of these areas, also known as 'hotspots'. Preference w at a point $\mathbf{s}$ in space can be calculated as follows

$$w(\mathbf{s}) = \frac{\upsilon(\mathbf{s})}{\alpha(\mathbf{s})} = r\frac{h_l}{1-h_l} = re^{\eta} \qquad\qquad 3.8$$

Eq. 3.8 holds for all values of r. As the number r of controls per observation increases, the response variable h becomes small, usually much smaller than

0.5. For these values of h the ratio $h/(h-1)$ is approximated well by h. I can therefore write that for large r,

$$w(\mathbf{s}) = r\frac{h_l}{1-h_l} \cong h_l r \propto h_l \qquad\qquad 3.9$$

Which is in agreement with the statement in eq. A4 in chapter 2, derived for r tending to infinity. Eq. 3.8 relies on accurately modelling accessibility $\alpha(\mathbf{s})$. To account for biases in $\alpha(\mathbf{s})$, I included distance to the haulout as an additional covariate in the model. Hence, plotting $w(\mathbf{s})$ from eq. 3.8, does not purely reflect the animals' preference for environmental variables. By breaking up the linear predictor into three additive components, preference can be re-written

$$\frac{\upsilon(\mathbf{s})}{\alpha(\mathbf{s})} = r\exp(\beta_0)\exp(\eta_{\text{distance}})\exp(\eta_{\text{environmental variables}}) \qquad\qquad 3.10$$

which implies

$$\frac{\upsilon(\mathbf{s})}{\alpha(\mathbf{s})\exp(\eta_{\text{distance}})} \propto \exp(\eta_{\text{environmental variables}}) \qquad\qquad 3.11$$

The expression on the left is preference, corrected for the biases in the accessibility model, and therefore $\exp(\eta_{\text{environmental variables}})$ can be used to generate spatial plots of preference.

3.2.8 Software

All data manipulation, analyses and plotting (including spatial plotting) was done using R (R Development Core Team 2004). ML parameter estimation was done using ADMB-RE (Otter Research Ltd. 2004).

3.3 Results

3.3.1 Response and environmental variables

From a total of 58 individuals caught on the Farne Isles, Abertay and the Isle of May, 42 made return trips to and from the Farne Isles and Abertay for a sufficiently long time to be included in the analysis. Tag life varied greatly between those individuals (minimum 2.5d, maximum 329d, average: 109d) leading to large differences in sampling intensity between individuals. These variations were accommodated by the mixed effects structure of the model (§2.2.3). Sampling intensity also varied temporally, with the most data in August and the least in February (Fig. 3.2). This means that pooling the data across the year will tend to bias the results if habitat preference of grey seals has a seasonal component.
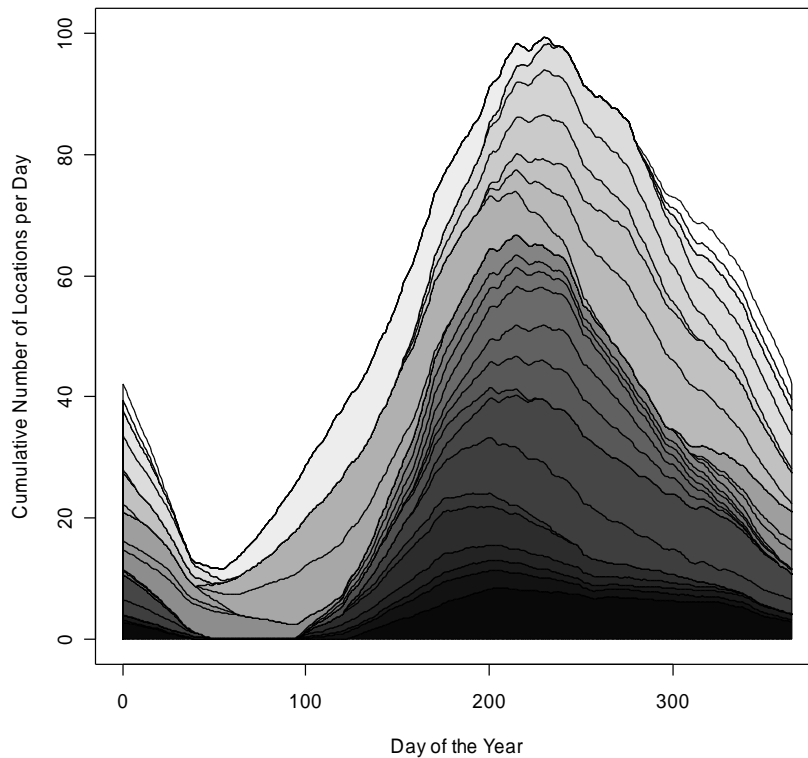
Fig. 3.2 Data collection intensity within a year and between individuals from the Farne Isles. This figure shows the cumulative number of location fixes for every day within a year. Estimates for each day were obtained using a 30d moving average. Each grey shade represents a different individual.

Following pre-processing of the satellite data (error-correction and path reconstruction), the Farne data set comprised a total of 2315 animal locations (Fig. 3.3a). I randomly select twice as many (i.e. 4630) points from the accessibility surface (Fig. 3.3b). The combined presence and absence data constituted the response data that were to be regressed against the environmental variables (examples shown in Fig. 3.4).
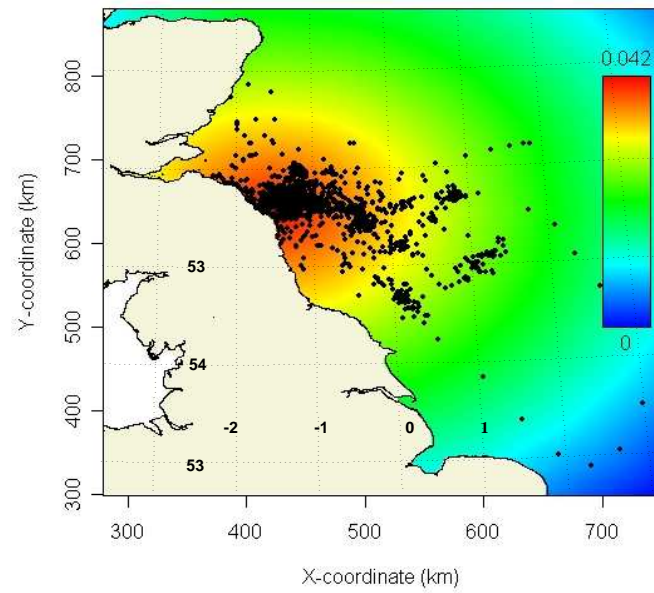
a.
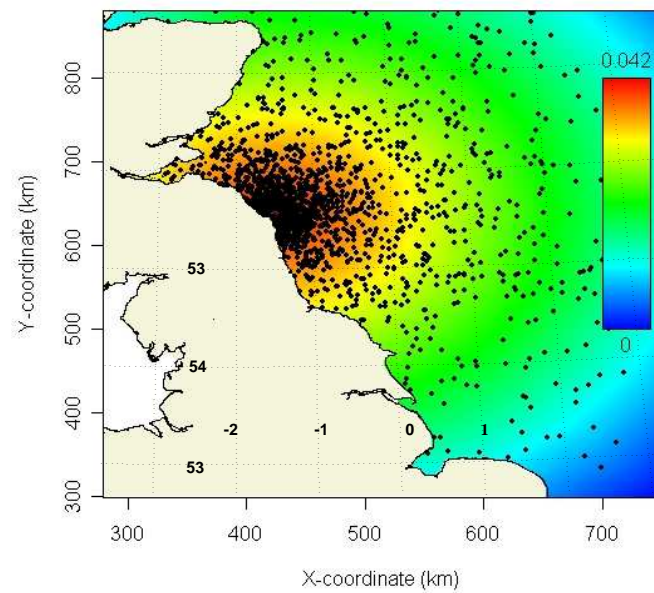
b.

Fig. 3.3 Spatial distribution of filtered grey seal locations (**a** - response = 1) and locations drawn from the accessibility surface (**b** - response = 0). The accessibility of each point in space is plotted in the background (**a** and **b**).
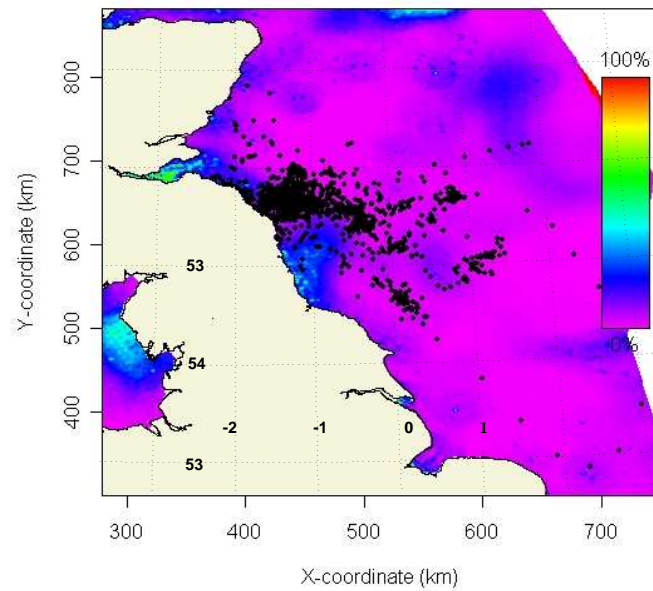
a
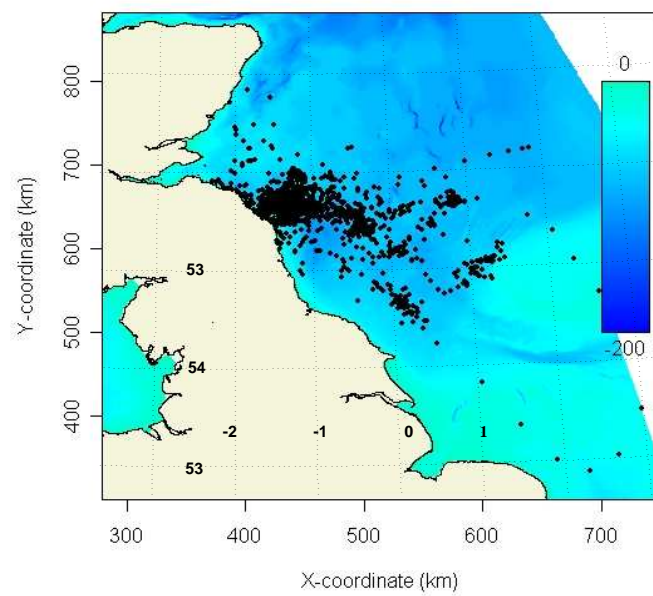
b

Fig. 3.4 Spatial plots of % mud in the sediment (**a**) and sea bottom depth (**b**).

3.3.2 Model selection

Forward model selection using the Farnes data suggested that a model containing all environmental variables (see Appendix B, table 1 for the parameter estimates) explained a significant proportion (31%) of the observed variation in the response data, taking model parsimony into account (Figs 3.5 and 3.6). The variables distance to the haulout site and % mud in the sediment explained most, accounting for 87% of the explained deviance. The saturated model also performed best in explaining the observed variation in the response for the Farnes validation data set. It was conceivable that the saturated model was too flexible due to the use of splines. According to the AIC, the sequential replacement of the splines by the un-transformed environmental variables led to a deterioration of the model. However, the model with no splines for gravel had the highest likelihood under both validation data sets. This model was used for further extensions.

I included individual characteristics to the random effects in an attempt to explain some of the observed individual variability in the response to different environmental variables. Based on the AIC an interaction between %mud and length, %mud and sex, distance to the haulout and length and %gravel and length led to better models (Figs 3.7 and 3.8). Under the validation data set, only the interaction between the nose-to-tail length of the individual and % mud in the sediment led to an improvement. This model was used to investigate the relationship between the response and the different environmental variables and to make spatial predictions of usage and preference.

Finally I assessed the predictive performance of the candidate models. For the Abertay data the saturated model was outperformed slightly by a simpler model not containing depth. However all inclusions of individual characteristics lead to a deterioration of the model.
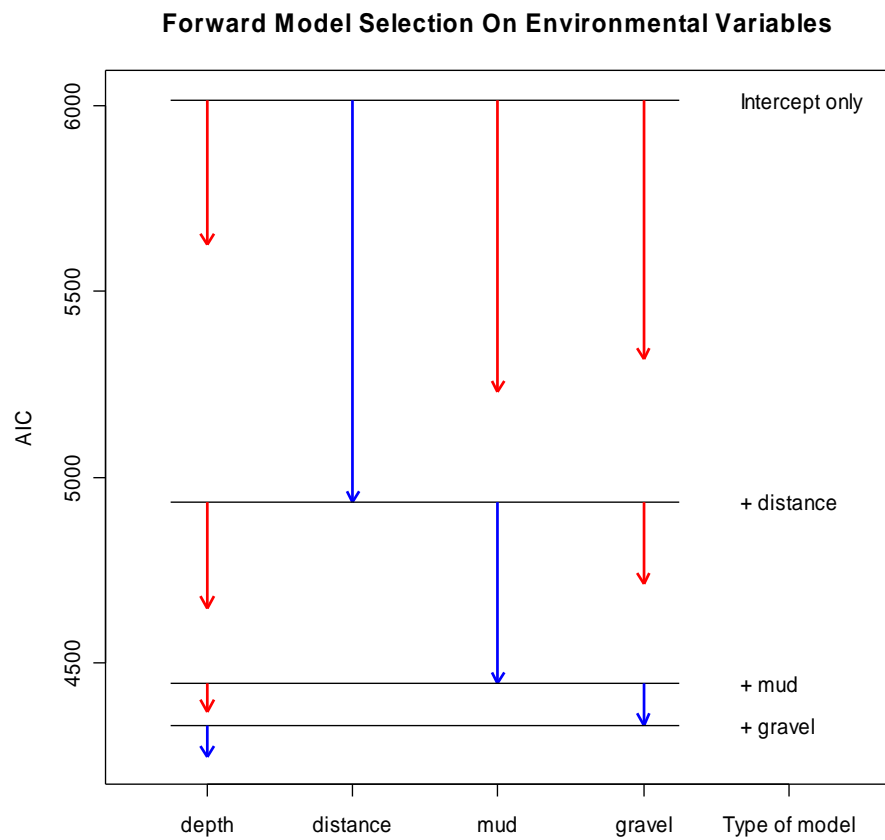
Fig. 3.5 Forward model selection on environmental variables. The length of the arrows indicates the change in AIC from the current model (horizontal line) as a result of adding an environmental variable not yet included in the model. Arrows for variables that lead to the largest improvement in AIC are coloured in blue.
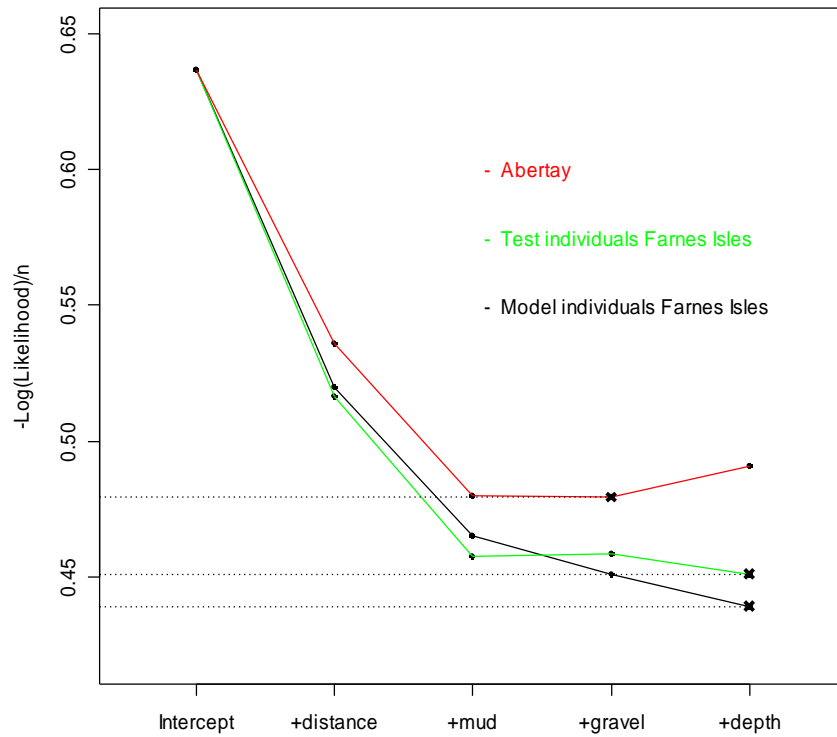
Fig. 3.6 Validation of models with only environmental variables, using test data from the same (Farnes Isles) and a different (Abertay) sub-population. Note that all parameter estimates used to calculate the likelihood are based on the 19 model-individuals from the Farnes Islands. Different models are arranged in order of increasing number of environmental variables along the x-axis. The y-axis represents the log-likelihood of the data under each model, standardised by the number of data points (n) in each of the three datasets. A black cross indicates the best model for each data set.
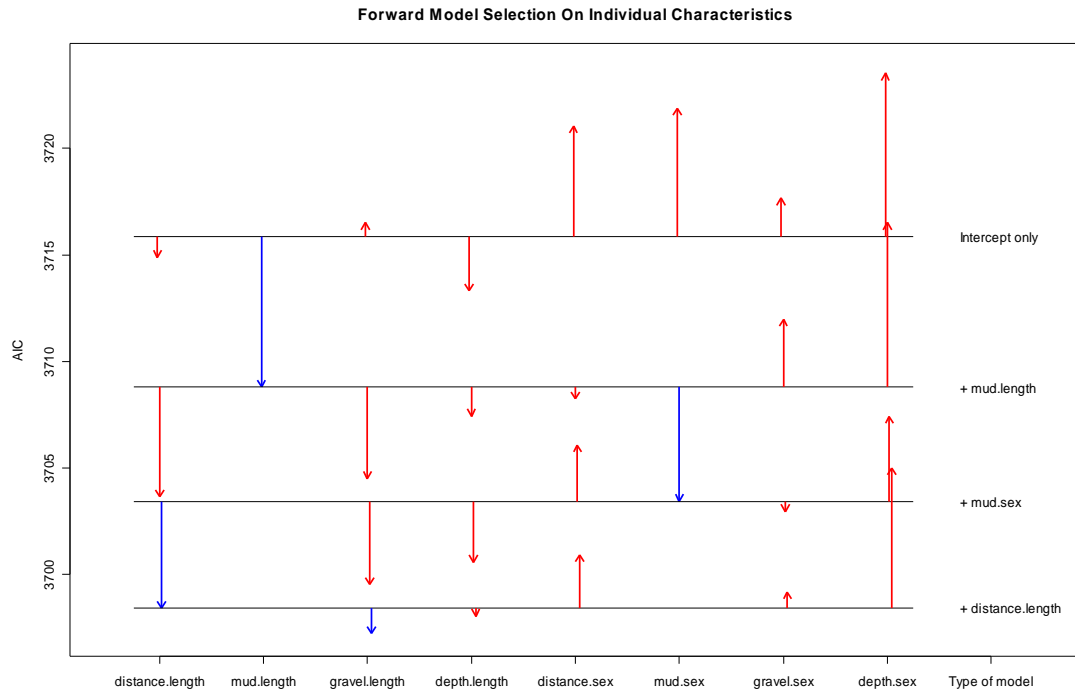
Fig. 3.7 Forward model selection on the interactions between individual characteristics (i.e. sex and nose-to-tail length) and environmental variables. The starting position is the best model arrived at by using environmental variables alone. In this case, this coincides with the saturated model (i.e. distance to the haulout, % mud, % gravel and depth) with splines for all variables except gravel. The arrows indicate the change in AIC from the current model (horizontal line) as a result of adding an interaction between an individual characteristic and environmental variable. Arrows for variables that lead to the largest improvement in AIC are coloured in blue.
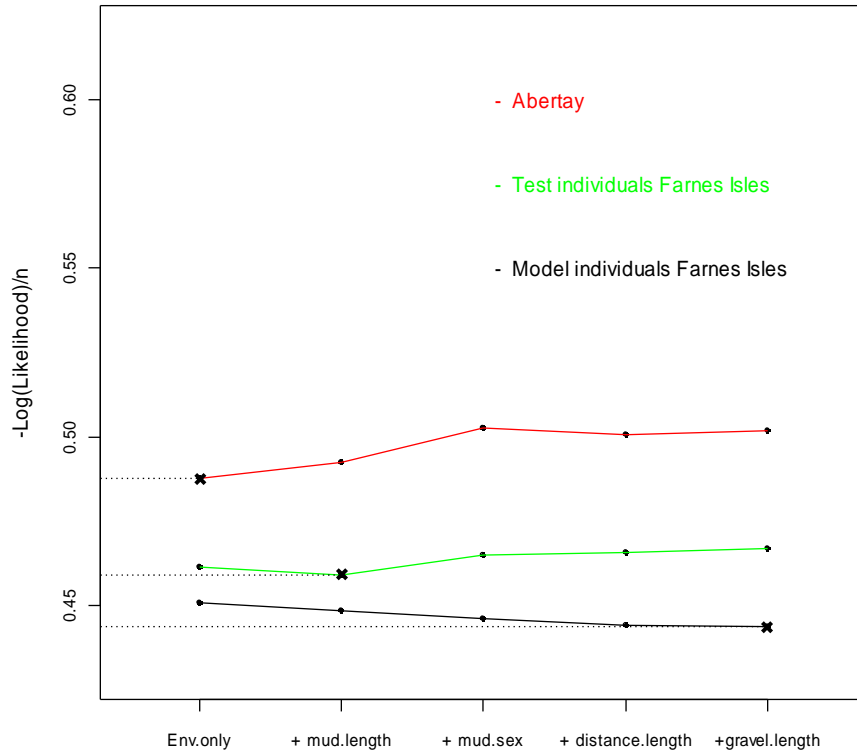
Fig. 3.8 Validation of models with added individual characteristics, using test data from the same (Farnes Isles) and a different (Abertay) sub-population. See also the legend of Fig 3.6. The *x*-axis indicate different models with increasing complexity (i.e. increasing number of interactions between individual characteristics and environmental variables) from left to right. The cross indicates the model with the lowest log(likelihood) for the corresponding data set.

It is possible that some important environmental variables were omitted from the analysis. This would reveal itself in the presence of spatial autocorrelation in the residuals (2.2.5). I constructed semi-variograms to investigate the spatial autocorrelation in the response and residuals for large (<40,000 m) and small (<10,000 m) spatial scales (Fig. 3.9)
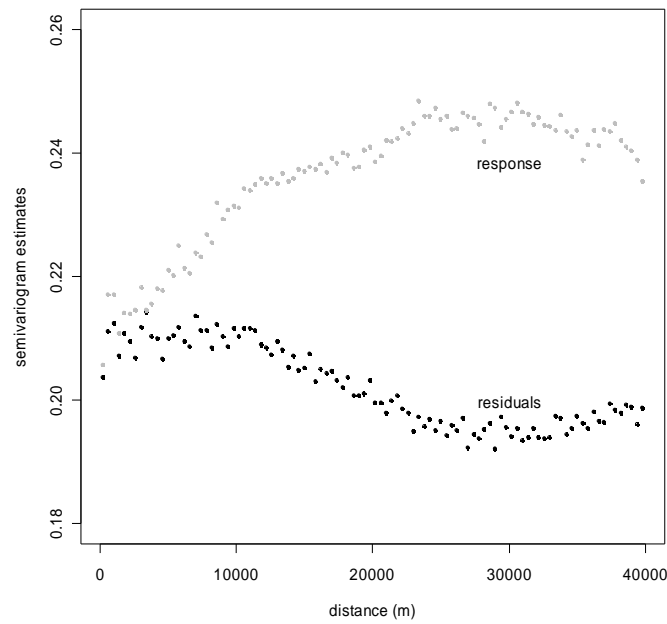
a.
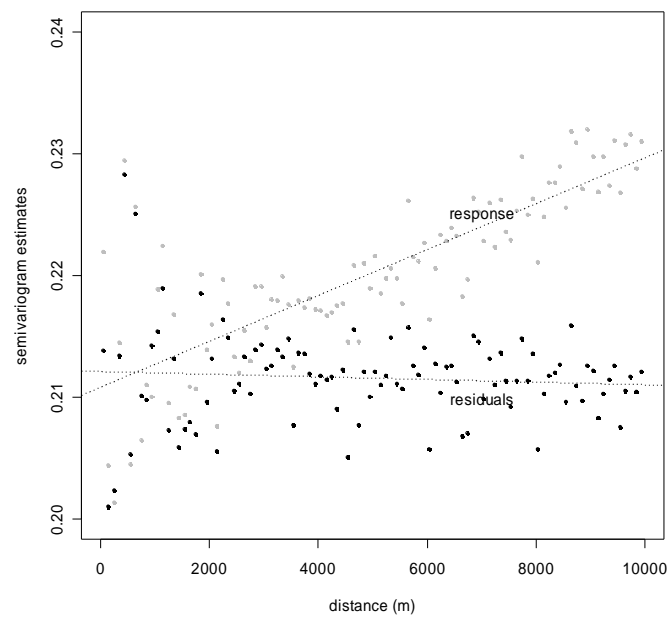
b.

Fig. 3.9 Semi-variogram of the response data (0 and 1) and the residuals (observed response - predicted) for up to 40km (**a**) and 10km (**b**) with the range of distance being partitioned into 100 distance classes in both cases. The plotted lines are linear regression models fitted to the data.

3.3.3 The effect of environmental variables on the response

The most important covariate is the distance to the haulout which has a negative relationship with the response, meaning that the observed usage far away from the haulout site is less than predicted by the accessibility model (Fig. 3.10a). The second most important environmental variable is mud and its interaction with the length of individuals. Fig. 3.10b shows that the animals' preference is highest for areas with mud content of approx. 3%. While small individuals change their response relatively little with increasing mud-contents (Fig. 3.10c), large animals have a well-defined peak in preference for those areas (Fig. 3.10d). Fig. 3.10e shows that grey seals also prefer gravely areas. Finally, the least important environmental variable depth, reveals that Grey seals have an increasing preference for areas up to 80 meters, after which the mean population preference decreases slightly. However, this pattern is confounded by large individual variability (Fig. 3.10f).
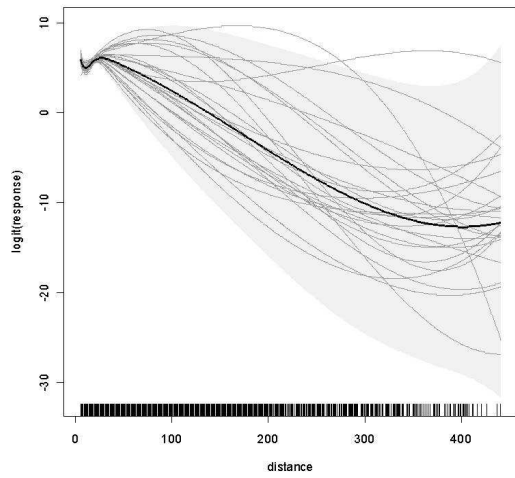
a.
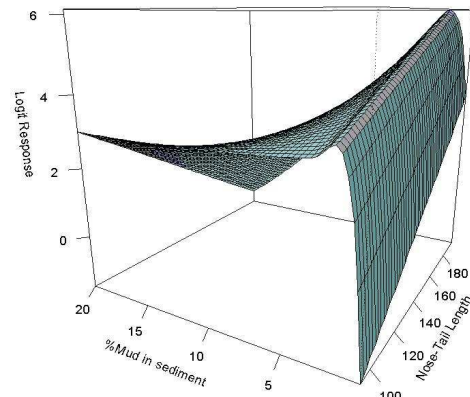
b.
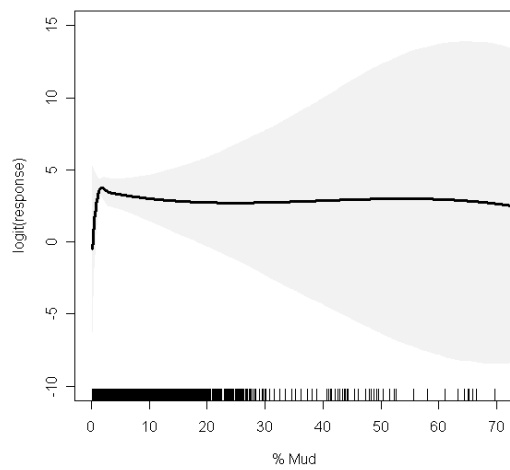
c.
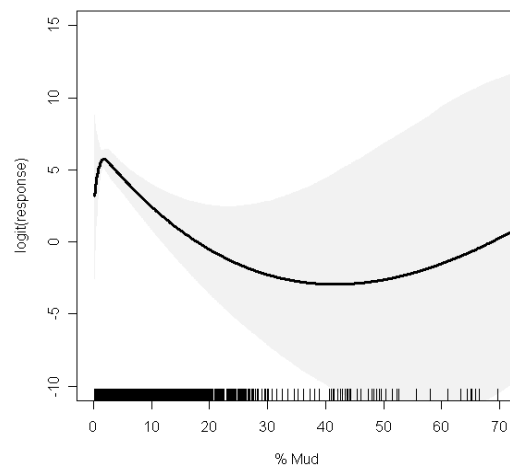
d.
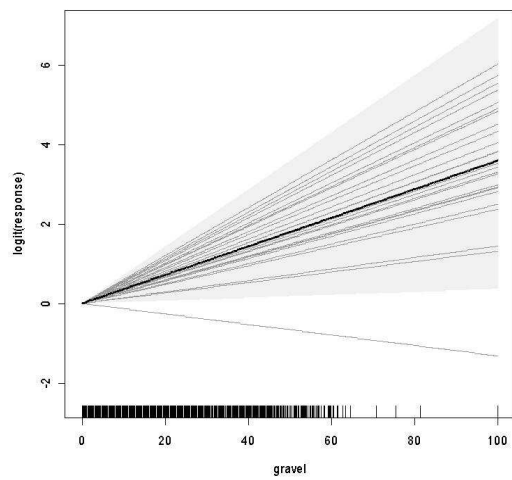
e.
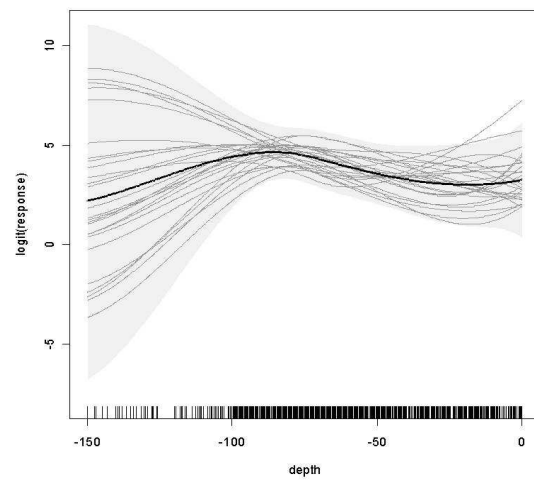
f.

Fig. 3.10 The effect of the covariates "distance to the haulout (km)" (a), the interaction between "%mud" and the individual characteristic "nose-to-tail length (cm)" (b), "%mud" for small (c) and large (d) individuals ("nose-to-tail length" is 100 and 180cm, respectively), "%gravel" (e) and "depth (m)" (f). In the plots for the single environmental variables (a, c, d, e and f) the 95% confidence limits are represented by the shaded areas, the grey lines (in a, e and f) represent individual specific responses and the black line the mean population responses (i.e. fixed effect).

3.3.4 Spatial prediction of usage and preference

Using the relations between the response and the environmental variables as outlined in the previous paragraph, I can estimate the rate of usage $v(\mathbf{s})$, for every point in space with known values for the environmental conditions (Fig. 3.11a). It is often difficult to interpret the absolute scale of a variable using colour plots alone. I therefore plotted one possible realization of a data set of observations that would be obtained from a heterogeneous spatial Poisson process with that rate (Fig. 3.11b). To indicate which areas Grey seals would use if all points in space were equally accessible, I plotted preference in space (Fig. 3.12). Finally, I used the model whose parameters were estimated using data from the Farnes, to make a spatial prediction of usage for seals making return trips from the Abertay (Fig. 3.13)
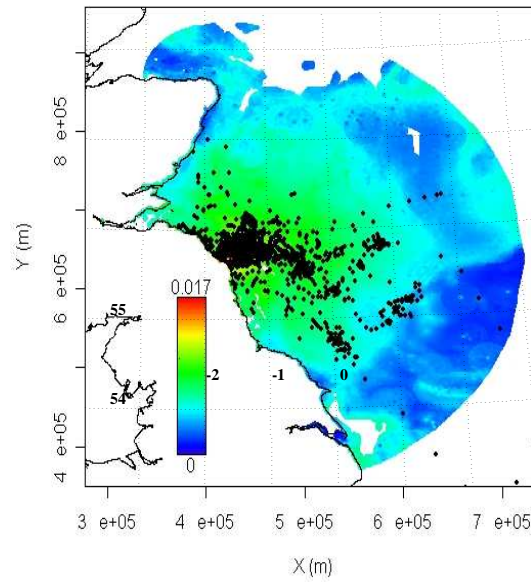
a.
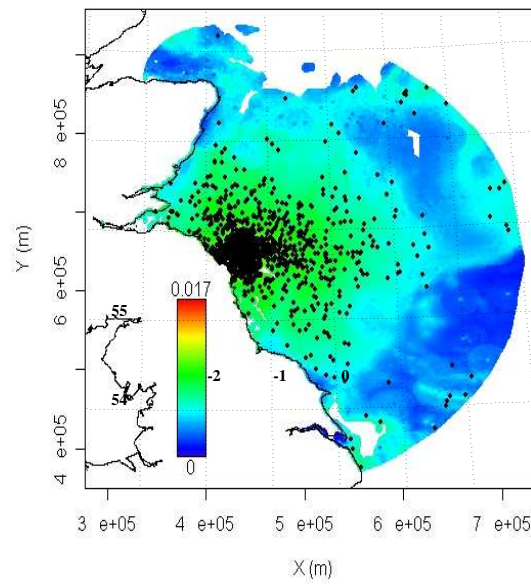
b.

Fig. 3.11 Spatial prediction of usage for the Farnes with the animal locations (**a**) and one possible realization for that same total number of locations using the usage predictions (**b**), plotted on top.
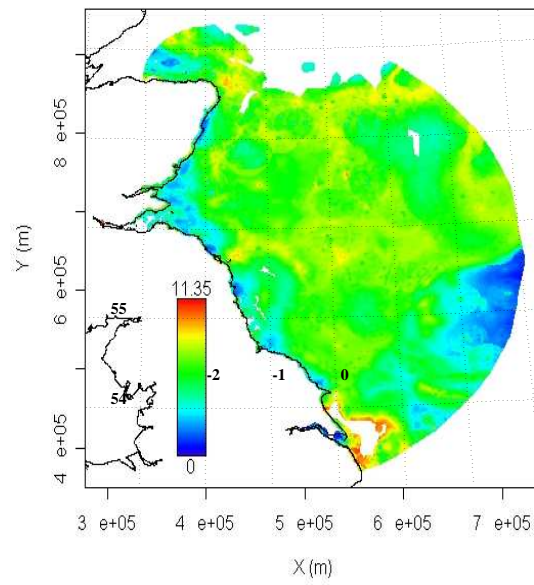
Fig. 3.12 Spatial prediction of preference corrected for unequal accessibility. Red indicates high preference and blue indicates low preference.
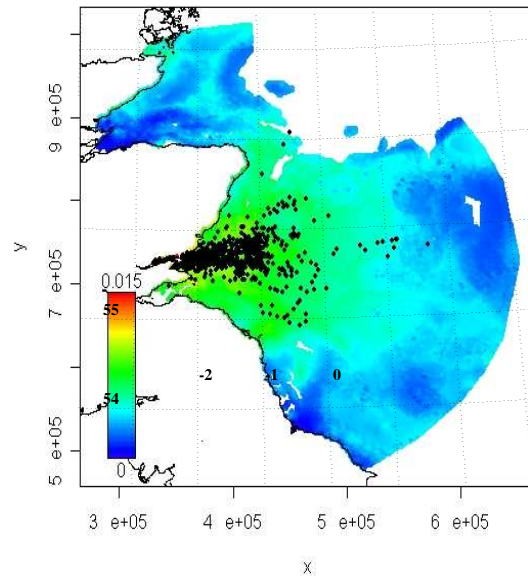
a.
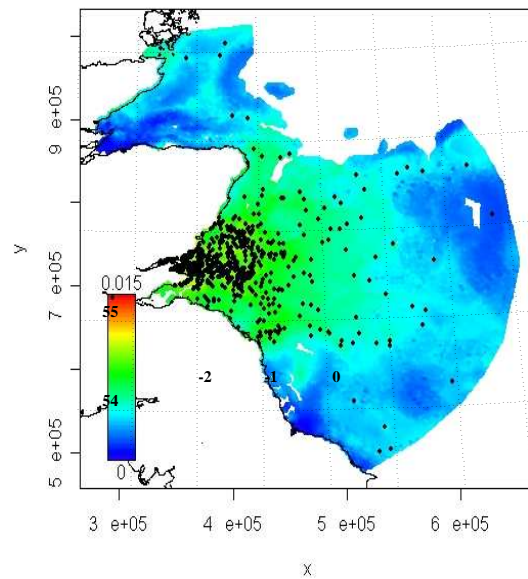
b.

Fig. 3.13 Spatial prediction of usage for Abertay using the Farnes Island model with the animal locations (**a**) and one possible realization for the same total number of locations using the usage predictions (**b**), plotted on top.

3.4 Discussion

Spatial ecologists generally try to understand where animals are, why they are there and where else they are likely to be. Matthiopoulos et al. (2004) focused on the first question for grey seals. Here, I extended that investigation to the other two questions. To achieve this, I relate their observations to prevailing environmental conditions and individual specific covariates and use these relationships to make predictions in space.

3.4.1 Habitat preference; why are Grey seals where they are?

All of the environmental variables examined in this case study helped to account for the variation in the response. Distance to the haulout and %mud were the most important. The negative relationship between the response variable h and distance indicates that grey seals concentrate their usage closer to the haulout than might be expected under the accessibility model. Avoidance of muddy areas could either be due to the fact that the seals' strategy to forage for sandeels by bio-turbation may not be effective in muddy substrates because of the rapid loss of visibility or due to the fact that sandeels also avoid muddy areas because it impairs the functionality of their gills (Wright et al. 2000, Holland et al. 2005).

Sea bottom depth might have been expected to be an important predictor of spatial usage because seals are predominantly benthic feeders whose useful foraging time at the sea bed decreases with increasing depths. However, the entire North Sea is relatively shallow and therefore unlikely to test the grey seals' diving capability. A second reason why depth might also have been expected to be important is its strong correlation with the density of sandeels. Wright et al. (2000) have shown that most sandeels around the Shetland Isles, in the North of Scotland, occur around depths of 50 to 60m. However, this relationship does not necessary hold across the North Sea. Also, if sandeels occurred in relatively deeper areas around the Farnes Isles, this might counter the additional cost of diving to such depths as described earlier.

3.4.2 Spatial predictions of usage and preference; Where else are grey seals likely to be?

The second major objective of this study was to make spatial predictions of usage (v). Those predictions can be classified into three categories in order of decreasing reliability: i) predictions within the area for which data is collected, ii) predictions outside the study area, but for similar environmental conditions as those used in constructing the model and iii) predictions outside environmental space.

In this study, I have shown that a model fitted to a subset of individuals, was able to capture the preferences and space-use of other individuals from that same sub-population. This carries the caveat that if patterns of space-use and preference change seasonally, then temporal variability in sampling effort will have biased the model's predictions towards the most intensively sampled months.

I have also shown that it is possible to predict the distribution of individuals from a different, albeit neighbouring, sub-population. This is particularly fortuitous because it has been known that differences in absolute availability of environmental conditions between areas can invalidate predictions of usage (Mysterud and Ims 1998). The similarity in environmental conditions facing the Farnes and Abertay sub-populations meant that the model's predictions were extrapolations in geographical, but not environmental space.

Extrapolations in environmental space are likely to be particularly unreliable. For example, the current data provide no hint as to the maximum depth that grey seals can dive to. Applied to the east coast of Scotland, this model would predict seal usage beyond the continental shelf.

3.4.3 Future research: additional environmental variables as candidate covariates

Generally there are two aspects that constitute a good candidate covariate: it must be relevant and knowable. A candidate covariate is relevant if it believed to have a strong relationship with the response variable. A covariate is knowable if it can be measured precisely in both space and time

or, alternatively, remains relatively constant. The main reason for choosing the environmental variables used in this study was the fact that they remain relatively constant. An alternative strategy would be to put more emphasis on the anticipated link between a candidate covariate and the response. Since grey seals go offshore to forage, their distribution is likely to be strongly related to fish distributions. Unfortunately, estimates of fish distributions are generally poorly known. A solution to this problem would be to first model fish distribution as a function of geophysical variables and then relate these modelled fish distribution with distribution with of grey seals.

3.4.4 Future research: Mechanistic modelling

In Chapter 2, I outlined the methodological progression from hypothesis testing to regression, GLMs, GAMs up to GAMMs. A similar process in biology where current knowledge on physiological, behavioural and ecological processes is used to form the basis for future research is less common and generally more difficult. This is especially true for studies on environmental preference, where e.g. different studies might include different environmental covariates, absolute availability of environmental conditions might be different between different study areas, individuals might behave differently at different times of the year and individuals and even sub-population might have different physiological restrictions and therefore show differences in environmental preference. Even if two sub-populations under study would behave identical, due to the stochastic nature of the data, fitting flexible regression functions to those data will generally lead to different functional forms. As a consequence it is extremely difficult to make proper comparisons between different studies presented in the literature.

An alternative approach is to use mechanistic models. For grey seals physiological models can be used to specify the cost of travelling to particular places and diving to specific depths. Dietary studies could be used to estimate the expected distribution of grey seals on the basis of the distribution of fish species. Alternatively grey seal distributions could be

related with spatial distributions of total nutrients on the basis of fish densities. Using such methods, one could more easily include prior knowledge and make more robust comparisons with other studies.

3.5 Rereferences

Argos (1989). Guide to the Argos System. Argos CLS, Toulouse.

Bonner, W.N. 1981. Grey seal. Handbook of Marine Mammals, Vol. 2 (eds S.H. Ridgeway & R.J. Harrison), pp. 111-144. Academic Press, London.

Bowen, W. D., D. Tully, D. J. Boness, B. M. Bulheier, and G. J. Marshall. 2002. Prey-dependent foraging tactics and prey profitability in a marine mammal. Marine Ecology-Progress Series **244**:235-245.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference. A practical Information-Theoretic Approach., 2nd edition edition. Springer-Verlag, New York.

Fletcher, R. 1987. Practical Methods of Optimization, 2nd ed. Wiley, Chichester.

Hammond, P. S., A. J. Hall, and J. Prime. 1994a. The diet of grey seals around Orkney and other island and mainland sites in northeastern Scotland. Journal of Applied Ecology **31**:340-350.

Hammond, P. S., A. J. Hall, and J. H. Prime. 1994b. The diet of grey seals in the Inner and Outer Hebrides. Journal of Applied Ecology **31**:737-746.

Holland, G. J., S. P. R. Greenstreet, I. M. Gibb, H. M. Fraser, and M. R. Robertson. 2005. Identifying sandeel Ammodytes marinus sediment habitat preferences in the marine environment. Marine Ecology-Progress Series **303**:269-282.

Isaaks, E. H., and R. M. Srivastava. 1990. Applied goestatistics. Oxford university press, Oxford.

Matthiopoulos, J. 2003. The use of space by animals as a function of accessibility and preference. Ecological Modelling **159**:239-268.

Matthiopoulos, J., B. McConnell, C. Duck, and M. Fedak. 2004. Using satellite telemetry and aerial counts to estimate space use by grey seals around the British Isles. Journal of Applied Ecology **41**:476-491.

McConnell, B. J., C. Chambers, K. S. Nicholas, and M. A. Fedak. 1992. Satellite Tracking of Gray Seals (Halichoerus-Grypus). Journal of Zoology **226**:271-282.

McConnell, B. J., M. A. Fedak, P. Lovell, and P. S. Hammond. 1999. Movements and foraging areas of grey seals in the North Sea. Journal of Applied Ecology **36**:573-590.

Mysterud, A., and R. A. Ims. 1998. Functional responses in habitat use: Availability influences relative use in trade-off situations. Ecology **79**:1435-1441.

Ollason, J. G. 1980. Learning to forage - Optimally? Theoret. Pop. Biol. **18**:44-56.

Otter Research Ltd. 2004. Random effects in AD Model Builder. ADMB-RE user guide. Otter Research Ltd, Sidney B.C.

Pinheiro, J. C., and D. M. Bates. 2000. Mixed-effects models in S and S-plus. Springer-Verlag, New York.

Prentice, R. L., and R. Pyke. 1979. Logistic Disease Incidence models and case-control studies. Biometrika **66**:403-411.

R Development Core Team 2004. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, URL http://www.R-project.org.

Ramsay. 2005. Functional Data Analysis. Springer, New York

Skaug, H. J., 2002. Automatic Differentiation to Facilitate Maximum Likelihood Estimation in Nonlinear Random Effects Models. Journal of Computational & Graphical Statistics 11: 458-470

Skaug, H., and D. Fournier. 2003. Evaluating the Laplace approximation by automatic differentiation in nonlinear hierarchical models, Unpublished manuscript: Inst. of Marine Research, Box 1870 Nordnes, 5817 Bergen, Norway.

SMRU 2005. Scientific Advice on Matters Related to the Management of Seal Populations: 2005. SCOS main advice report.

Thompson, D., P. S. Hammond, K. S. Nicholas, and M. A. Fedak. 1991. Movements, diving and foraging behaviour of grey seals {IHalichoerus grypus}. Journal of Zoology **224**:223-232.

Thompson, P. M., B. J. McConnell, D. J. Tollit, A. Mackay, C. Hunter, and P. A. Racey. 1996. Comparative distribution, movements and diet of harbour and grey seals from the Moray Firth, NE Scotland. Journal of Applied Ecology **33**:1572-1584.

Thompson, P. M., D. Tollit, A. Mackay, H. Corpe, K. Grellier, and B. Greig. 1995. The behaviour and ecology of common and grey seals in the Moray Firth.

Vincent, C., B. J. McConnell, V. Ridoux, and M. A. Fedak. 2002. Assessment of Argos location accuracy from satellite tags deployed on captive gray seals. Marine Mammal Science **18**:156-166.

Wood, S. N. 2001. mgcv: GAMs and Generalized Ridge Regression for R. R News **Vol 1**:20-25.

Wood, S. N. 2006. Generalized Additive Models: an introduction with R. in press.

Wright, P. J., H. Jensen, and I. Tuck. 2000. The influence of sediment type on the distribution of the lesser sandeel, Ammodytes marinus. Journal of Sea Research **44**:243-256.

Zhang, D. W., and M. Davidian. 2004. Likelihood and conditional likelihood inference for generalized additive mixed models for clustered data. Journal of Multivariate Analysis **91**:90-106.

4. MODELLING SPACE-USE AND HABITAT PREFERENCE IN NORTHERN GANNETS (MORUS BASSANUS)

Animals need food resources to meet their requirements for survival, growth and reproduction (Manly et al. 1993). The distribution of food in the marine environment varies considerably in both space and time and is therefore hard to predict (Barnes and Hughes 1999). Correctly gauging the spatiotemporal availability of food is essential for seabirds because they have to contend with the high energetic costs of flight while foraging. This is particularly true for Northern Gannets, the largest pelagic seabirds breeding in the north Atlantic, whose size brings a high mass-specific expenditure during flight (Birt-Friesen et al. 1989). During the breeding season, when adults also need to provision for their chick, an efficient foraging strategy is even more important.

Burger (1980) postulated that the delayed sexual maturation observed in gannets may be due to the lack of experience of young adults in efficiently provisioning for offspring. If this were true, it would indicate that foraging efficiency increases with age and experience. Although this could be due to an increase in the birds' skill at catching prey, it could equally be due to an increase in their ability to locate it, for example by more successfully reading environmental cues. The first objective of this chapter is to investigate if space use in gannets correlates with static environmental variables or changing visual cues. Its second objective is to use this information to make

predictions of space use to be employed in the management and conservation of the species.

4.1 Gannet natural history

The largest population of northern Gannets is on Bass Rock in the Frith of forth, SE Scotland (56˚6'N, 2˚36'W). The colony contains more than 70,000 breeders and several thousand non-breeders and is still increasing at ~5% per year. From their breeding site, chick-rearing adults make trips to sea lasting, on average, 32.2 h (SD = 13.2 h). The furthest points of these trip are an average 232 km (SD = 100 km) from the breeding site (Hamer et al. 2000).

During these trips in 1998, Gannets from Bass rock predominantly fed on Mackerel (*Scomber scrombrus* - 30.8 % of biomass in regurgitates), Herring (*Clupea harengus* - 20.3%), Sandeel (*Ammodytes marinus* - 17.9%), Sprat (*Sprattus sprattus* - 12.6%) and Gadidae (16.4%) such as haddock (*Melangogrammus aeglefinus*), whiting (*Melangogrammus merlangus*) and cod (*Gadus morhua*) (Hamer et al. 2000).

4.2 Methods

4.2.1 Response variable

In 1998, 17 chick-rearing adults were caught on Bass Rock (56˚6'N, 2˚36'W), SE Scotland (Hamer et al. 2000, Hamer et al. 2001). Only individuals with hatching dates ± 2 wk from the modes were captured and Platform Terminal Transmitters (PTTs) were attached to each bird for a duration of 14 to 23 days (average 16 days). Similar to the SRDLs used in the grey seal case study (see chapter 3 and 5), the PTTs also rely on the Argos satellite system. The set of animal telemetry locations contained a small number of highly erroneous location fixes which were removed by applying the filtering algorithm described by McConnell et al. (1992), using a maximum speed parameter of 22.5 m/s (Hamer et al. 2000), well above the average flight speed of 15.3 m/s (Pennycuick 1987). Some remaining error in the data was

treated using a smoothing algorithm developed (M. Lonergan unpublished) within the MGCV (Wood 2001) package in R (see §3.2.1). I used the output of this algorithm to interpolate the animals' positions at 3-hourly intervals. Obtaining temporally regular estimates of the animal's position also helped to reduce biases due to the variable observability of the animals (§ 2.2.7). Because I was primarily interested in foraging behaviour I restricted my attention to those locations that were outwith 5km distance of their breeding site.

The error-corrected and temporally interpolated data set of satellite locations represented a sample of points in geographical space visited by the tagged gannets (presence data). Under the case-control paradigm, to provide the contrast necessary for modelling preference, these had to be complemented with a set of points not visited by the study animals (absence data). The Bernoulli response variable in eq. 2.5 took the value 1 wherever there was an animal present and 0 where there wasn't.

To account for the unequal accessibility of points in space, the absence points can be drawn from a spatial surface of accessibility (see also §2.2.2) which is either based on an individual movement (Matthiopoulos 2003) or a diffusion model. Alternatively, if the necessary parameters of movement are not readily available as is the case here, absences can be selected uniformly randomly from space and the distance to the central-place included as an explanatory variable. This uniform selection of absences results in areas close to the breeding site being as well-sampled as the more remote, less accessible ones. This means that a large number of absences are required in total to represent fine-scale patterns of usage close to the central-place, leading to insurmountable computational problems with the full model. To bypass these problems, as a prequel to the analysis, I set up a model using uniformly selected absences and distance from the breeding site as the only covariate of usage. I then used the predictions of this model as the accessibility surface from which the absence data for the full model were randomly selected.

4.2.2 Explanatory variables

For each set of spatial coordinates contributing to the response data, I provided values for the explanatory variables corresponding to conditions at these coordinates. I selected environmental variables on the basis of possible biological relevance and availability of data. I opted for three static variables: sediment type, sea depth, distance to coast and distance to the breeding site and two non-static variables that are visible from the air: chlorophyll-a concentration and water opaqueness, also known as diffuse attenuation.

Gannets might have a preference for certain sediment types as an indirect consequence of the habitat preference of sediment-burrowing prey such as sandeels (Wright et al. 2000, Holland et al. 2005). Sediment type is also a reflection of prevailing current conditions (e.g. high mud content in the sediment reflects slow currents). Strong currents imply a large net influx of zooplankton on which fish can feed but they also increase the energetic cost of mobility for fish. I used the same GIS layers for sediment generated for the grey seal analysis (§3.2.2).

The relationship between the spatial distribution of gannets and sea depth is difficult to anticipate. Shallow areas among deeper areas will generally be characterised by stronger currents and therefore higher net influx of plankton. On the other hand, these regions might be a more exposed to breeding seabirds and hence characterised by higher risk. Deeper areas might also offer a wider variety of currents within the water column giving the fish more options from which to chose in trying to improve their foraging success. I used the GIS layer for bathymetry derived for the grey seal analysis (§3.2.2).

Because diving gannets catch prey using visual cues at depths of up to 34 meters (Brierley & Fernandes 2001), the clarity of water will improve the probability of catching a fish conditional on it being there. On the other hand, clear water might be the result of low plankton concentration and therefore indicate a low probability of fish being there. The variable "diffuse attenuation" was used as an indicator of water clarity, while chlorophyll-a concentration is an indicator of primary productivity. Both variables were derived from weekly composite of Seawifs satellite images and measured at

approximately 9km spatial resolution. Due to cloud cover, data did not exist for all cells. However with the aid of spatial and temporal correlation in the data, these missing values could be interpolated in both space and time from neighbouring cells. To do this, I fitted a GAM with a smooth interaction between latitude and longitude and the date at which the data was collected, by means of tensor product smoothers available in the MGCV package in R (Wood 2001). The tensor product function in MGCV works by first constructing a set of k basis spline functions for each variable in the model (in this example, latitude, longitude and time). It then calculates the products of all possible combinations of basis functions (in this case, leading to a total of k^3 interaction terms). I used the resulting model to make predictions for every cell for which data was missing due to cloud cover. The effect of tensor product smoothing is clearly visible in Fig. 4.3 b and e. The Pixel Intensity (PI) was stored using values 1 to 255, but could be converted into absolute measures of chlorophyll-a concentration (CHLO) and diffuse attenuation (K490) using the following functions

$$CHLO = 10^{-2+0.01 \cdot PI_{CHLO}} \qquad\qquad 4.1$$

$$K490 = 10^{-2+0.011 \cdot PI_{K490}} \qquad\qquad 4.2$$

Finally, there might be several reasons why distance to coast is important to the gannets. They might avoid coastal areas as a result of inter-specific competition with short-ranging species or prefer them because of their higher productivity. They might also use the coast line as an orientation aid. Distance from the gannet's breeding site was included to account for residual issues of accessibility.

4.2.3 Model structure

The basic structure of the model follows from using a binomial response variable (h) with a logit link

$$\hat{u}_l \sim B(1, h_l)$$

$$h_l = g^{-1}(\eta_l) = \frac{e^{\eta_l}}{1 + e^{\eta_l}}$$

4.3

To make the model flexible (§2.2.8), the linear predictor (η_l) follows the structure of a GAM for a total of six explanatory variables

$$\eta_l = b_0 + s_1(\text{distance}_l) + s_2(\text{depth}_l) + s_3(\text{mud}_l) + s_4(\text{gravel}_l) +$$
$$s_5(\text{diffuse attenuation}_l) + s_6(\text{chlorophyll}_l)$$
$$= b_0 + \sum_{k=1}^{6} b_{1,k} f_{1,k}(\text{distance}_l) + \cdots$$

4.4

where each spline $s(\cdot)$ is a flexible function of each explanatory variable, and $b_{j,k}$ is the random effect parameter for the kth out of 6 basis function of the spline for the jth explanatory variable. Construction of splines followed the protocol developed for grey seals (§3.2.3).

In the most saturated form of the model, each parameter, was expressed as a function of individual characteristics

$$b_{j,k,m} = \beta_{j,k,0} + \beta_{j,k,2} sex_m + \beta_{j,k,3} mass_m + v_{j,k}$$

4.5

where m refers to a specific individual and the individual characteristic mass was recorded at re-capture.

4.2.4 Parameter estimation

The parameters that needed to be estimated were the model's coefficients β_j and the variances σ_j^2 of the variance-covariance matrix ψ, of which the covariances were assumed to be 0. Parameter estimation was done using maximum likelihood methods which are described in detail in §3.2.4.

4.2.5 Model selection and model validation

I adhered to the model selection protocol derived for grey seals. I used data from 11 individuals for model fitting and the remaining 6 individuals for model selection.

4.2.6 Spatial prediction of usage

I used the same rationale behind the grey seal usage predictions. Because the total population size of breeding adults on Bass Rock was known and the proportion of time spent foraging outside the 5km zone ($p_{\text{at sea}}$) could be estimated from the telemetry data, I was able to estimate the at-sea density of gannets within a spatial unit *c* (e.g. a 1 km cell) as follows

$$u_c = N p_{\text{at sea}} \int_{\mathbf{x} \in c} \upsilon(\mathbf{s}) d\mathbf{x} \qquad 4.6$$

4.2.7 Spatial prediction of preference

As in the grey seal study, predictions of preference were generated from the relationship

$$\frac{\upsilon(\mathbf{s})}{\alpha(\mathbf{s}) \exp(\eta_{\text{distance}})} \propto \exp(\eta_{\text{environmental variables}}) \qquad 4.7$$

4.2.8 Software

All data manipulation, analyses and plotting (including spatial plotting) was done using R (R Development Core Team 2004). ML parameter estimation was done using ADMB-RE (Otter Research Ltd. 2004a, b). Transformation of satellite images in .hdf format to ASCII raster was done using Manifold.

4.3 Results

4.3.1 Response and environmental variables

A total of 17 individuals were caught, equipped with satellite transmitters and released within a 45 day period (11 July to 25 August). Following track filtering the raw data yielded a total of 1293 locations. I augmented these with twice as many (i.e. 2586) points selected uniformly randomly from space for which environmental data is available (Fig. 4.1a). I then extracted the relationship between the response variable presence/absence and distance. This model was used to calculate the rate of usage $v(\mathbf{s})$ at all points in space which was than treated as the accessibility surface from which new absences were extracted (Fig. 4.1b) for the full model.

The combined presence and absence data constituted the response data that were to be regressed against the environmental variables (examples shown in Figs 4.1b, 4.2 & 4.3).

a.   b.
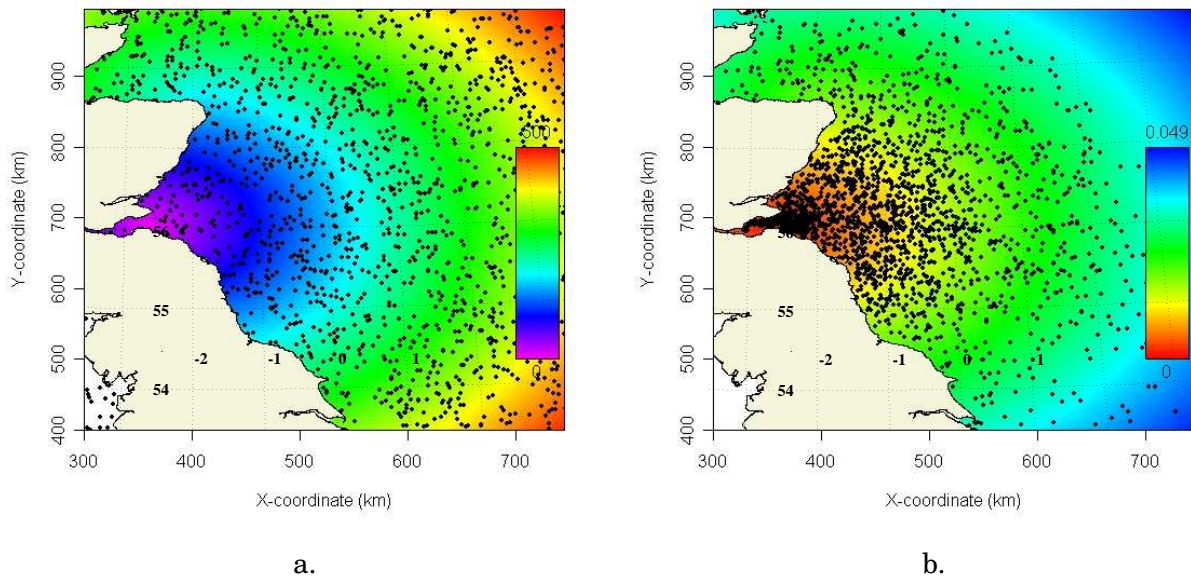
Fig. 4.1 Distance to the breeding site and absences drawn uniformly from within the study area (**a**). Surface of accessibility and absences drawn uniformly from it (**b**).

a.

b.

c.
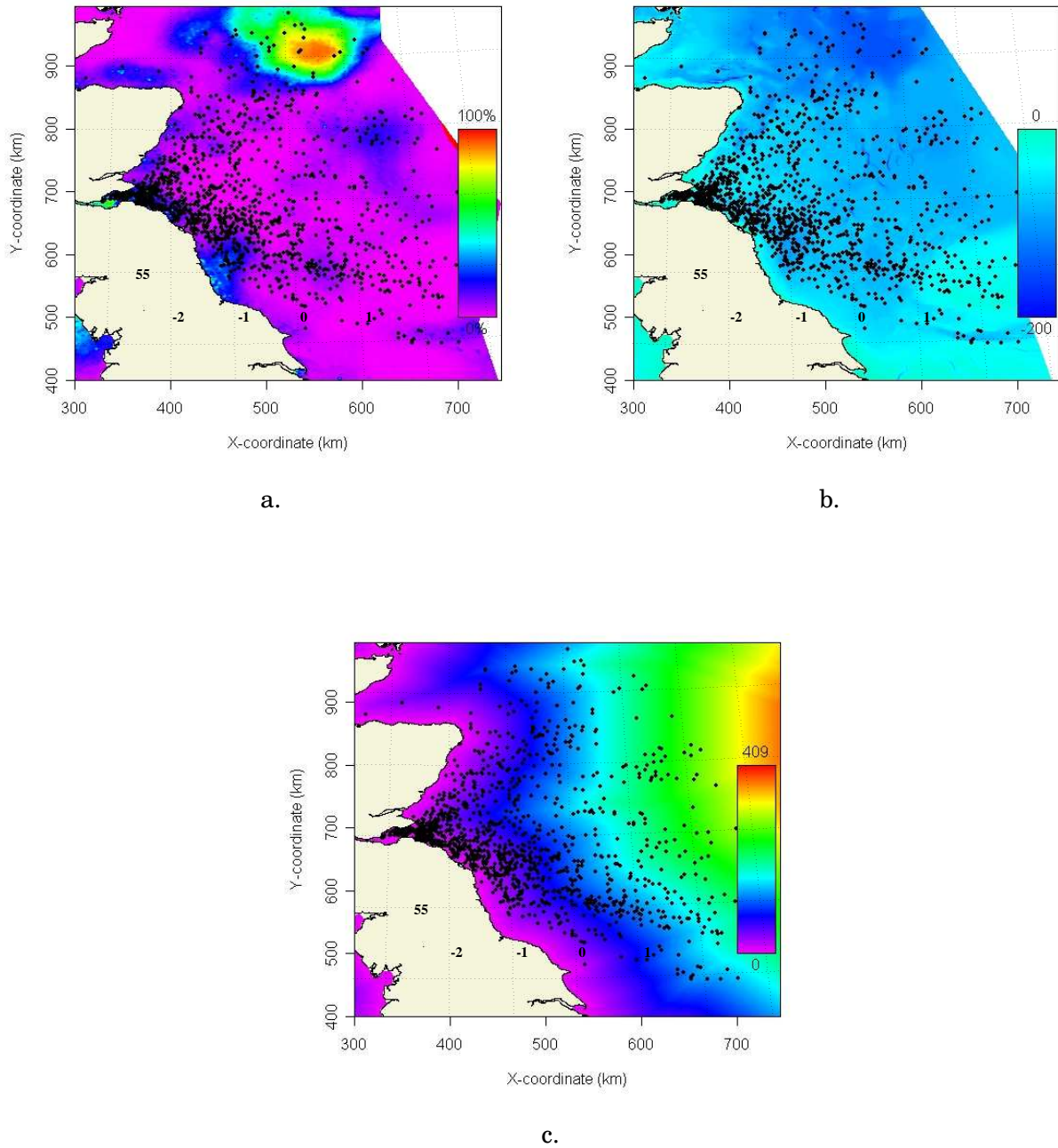
Fig. 4.2 Spatial representation of % mud in the sediment (**a**), depth (**b**) and distance to the coast line (**c**).
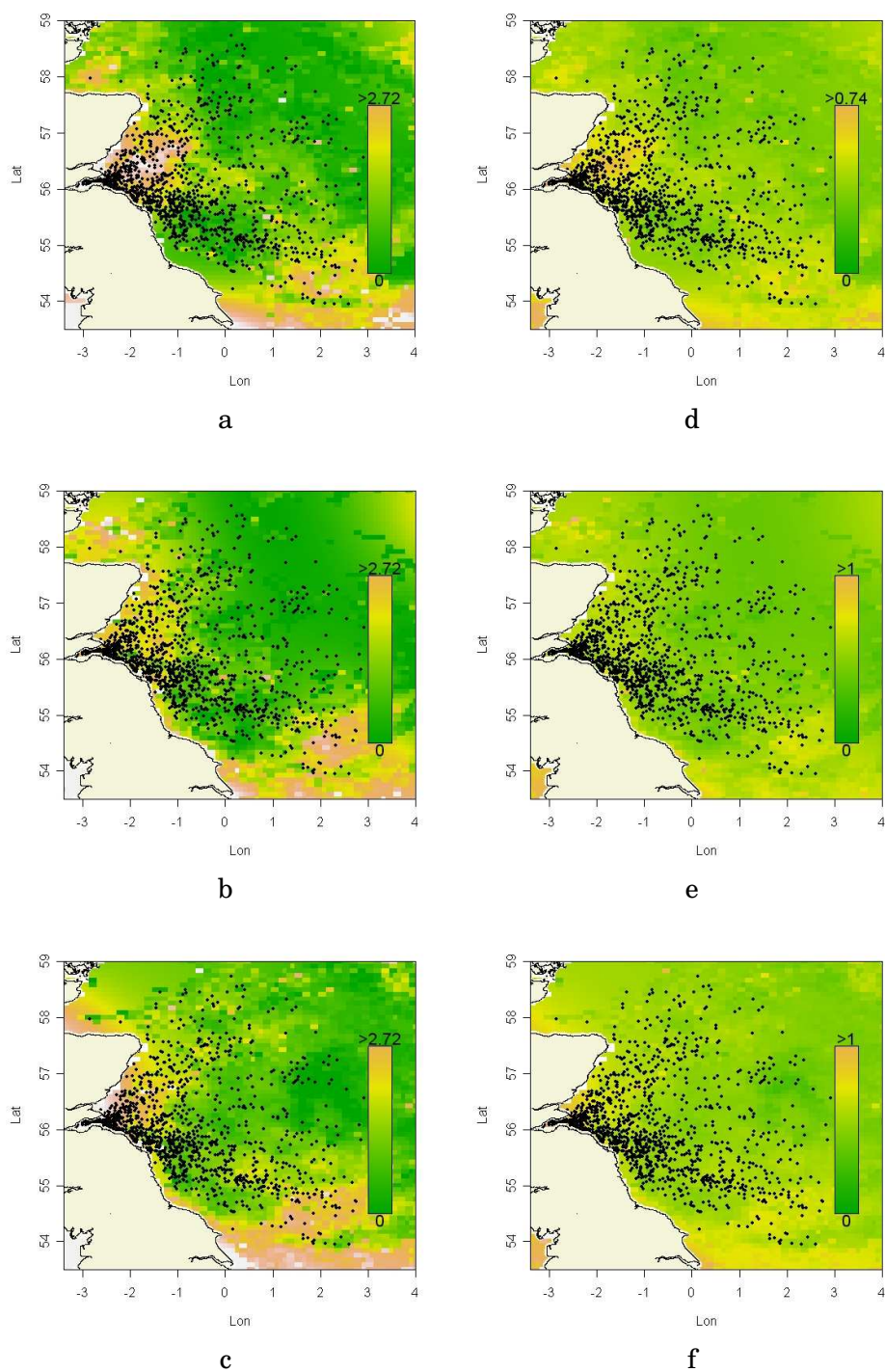
Fig. 4.3 Satellite images of Chlorophyll concentrations (**a**, **b** & **c**) and diffuse attenuation (**d**, **e** & **f**) for three weeks each starting with the on day 193 (**a** & **d**), 209 (**b** & **e**) and 225 (**c** & **f**) of 1998.

4.3.2 Model selection

Forward model selection suggested that a model containing all environmental variables explained a significant proportion of the variation in the response (Figs 4.4 & 4.5), but this was only 7%. Due to computational restrictions I was limited to fitting models containing up to 5 (instead of 7) covariates. Validation of the candidate models generated by this process (Fig. 4.4) indicated that a simpler model (see Appendix B, table 2 for the parameter estimates) containing distance to the coast, depth and distance to the breeding site best explained the response data from the validation individuals. This model was used for further analysis.

It was conceivable that this model was too flexible due to the use of splines. The sequential replacement of splines by the un-transformed environmental variables led to a deterioration of the model's AIC under both the fitting and test data sets.

To explain some of the residual variation I also included individual characteristics (eq 4.5). Based on the AIC, an interaction between depth and gannet mass led to a slightly improved model (2683.2 compared to 2683.6). However, under the validation data set the log-likelihood of the model decreased from -819 to -832. Therefore, I used the model with no individual-specific covariates to investigate the relationship between the response and the different environmental variables and to make spatial predictions of both usage and preference.
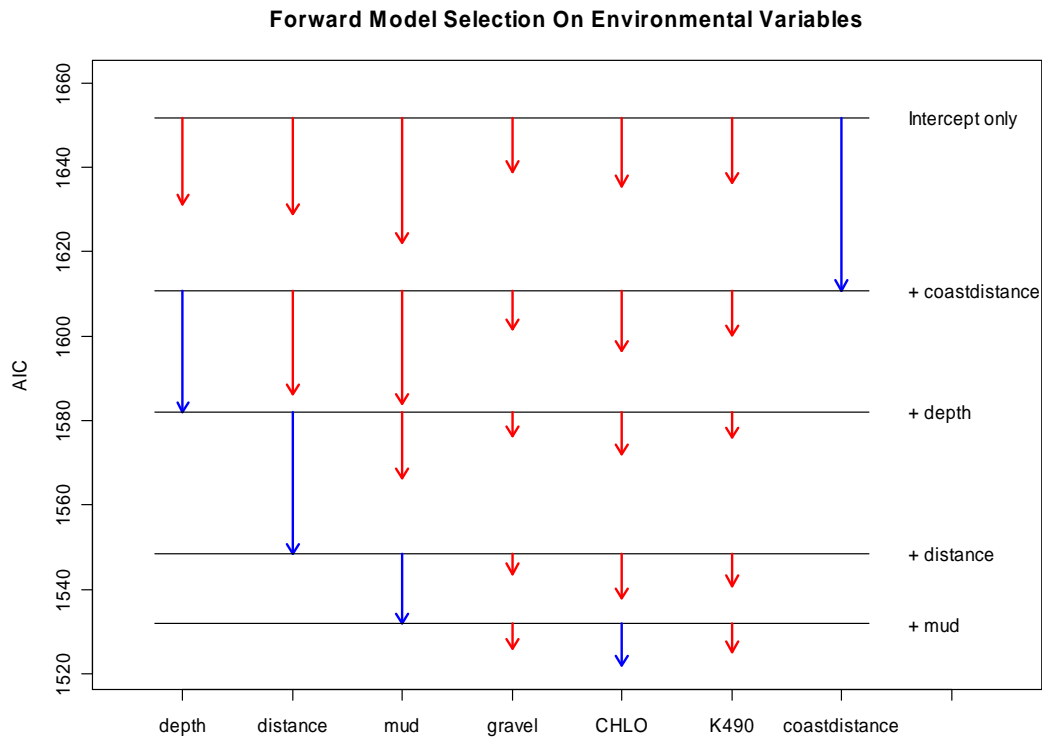
Fig 4.4 Forward model selection on the environmental variables; depth, distance to the breeding site, percentage mud and gravel in the sediment, chlorophyll concentration, diffuse attenuation and distance to the coastline denoted by "depth", "distance, "mud", "gravel", "CHLO", "K490" and "coast distance". The arrows indicate the change in AIC between models (horizontal lines) as a result of adding an environmental variable. Arrows for variables that lead to the largest improvement in AIC are coloured in blue.
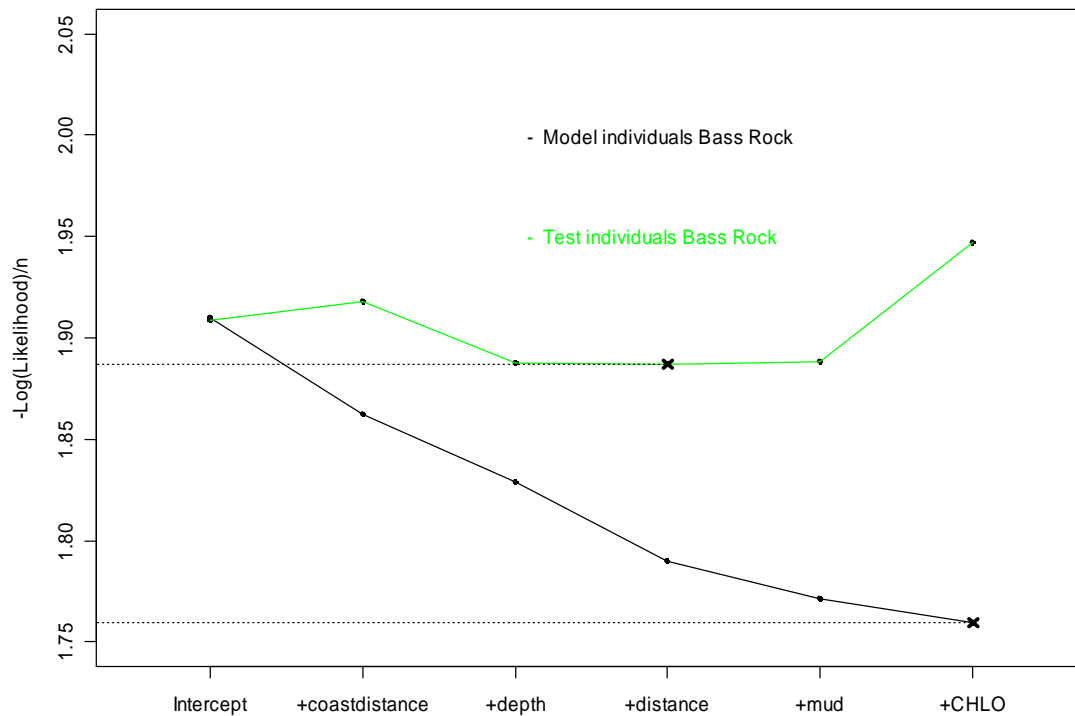
Fig 4.5 Validation of models containing only environmental variables using the test data set. Note that all parameter estimates used to calculate the likelihood are based on the 11 individuals from Bass Rock. Different models are arranged in order of increasing number of environmental variables along the x-axis. The y-axis represents the log-likelihood of the data under each model, standardised by the number of data points (n) in each of the three datasets. The black crosses indicate the best model for each data set.

It is possible that some important environmental variables were omitted from the analysis. This would reveal itself in the presence of spatial autocorrelation in the residuals (§ 2.2.5). I constructed semi-variograms to investigate the spatial autocorrelation in the response and residuals for large (<40,000 m) and small (<5,000 m) spatial scales (Fig. 4.6). These indicated that spatial autocorrelation is present only on large spatial scales, although some some of it is accounted for by the covariates.

a.   b.
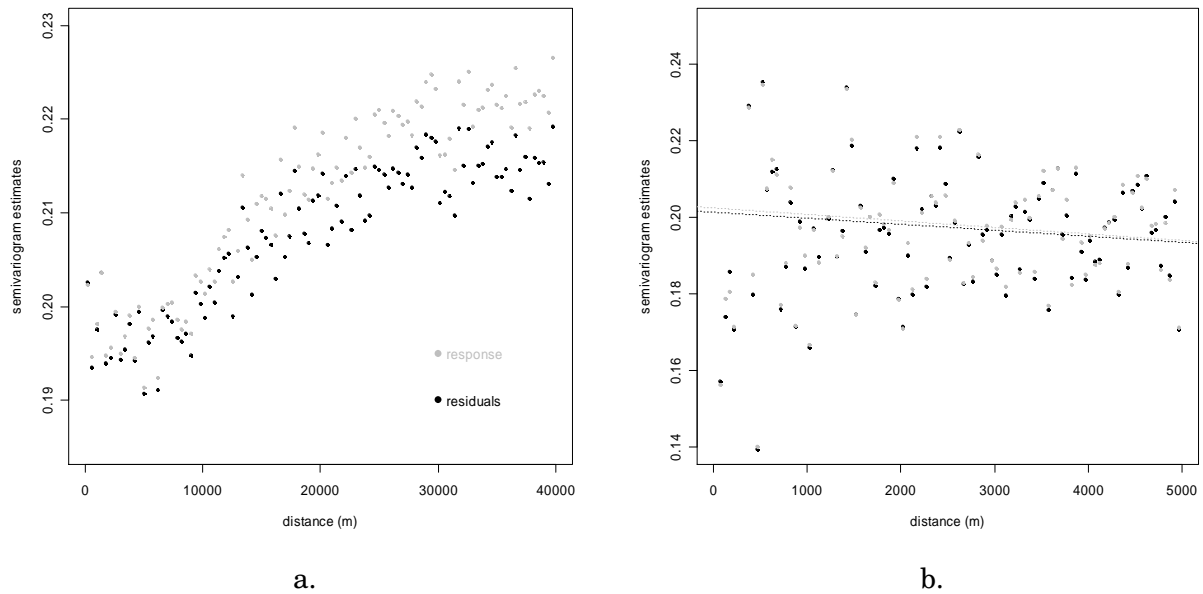
Fig 4.6 Semi-variogram of the response data (0 and 1) and the residuals (observed response - predicted) for up to 40km (**a**) and 5km (**b**) with the range of distance being partitioned into 100 distance classes in both cases. The plotted lines are linear regression models fitted to the data. This trend is not significant (p-value = 0.41).

4.3.3 The effect of environmental variables on the response.

The most important covariate was the distance to the coast which had a negative relationship with the response for distances up to 20 km, meaning that the observed usage away from the coast is less than predicted by the accessibility model (Fig. 4.7a). The second most important environmental

variable was depth. Fig. 4.7b shows that the animals' preference is highest for areas with a depth of about 80 meters. Finally Fig. 4.7c shows that gannets generally prefer areas that are closer to the breeding site than predicted by the accessibility model. This could be due to the fact that the knot positions of the splines used for this variable are different from those used to estimate the accessibility surface.

a.

b.

c.
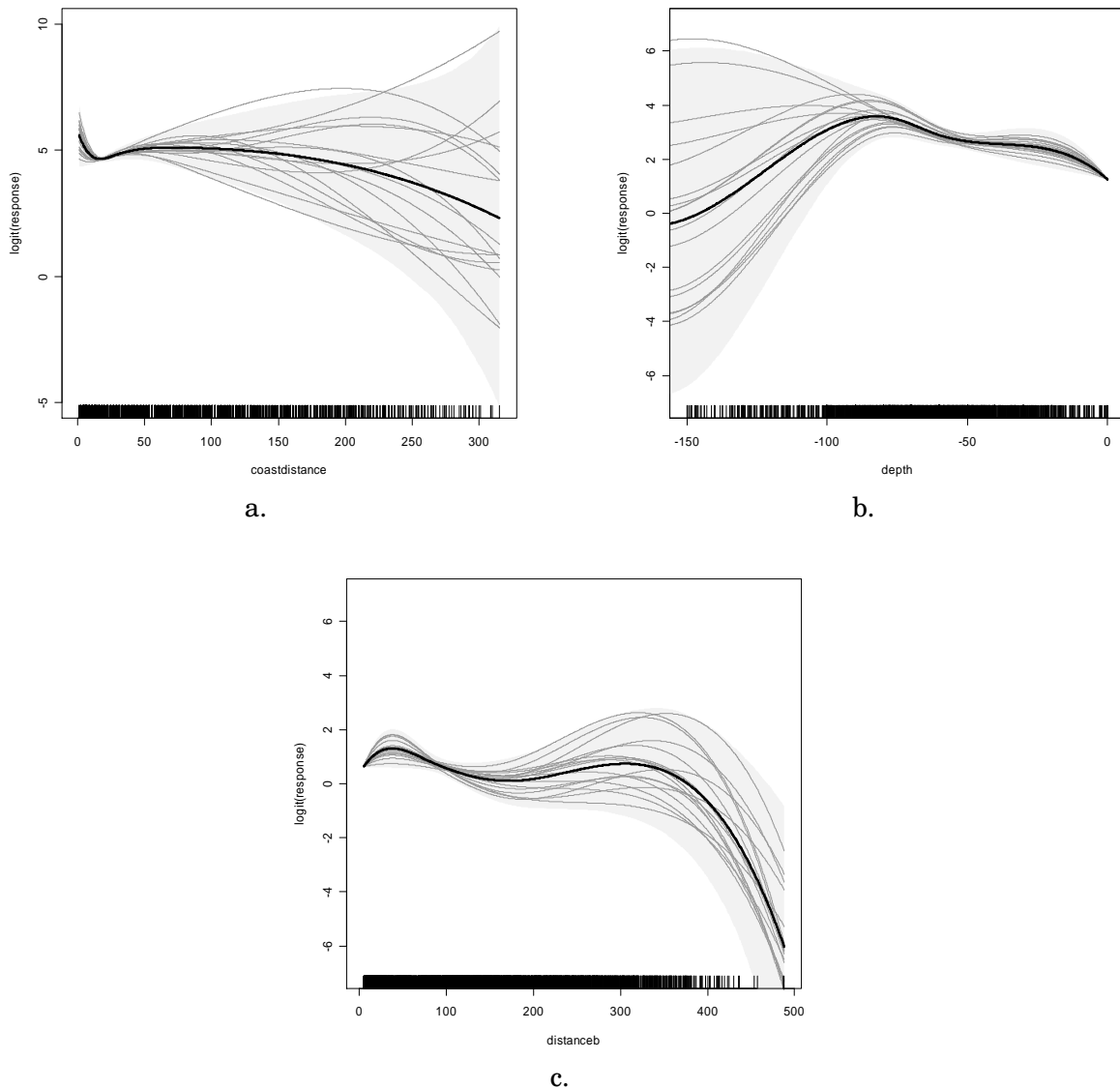
Fig. 4.7 The effect of the covariates distance to the coast-line (**a**), depth (**b**) and distance to the breeding site (**c**). The 95% confidence limits are represented by the shaded areas , the grey lines represent individual specific responses and the black line the mean population responses (i.e. fixed effect).

4.3.4 Spatial prediction of usage and preference

Using the relations between the response and the environmental variables as outlined in the previous paragraph, I can estimate the rate of usage (v), for every point in space with known values for the environmental conditions (Fig. 4.8a and b). It is often difficult to interpret the absolute scale of a variable using colour plots alone. I therefore plotted one possible realization of a data set of observations that would be obtained from a heterogeneous spatial Poisson process with that rate (Fig. 4.8c). This realization was based on a total adult population size of 70,000 and a proportion spent at sea of 50%.

a.

b.

c.
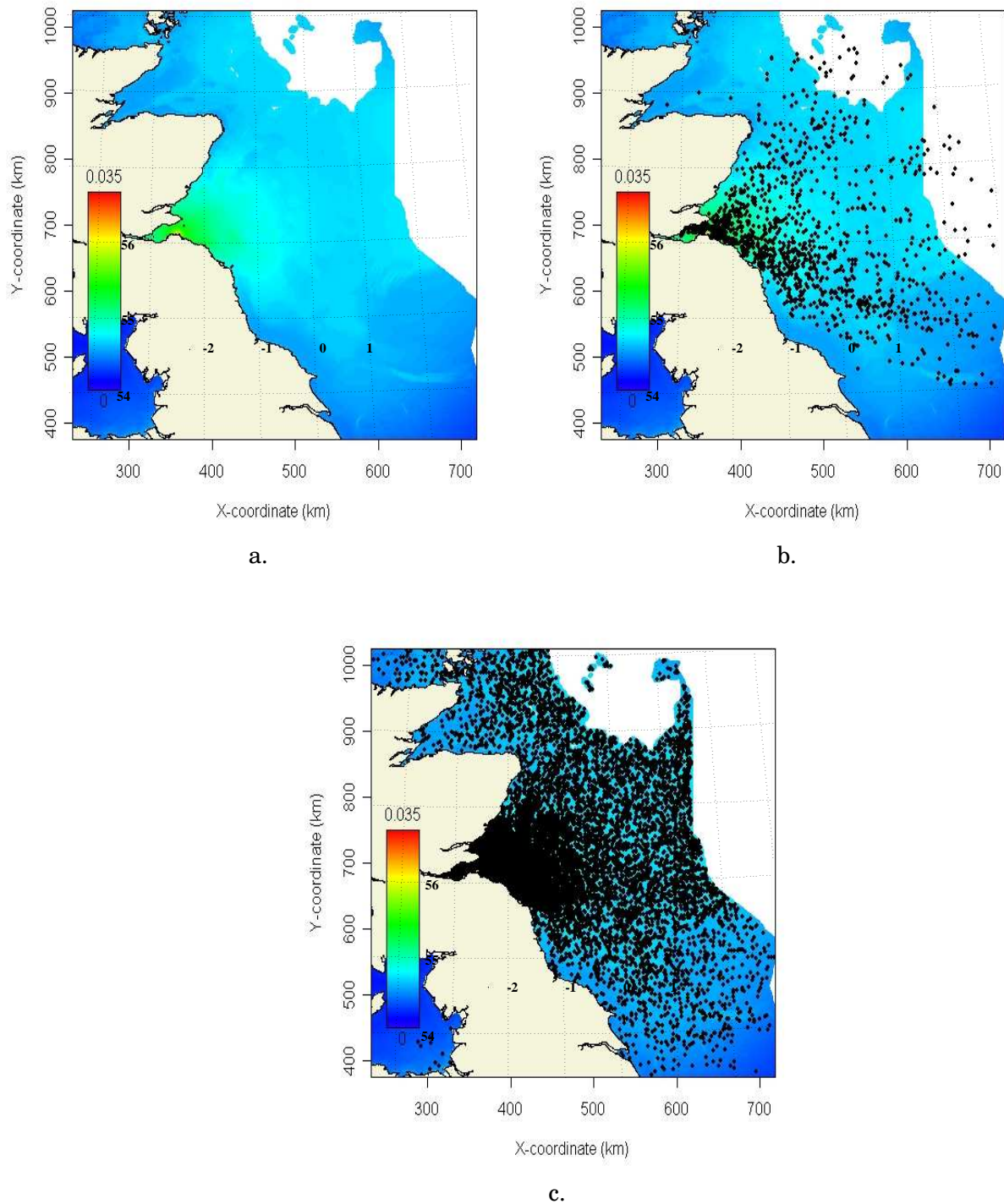
Fig 4.8 Spatial prediction of usage for the Bass rock without (**a**) and with (**b**) telemetry locations and one possible realization for the adult breeding population of Bass Rock using the usage predictions (**c**) plotted on top

To indicate which areas Gannets would use if all points in space were equally accessible, I plotted preference in space (Fig. 4.9).
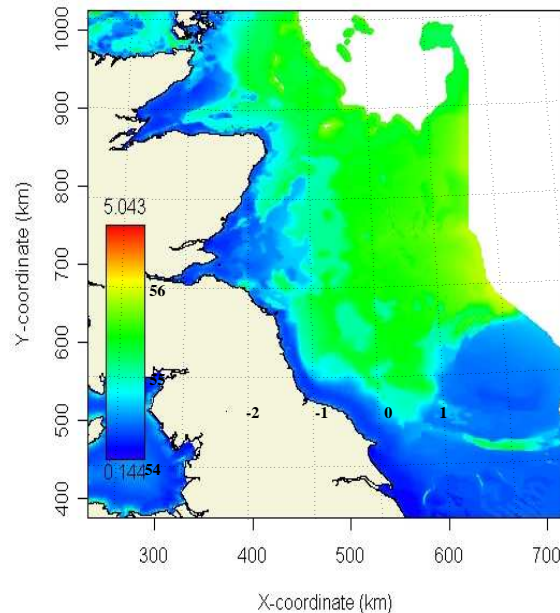
Fig 4.9 Spatial prediction of preference accounted for unequal accessibility and the effect of distance to the haulout site on the response.

4.4 Discussion

4.4.1 Habitat preference; Why are Gannets where they are?

Probably the most important conclusion of this study is that the biological and physical environmental variables I considered to be of importance (i.e. sediment type, depth, chlorophyll concentrations and diffuse attenuation), explain relatively little of the observed distribution of gannets at sea, especially when compared with the effect of these covariates on the spatial distribution of grey seals. There are several reasons for this:

Gannets predominantly feed on pelagic prey, such as mackerel, herring and sandeel (mostly 0-group individuals). The spatial and temporal distribution of prey items for those fish species (such as phytoplankton,

zooplankton and mesoplankton) is largely driven by physical oceanographic and atmospheric processes. These processes vary strongly in both space and time, and so do the distributions of the pelagic fish species that depend on them. If gannets cannot rely on predictable cues like the ones used in this study, they might spend considerable time searching for those resources, and the gannets' spatial distribution might strongly match the expected distribution under the accessibility model. Even if the spatial distribution of undisturbed prey is predictable by means of environmental cues, these may not be as useful if prey hotspots have been depleted by the prolonged action of gannets or other marine predators. Indeed, Lewis et al. (2001) show that intraspecific competition in larger gannet colonies increases the duration of foraging trips as individuals search for food further afield.

The covariates that partly explain the gannet's spatial distribution, are in order of importance, distance to the coast, depth and distance to the breeding site. Preference is high near the coast and declines sharply within the first 20km away from it. This may be due to higher levels of primary productivity originating from nutrient input from rivers and the strong vertical mixing (upwells) occurring near the coast. Another, explanation for the importance of distance from the coast can be found in the way gannets orientate to offshore foraging areas by flying parallel to the coast for a considerable part of their trip (Hammer et al. 2001).

Although previous studies have shown that there is sex-differentiation in dive depth (Lewis et al 2002), we found that neither the mass nor the sex of the individuals explained individual variability in habitat preferences.

4.4.2 Spatial predictions of usage and preference; Where else are Gannets likely to be?

Although few of the environmental variables and none of the individual characteristics were retained in the final model, it nevertheless provided an accurate description of total at-sea distribution. Whether this information is appropriate for managing the population depends on the objectives of management and conservation. For example, investigating the extent of seabird - fisheries competition might require more specific information about

the spatial distribution of actual feeding events. In contrast, risk assessment for other human activities in the North Sea such as oil drilling or the development of offshore wind farms, might require estimates about the overall at-sea distribution, such as the ones provided here.

4.4.3 Future research: additional environmental variables as candidate covariates

One of the main spatial features seen in the raw telemetry data is the gannets' tendency to forage in areas south-east and north-east off Bass-rock (Hamer et al. 2001), a pattern that is not reflected in this model's spatial predictions. In addition there might be more fine-scale features in the telemetry location data that changes over time caused by changing environmental conditions and that are therefore not revealed by the total composite of the telemetry locations (fig 4.2). We will discuss which other variables could explain those patterns.

Previous studies have shown that fish densities accumulate at fronts, which act as a natural barrier (Barnes and Hughes 1999). Also, during the summer months, regions with strong upwelling brings both nutrients and plankton, and therefore also fish assemblies closer to the surface. Both fronts and upwelling regions will be visible on Sea Surface Temperature satellite imagery from NOAA-Advanced Very High Resolution Radiometer (AVHRR) as strong temperature gradients and cold spots, respectively. Another covariate that could explain the observed distribution would be an autoregressive term, which would not only deal with problems of spatial and temporal autocorrelation in the response data (see §2.2.5 and §2.2.6), but might also explain potentially existing positive (e.g. aggregate feeding) or negative (e.g. resource competition) density dependent effects. Another variable that might explain the direction of flight and therefore also the actual at sea distribution is wind direction and velocity.

Gannets forage at sea to feed, and although their distribution is most likely to be related to fish, one could argue that gannets might actually respond to other cues like the ones presented above. However, this distinction can only be resolved using model selection techniques like the one

suggested in this paper and only if estimates of fish distributions are available. Such data is generally sparse in both space and time. Alternatively fish distributions could be related to environmental variables, using a similar approach as the one suggested in this study. In recent years considerable progress is made in developing physical oceanographic and biological models such as POL2dERSEM (Allen et al. 2001) and ECOSMO (Schrum et al in press). Predictions include, estimates for both phyto- and zoo-plankton biomass and productivity. Future research should focus on modelling fish distribution either empirically or mechanistically using these models, if we are ever able to understand the structure and functioning of marine ecosystem of which gannet form an important component.

4.5 References

Allan, J. I., Blackford, J., Holt, J., Proctor, R. Ashworth, M., and J. Siddorn. 2001. A highly spatially resolved ecosystem model for the North West European Continental Shelf. Sarsia 86: 423-440.

Barnes, R. S. K., and R. N. Hughes. 1999. An introduction to Marine Ecology, Third edition. University Press, Cambridge.

Birt-Friesen V.L., Montevecchi W.A., Cairns D.K., and S.A. Macko. 1989. Acitvity-specif metabolic rates of free-living northern gannets and other seabirds. Ecology 70: 357-367

Brierley, A., and P.G. Fernandes. 2001. Diving depths of Northern Gannets: Acoustic observations of Sula bassana from an autonomous underwater vehicle. The Auk 118(2): 529-534.

Burger, J. 1980. The transition to independence and post-fledging parental care in seabirds. Behav. Mar. Anim 4: 366-447.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference. A practical Information-Theoretic Approach., 2nd edition edition. Springer-Verlag, New York.

Hamer, K. C., R. A. Phillips, J. K. Hill, S. Wanless, and A. G. Wood. 2001. Contrasting foraging strategies of gannets Morus bassanus at two North Atlantic colonies: foraging trip duration and foraging area fidelity. Marine Ecology-Progress Series 224:283-290.

Hamer, K. C., R. A. Phillips, S. Wanless, M. P. Harris, and A. G. Wood. 2000. Foraging ranges, diets and feeding locations of gannets Morus bassanus in the North Sea: evidence from satellite telemetry. Marine Ecology-Progress Series 200:257-264.

Holland, G. J., S. P. R. Greenstreet, I. M. Gibb, H. M. Fraser, and M. R. Robertson. 2005. Identifying sandeel Ammodytes marinus sediment habitat preferences in the marine environment. Marine Ecology-Progress Series 303:269-282.

Lewis, S., Sherratt, T.N., Hamer, K.C., and S. Wanless. 2001. Evidence of intra-specific competition for food in a pelagic seabird. Nature 412: 816-819.

Lewis, S., Benvenuti, S., Dall'Antonia, L., Griffiths, R., Money, L., Sherratt, T.N., Wanless, S., and K.C. Hamer. 2002. Sex-specific foraging behaviour in a monomorphic seabird. Proc. R. Soc. London. B 269: 1687-1693.

Manly, B. F. J., L. L. McDonald, and D. L. Thomas. 1993. Resource selection by animals: statistical design and analysis for field studies. Chapman and Hall, London.

Matthiopoulos, J. 2003. The use of space by animals as a function of accessibility and preference. Ecological Modelling 159:239-268.

McConnell, B. J., C. Chambers, K. S. Nicholas, and M. A. Fedak. 1992. Satellite Tracking of Gray Seals (Halichoerus-Grypus). Journal of Zoology 226:271-282.

Otter Research Ltd. 2004a. An introduction to AD Model Builder Version 7.1.1. For use in Non-linear Modelling and statistics. Otter Research Ltd, Sidney B.C.

Otter Research Ltd. 2004b. Random effects in AD Model Builder. ADMB-RE user guide. Otter Research Ltd, Sidney B.C.

Pennycuick, C.J. 1987. Flight of aucks (Alcidae) and other northern seabirds compared with southern Procellariiformes: ornithodolite observations. J Exp Biol 128: 335:347

Pinheiro, J. C., and D. M. Bates. 2000. Mixed-effects models in S and S-plus. Springer-Verlag, New York.

R Development Core Team 2004. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, URL http://www.R-project.org.

Schrum, C., Alekseeva, I., and M. St. John. In prep. Development of a coupeled physical-biologicl ecosystem model ECOSMO Part I: Model description and validation for the North Sea.

Wood, S. N. 2001. mgcv: GAMs and Generalized Ridge Regression for R. R News Vol 1:20-25.

Wood, S. N. 2006. Generalized Additive Models: an introduction with R. in press.

Wright, P. J., H. Jensen, and I. Tuck. 2000. The influence of sediment type on the distribution of the lesser sandeel, Ammodytes marinus. Journal of Sea Research 44:243-256.

_____

5.   MODELLING SPACE-USE AND HABITAT PREFERENCE IN GREY SEALS (HALICHOERUS GRYPUS) OFF THE WEST COAST OF SCOTLAND

_____

Human exploration and exploitation of the marine environment through activities such as fishing (Wickens et al. 1992; Lunneryd et al. 2003; Read & Brownstein 2003), military exercises (Jepson et al. 2003; Goold 1998; Goold & Fish 1996), drilling for oil and gas and development of wind farms (Wursig & Greene 2002; Harwood & Wilson 2001), can often come into conflict with marine mammal populations. This can be particularly acute in regions where intensive human activities overlap with hotspots in the distribution of marine mammals. A simple but effective mitigation measure is to reduce human activities in areas with high densities of marine mammals.

Estimates of the distribution of many marine mammal species can be made using visual line-transect methods (Hammond et al. 2002; Macleod et al. 2003). However, these are not very effective for grey seals because they spend most of their time under water, they are difficult to detect while at the surface, they forage at remote areas and are relatively sparse. Instead, grey seals can be individually tracked using Satellite Relay Data Loggers

(SRDL - Fedak and McConnell 1993). A major disadvantage of this technique is that, due to logistic and financial constraints, very few such devices can be deployed and sampling effort is often unbalanced across different geographic regions and types of animals. This is problematic, because conservation and management of grey seals requires unbiased estimates of their spatial distribution within an entire region of interest.

Although individual grey seals vary considerably in how they use space, their collective space-use is ultimately shaped by the availability of resources and environmental restrictions. In this chapter I firstly aim to model the effect of environmental variables (both static and time-variant) on the distribution of grey seals foraging off the west coast of Scotland. Secondly, I aim to combine the predictions of the habitat model with aerial counts of animals at haulouts in order to estimate the expected abundance of grey seals at sea.

5.1 Facts about the natural history of grey seals on the west coast of Scotland.

The Inner and outer Hebrides are the most important regions for grey seals, containing an estimated 49% of the total UK population (Hammond et al. 1994). During the breeding season (late September to early October), which starts approximately 1 ½ months earlier than in the east coast, large numbers aggregate on land (King 1983). During the remainder of the year, individuals frequently aggregate on coastal sites, known as haulout sites.

From these haul-out sites they make foraging trips to sea, feeding on a variety of fish species. Most grey seal telemetry data came from seals foraging from halout-sites in the southern part of the Inner Hebrides (area 1 Fig 5.1). In this region, the diet of Grey seals (determined from scat-samples collected from April until September 2002), predominantly consists of Sandeels (*Ammodytes marinus* - 59.57% of diet expressed in weight), Atlantic Cod (*Gadus morhua* - 14.86%), Lemon Sole (*Microstomus Kitt* - 10.09%) and Bullrout (*Myoxocephalus scorpius* - 8.12%) (SMRU unpublished).
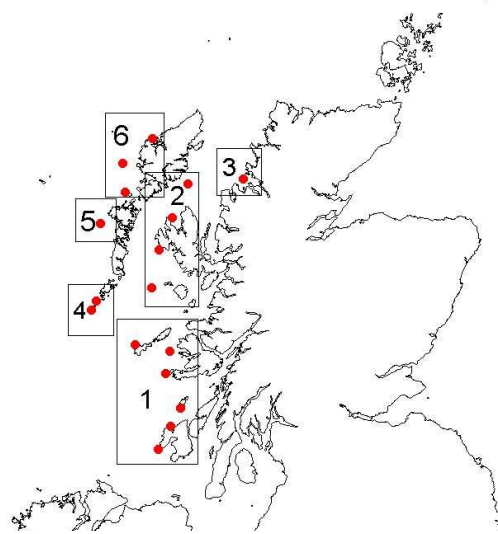
Fig. 5.1 Haul-out sites from which scat samples were collected in 2002. Area 1 contains the haul-out sites from which individuals were caught and equipped with satellite transmitters.

5.2 Methods

5.2.1 Response variable

In March and April 2003 and 2004 grey seals were caught at Islay (55°54N, 6°20W), Collonsay (56°01N, 6°15W) and Tiree (56°30N, 7°00W). Seals were anaesthetised, measured and equipped with Satellite Relay Data Loggers (SRDLs). The SRDLs sent UHF signals to two polar-orbiting satellites with an Argos satellite system. The location of the animal was determined using the frequency Doppler shift of the signal (Argos 1989). These estimates contained a small number of highly erroneous location fixes which were removed by applying the filtering algorithm described by (McConnell et al. 1992), using a maximum swim speed parameter of 5 m/s. Some remaining error in the data was treated using a smoothing algorithm developed (M. Lonergan unpublished) within the MGCV (Wood 2001) package in R (see §3.2.1). I used the output of this algorithm to interpolate the animals' positions at 6-hourly intervals. Obtaining temporally regular estimates of the animal's position also helped to reduce biases due to the variable observability of the animals (§ 2.2.7).

Not every telemetry observation was used in defining the response variable. In summary, only those telemetry locations collected in July, August or September that were further than 5km from the haulout site and observed during a return trip were used to model habitat preference. The motivation for each of these decisions is given below:

Excluding locations within 5km of a haulout sites.— Many of the smoothed animal locations are on, or very close to land. Grey seals tend to aggregate on land, as well as in the water close to land, to rest and perhaps socialise (pers. obs.). Such activities are unlikely to be related to the physical and biological variables used in this study, and if they are, the complex topography of inshore areas and the lack of fine-scale environmental data combined with the relative large errors in the estimated animal locations, would not allow me to correctly model habitat preference. Therefore, I restricted my attention to those telemetry locations further than 5km from any haul-out site.

Excluding telemetry observations outwith July, August and September.— The grey seal preference for prey changes as a consequence of seasonal changes in prey availability (Smout 2006). It is therefore likely that the spatial distribution of grey seals will also vary seasonally. Accounting for these changes would require including interaction terms between (a possibly non-linear function of) time and every environmental variable. This would make unfeasible computational demands on parameter estimation. Furthermore, under the second objective of this chapter, estimates of at-sea abundance will be made with the aid of haulout counts. These are only available for the month of August. Finally, these months immediately precede the breeding season during which females need to facilitate a growing foetus and build up large energy supplies for milk production later in the year. Hence, this period may play an important role in determining some of the population's demographic rates and is therefore very important from a conservation and management perspective.

Return trips observations only.— Under the case-control paradigm, to provide the contrast necessary for modelling preference, the error-corrected and temporally-smoothed data of satellite locations (presence data) has to be complemented with a set of points not visited by the study animals (absence data). For central-place foragers, the absence points can be drawn from a spatial surface of accessibility (Matthiopoulos 2003b). The accessibility of a point in space is defined as the likelihood of that point being visited by individuals that show no preference for environmental conditions, but are subject to physical restrictions to movement (obstacles, swimming speed, trip durations, and the start and end point of a trip). The start and end points of a trip can be different (transitory trips), but are more often (McConnell et al. 2000) the same (return trips). Matthiopoulos (2003b) estimated the likelihood of a point in space being visited by individuals making return trips and showed how to calculate the equivalent likelihood for transitory trips. Generating a map of aggregate spatial use for the population using both return and transitory trips, requires appropriate weighting of trips performed between all pair-wise combinations of haulouts. The weights for this operation must relate to the relative frequency of occurrence of trips between any two haulouts. This information is not yet available. I therefore restricted my attention to return trips and excluded transitory trips from my analysis data frame. On the east coast of Scotland the proportion of transitory trips was small, namely 14%. In contrast, the proportion of transitory trips might be greater for grey seals foraging off the west coast of Scotland. This may be because, the larger number of haulout sites on the western coast results in smaller average distances between them and makes it less costly for seals to travel between them or they appear to have travelled between them due to observation error. To prevent excessive censoring of data, I clustered haulout sites into groups. Model-based clustering techniques exist that can automatically determine the number of clusters by means of theoretical information criteria (e.g. BIC - Ter Braak et al. 2000). Because I had to calculate the at-sea distance to every cluster, I was not just interested in the the optimum number of clusters, but I was also restricted by computational costs. Therefore, I assumed a fixed (though

arbitrary) number of 25 clusters. The clustering method is described in Appendix 5.A.

5.2.2 Explanatory variables

I provided values for the explanatory variables corresponding to conditions at each set of spatial coordinates contributing to the response data. I selected environmental variables on the basis of possible biological relevance and availability of data. Since grey seals go out to sea to forage, a useful set of covariates would be the spatial distributions of their prey. However, fish distributions are notoriously variable in space and time and the data and models that could be used to describe these distributions are scarce or unavailable. Instead, I opted for more precisely measured covariates such as the abundance and accessibility of phyto-and zooplankton (the major food source for species such as sandeel), the distribution of sediment components (a determinant of the burying conditions for ground-fish) and time restrictions on horizontal and vertical movement of grey seals.

Phyto- and zooplankton abundance.— The fish species on which grey seal prey, feed on phytoplankton, zooplankton and other species of fish and this may lead to a positive correlation between the distribution of seals and the distribution of their prey's food sources. One proxy for the abundance of phytoplankton is chlorophyll concentration which can be measured remotely by MODIS (MODerate resolution Imaging Spectroradiometer) and SEAWIFS (Sea-viewing Wide Field-of-view Sensor) satellite images which were extracted from http://oceancolor.gsfc.nasa.gov/. For every value of the response, given its spatial position and time of observation, I extracted the corresponding chlorophyll values from 8-daily composites. If, due to extensive cloud cover, satellite observations were missing I used values from monthly composites instead. The chlorophyll concentrations from both satellite sources were averaged.

Zooplankton concentrations cannot be measured remotely and their abundance is not necessary proportional to phytoplankton abundance due to delayed responses. Instead, I used estimates of zooplankton abundance

generated by a hydrodynamic model, the POL-3DB (Proctor & James 1996; Holt & James 2001) baroclinic model with the European Regional Seas Ecosystem Model, ERSEM (Baretta et al. 1995). This model describes the biogeochemical cycling of carbon, nitrogen, phosphorous and silicate through both the pelagic and benthic ecosystem and the coupling between them (Allen et al. 2001).
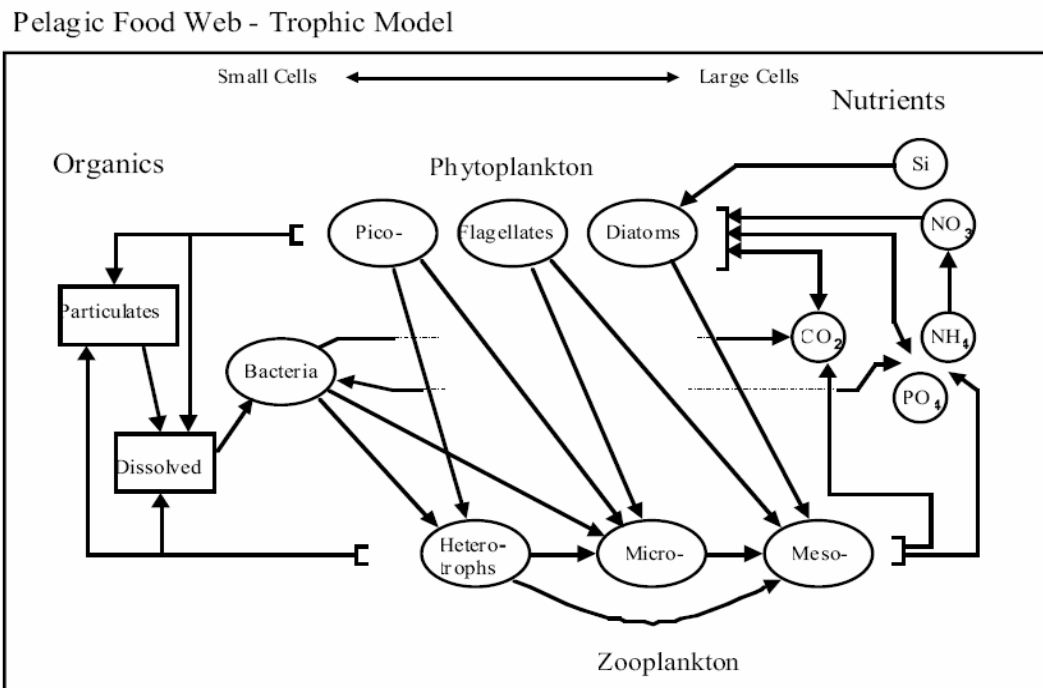
Fig 5.2 The pelagic food web of the ERSEM model (from Allen et al. 2001).

Phyto and zooplankton accessibility.— For fish to able to exploit the abundance of phyto- and zoo-plankton efficiently these resources need to be aggregated in sufficiently high densities. The measures on zooplankton and phytoplankton abundance described in the previous section do not quantify spatial (described by latitude and longitude) aggregations within cells neither do they quantify vertical aggregations in the water column. Aggregations generally occur in places with steep temperature, salinity and density gradients.

Although temperature is measured at the same spatial resolution as chlorophyll concentrations, measuring horizontal variability in temperature can reveal information about the dynamic nature of that cell and therefore

provide an indiciation of the amount of heterogeneity within that cell. Horizontal gradients can be observed remotely using satellite imagery. Using satellite measures of sea surface temperature, I estimated the gradient α for every cell in space using the eight surrounding cells by means of the following equations (Burrough 1986)

$$\partial x = \frac{\left(z_{i-1,j+1} + 2z_{i,j+1} + z_{i+1,j+1}\right) - \left(z_{i-1,j-1} + 2z_{i,j-1} + z_{i+1,j-1}\right)}{8\partial l_x}$$

$$\partial y = \frac{\left(z_{i-1,j-1} + 2z_{i-1,j} + z_{i-1,j+1}\right) - \left(z_{i+1,j-1} + 2z_{i+1,j} + z_{i+1,j+1}\right)}{8\partial l_y} \qquad 5.1$$

$$\alpha = \arctan\left(\sqrt{\partial x^2 + \partial y^2}\right)\frac{180}{\pi}$$

where z is the SST value and ∂l_x and ∂l_y are the width of the cells in the x and y directions respectively.

Vertical aggregations in the the water column cannot be visualized using satellite imagery. Instead, I used the POL-3DB model predictions to estimate the difference in surface and sea-bottom temperature as a measure of stratification. On average, deeper areas will have larger differences between surface and bottom temperature independently of the amount of mixing that takes place. But, because depth is also included as a candidate covariate in the model, it should account for some of this effect.

Fish burying habitat preference.— Benthic species are the main sources of food for grey seals. For the Shetland islands and East coast, Wright et al. (2000) and Holland et al. (2005) have shown that the sandeel distribution is closely related to sediment type: Sandeels appear to avoid mud. It is believed that this strategy prevents their gills from getting clogged.

Sediment type data come from 3 sources: i) maps of kriged sediment core samples (see §3.2.2), ii) BGS classified sediment conditions and iii) UKHO sea bottom texture side-scan data. The latter two sources were stored as categorical data in both environmental and geographical space and the classes were transformed to values for the percentages of mud, gravel and sand in the sediment using the Folk-classification (Folk 1980). For any point

in space, use of these data sets, yielded at least one and usually three estimates of sediment composition. When more than one estimates were available I used the mean percentage of each category (gravel, mud or sand). Some points were classified as "Rock" by either the UKHO side-scan or BGS habitat classification. I treat rock as a dummy variable; its value is 1 if rock is present and 0 if absent.

Grey seal foraging restrictions.— Since grey seals predominantly dive to the sea bottom, they spend more time travelling to depth in deeper waters, shortening their effective foraging duration. In the extreme, some depths might not be accessible at all and should therefore be avoided. Depth was therefore a relevant candidate covariate. In addition to these physiological and temporal restrictions, some studies (Wright et al. 2000, Holland et al 2005) have shown that fish distributions are not uniformly distributed with depth. For example, Wright et al. 2000 showed that sandeels have a peak preference for depths in the range of 50 to 60 meters.

Distance from the haulout might also impose limitations to usage. This should, to a large extent, be captured by the accessibility model. This model includes information on the distribution of trip durations derived from tagging data collected around the UK. However, Matthiopoulos et al. (2004) indicate that these can be different between haulout sites, leading to slightly different accessibility surfaces. These deviations from the accessibility model will lead to residual under/over-prediction as a function of distance from the haulout. To account for the these effects, I included the swimming distance to the haul-out site as a covariate.

5.2.3 Model structure

The basic structure of the model follows from using a binomial response variable (h) with a logit link

$$\hat{u}_l \sim \text{B}(1, h_l)$$
$$h_l = g^{-1}(\eta_l) = \frac{e^{\eta_l}}{1 + e^{\eta_l}}$$

5.2

To make the model flexible (§2.2.8), the linear predictor (η_l) follows the structure of a GAM for a total of twelve explanatory variables

$$
\begin{aligned}
\eta_l = b_0 &+ s_1(\text{distance}_l) + s_2(\text{depth}_l) + s_3(\text{mud}_l) + s_4(\text{gravel}_l) + \\
&s_5(\text{chlorophyll}_l) + s_6(\text{sst}_l) + s_7(\text{bottom speed}_l) + \\
&s_8(\text{temperature stratification}_l) + s_9(\text{temperature gradient}_l) + \\
&s_{10}(\text{zooplankton}_l) + s_{11}(\text{phytoplankton}_l) + \\
&s_{12}(\text{bottom speed} * \text{zooplankton}_l) \\
= b_0 &+ \sum_{k=1}^{6} b_{1,k} f_{1,k}(\text{distance}_l) + \cdots
\end{aligned}
$$

$$5.3$$

where each spline $s(\cdot)$ is a flexible function of each explanatory variable, and $b_{j,k}$ is the random effect parameter for the kth out of 6 basis function of the spline for the jth explanatory variable. Construction of splines followed the protocol developed for grey seals (§3.2.3). In the most saturated form of the model, each parameter, was expressed as a function of individual characteristics

$$
b_{j,k,m} = \beta_{j,k,0} + \beta_{j,k,2}\text{sex}_m + \beta_{j,k,3}\text{length}_m + v_{j,l} \qquad 5.4
$$

where *m* refers to a specific individual.

5.2.4 Parameter estimation

I estimated the model's coefficients β_j and the variances σ_j^2 of the random effects by maximum likelihood. Parameter estimation is described in detail in §3.2.4

5.2.5 Model selection and model validation

I adhered to the model selection protocol derived for grey seals from the East coast of Britain, in section §3.2.5. I used data from 14 individuals for model fitting and the remaining 13 individuals for model selection.

5.2.6 Spatial prediction of usage

I used the same rationale behind the grey seal usage predictions made for the east coast of Scotland (§3.2.6). To scale prediction of usage up to total abundance at sea, I first needed to calculate the total population size N_i associated with each haul-out site i,

$$N_i = \frac{n_i}{p_{land}} \qquad 5.5$$

where n_i is the number of seals observed during the 96-97 aerial survey and $p_{land} \cong 0.21$ is the proportion of animals that are expected to be hauled-out during these counts. It is calculated by dividing the total number of individuals observed during the 96-97 survey, which is 24,047, by the total UK population size (116,000) which is based on annual pup-production (Sea Mammal Research Unit 2002). The expected density of seals at sea within a spatial unit c (e.g. a 1km grid cell) can be estimated as follows

$$u_c = \sum_{\text{all } i} N_i p_{\text{at sea}} \int_{\mathbf{s} \in c} \upsilon(\mathbf{s}) d\mathbf{s} \qquad 5.6$$

Where $p_{at\,sea} \cong 0.613$ is the proportion of time spent foraging outside a 5km range of the haul-out.

5.2.7 Spatial prediction of preference

As in the east coast study (§3.2.7), predictions of preference were generated from the relationship

$$\frac{\upsilon(\mathbf{s})}{\alpha(\mathbf{s})\exp(\eta_{\text{distance}})} \propto \exp(\eta_{\text{environmental variables}})$$ 5.7

5.2.8 Software

Most data manipulation, analyses and plotting (including spatial plotting) was done using R (R Development Core Team 2004). ML parameter estimation was done using ADMB-RE (Otter Research Ltd. 2004a, b). Transformation of satellite images in .hdf to ASCII raster was done using Manifold. Extracting of the environmental conditions for points in space and time was done using an ArcGis extension developed in-house by the Scottish Association of Marine Science (SAMS).

5.3 Results

5.3.1 Response and environmental variables

A total of 27 individuals were caught, equipped with satellite transmitters, released and observed from July 1st to September 30th in either 2002 or 2003. Following track filtering, a total of 38.7% locations were within 5km of a haulout site and were therefore excluded from the analysis (see §5.2.1). The remaining locations were part of a trip returning to the same haulout (i.e. return-trip) or a different one (i.e. transitory trip). To determine which ones would be classified as return trips I grouped the total number haul-out sites (483) into 25 clusters (Fig. 5.3a).
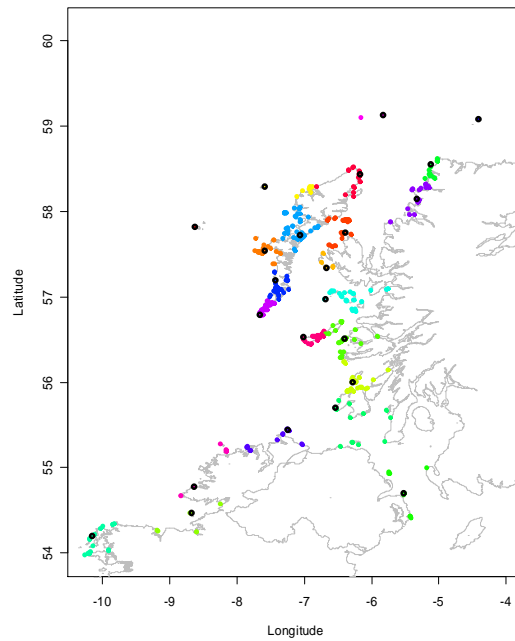
a.
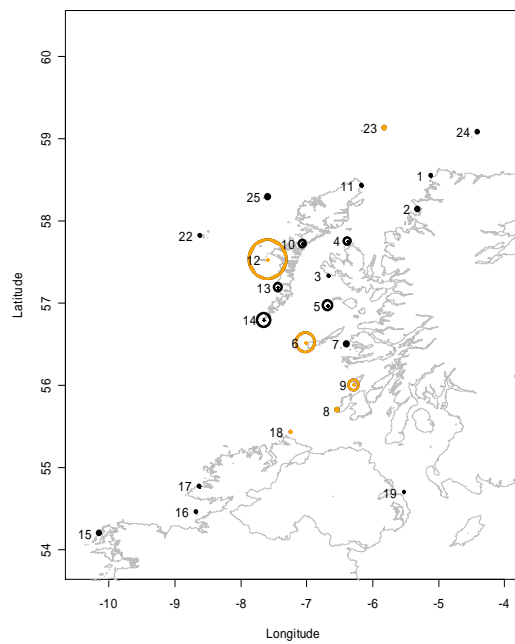
b.

Fig. 5.3 All haul-out sites on which grey seals were observed in August 1996 or 1997, colour-coded by cluster which are identified by a black dot (**a**). All 25 clusters with point sizes being proportional to the number of individuals. Orange points are clusters for which some telemetry data used in this study was available (**b**).

After clustering the haul-sites, 65% of the at-sea locations belonged to return-trips (Fig. 5.4a) corresponding to a total of 4947 locations. Every location (with known latitude, longitude and time), was augmented by one absence point in space selected from the accessibility surface of the relevant haulout cluster, and a time, identical to the time of the corresponding animal location (Fig. 5.4b).

The combined presence and absence data constituted the response data that were to be regressed against the environmental variables (examples shown in Fig. 5.5).
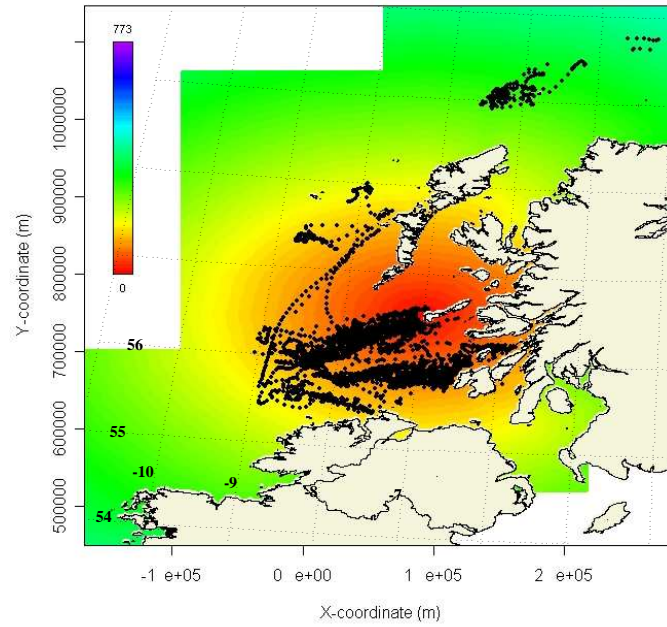
a.
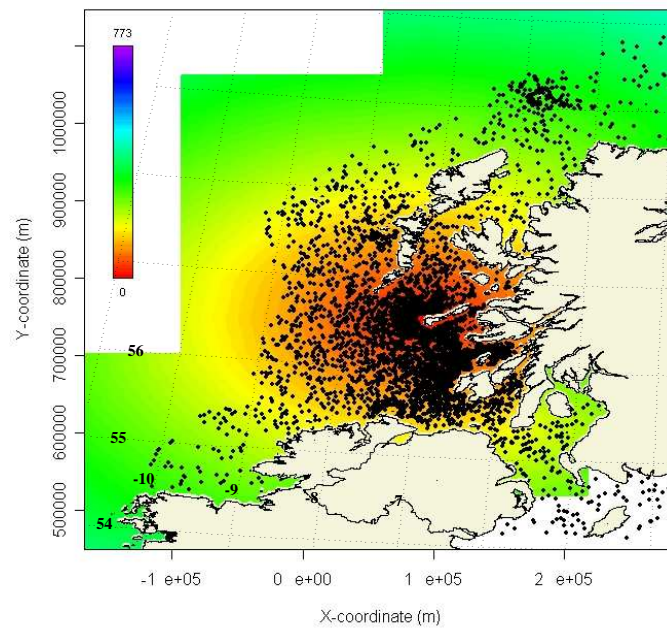
b.

Fig. 5.4 All animal locations used in this study (a) and pseudo-absences drawn from the accessibility surfaces (b) plotted on top of the swimming distance from the Tiree haulout (cluster 6 in Fig. 5.3b).

a. Bottom speed

b. Temperature stratification

c. Zooplankton

d. Phytoplankton
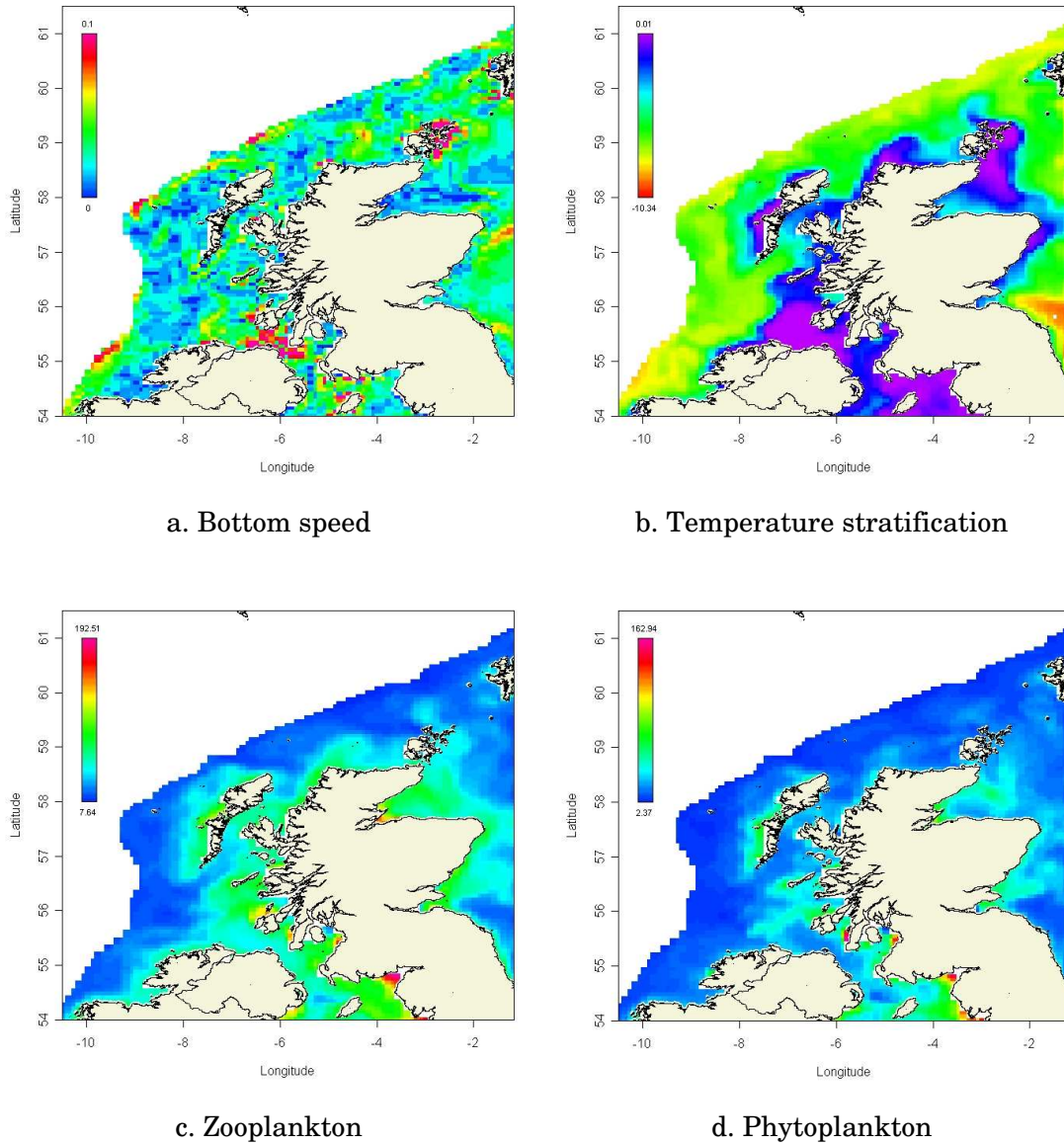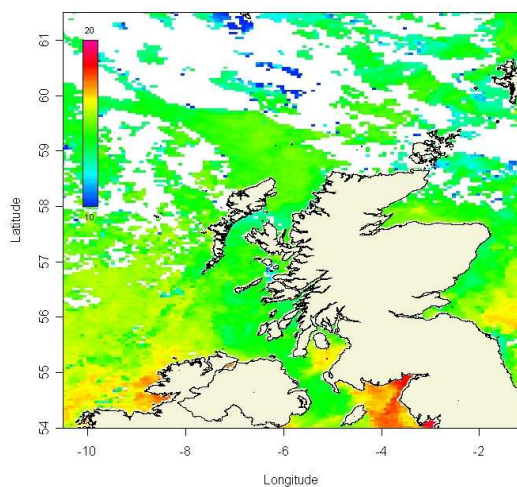
e. 8-day sea surface temperature
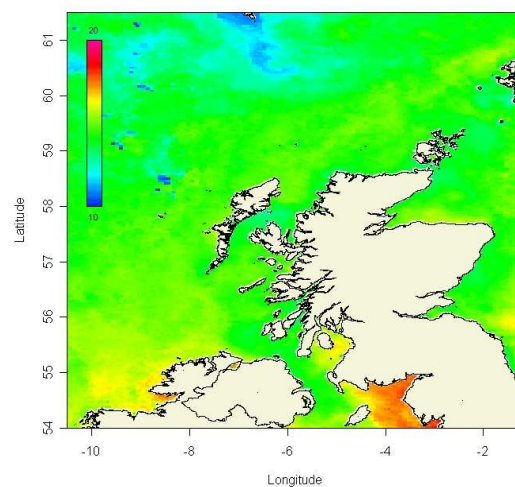
f. Monthly sea surface temperature
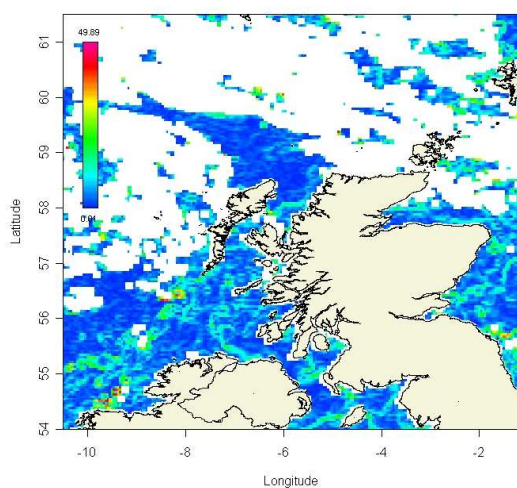
g. 8-day temperature gradient
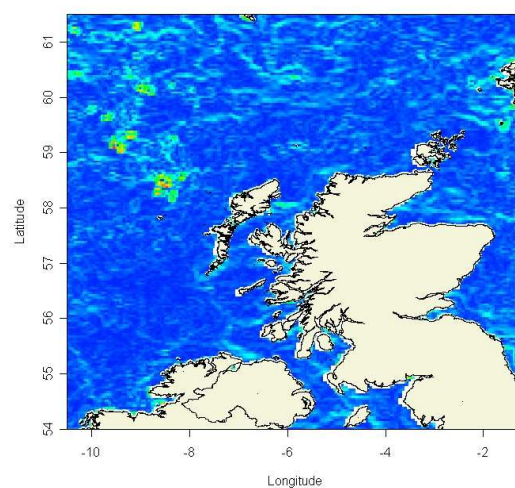
h. Monthly temperature gradient
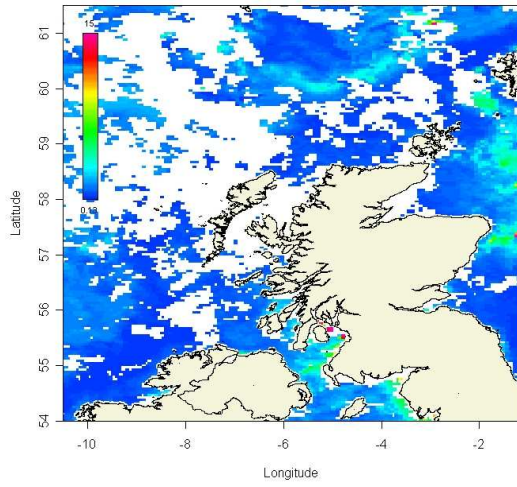
i. 8-day chlorophyll concentration
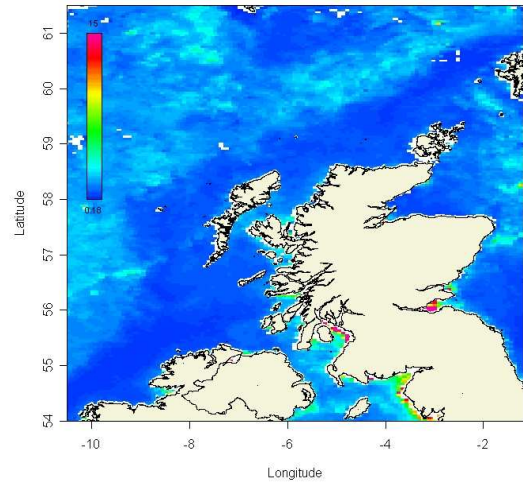
j. Monthly chlorophyll concentration
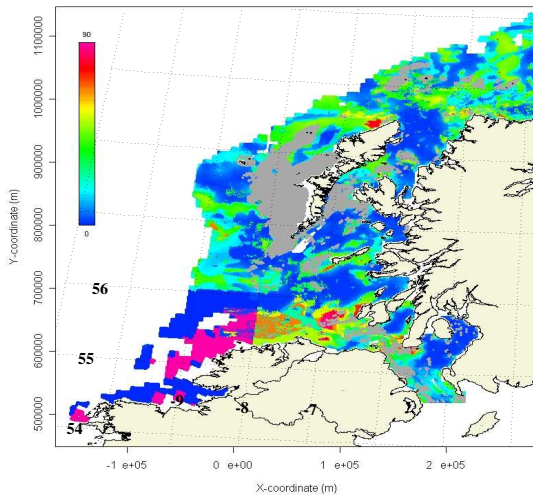
k. Gravel concentration
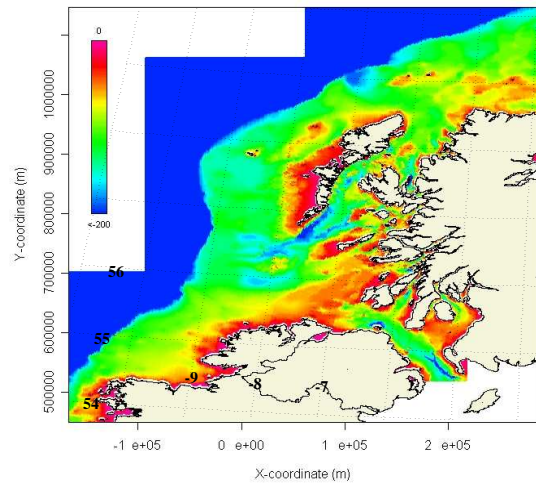
l. Depth

Fig. 5.5 Spatial representation of the bottom speed in m/s (**a**), temperature stratification expressed as the difference between surface and bottom temperature (**b**), zooplankton (**c**) and phytoplankton (**d**) concentrations, MODIS satellite derived 8 day (**e**) and monthly (**f**) sea surface temperature, estimated 8 day (**g**) and monthly (**h**) temperature gradient in degrees, MODIS satellite derived 8-day (**i**) and monthly (**j**) chlorophyll concentrations, percentage gravel in the sediment (colour coded) and the presence of rock (grey) (**k**) and depth (**l**).

5.3.2 Model selection

Due to computational constraints, I was limited to fitting models containing up to 5 (instead of all 12) covariates. Forward model selection suggested that a model containing the first five most important environmental variables explained 51% of the variation in the response (Figs 5.6 & 5.7). Validation of the candidate models generated by this process indicated that all five covariates significantly explained some of the variability in the response of the test data (Fig. 5.7).

It was conceivable that this model was too flexible due to the use of splines. The sequential replacement of splines by the un-transformed environmental variables led to a deterioration of the model's AIC under both the fitting and test data for all variables, except for the covariate sand. In that case, replacing the smooth function of sand by its linear term reduced the AIC from 3272.88 to 3267.94

To explain some of the residual variation, I also included individual characteristics (eq 5.4). Including an interaction between depth and sex of the animal, lead to a decrease in AIC from 3270 to 3244. An additional interaction between stratification and sex reduced the AIC even further to 3234. However, including the individual-specific effects led to a deterioration of the model under the test data; the negative log-likelihood of the test data set increased from 2038 to 2052 and to 2066, respectively. Therefore, I used the model with no individual-specific covariates to investigate the relationship between the response and the different environmental variables and to make spatial predictions of usage and preference. The parameter estimates, the parameter standard deviations and correlations of a simpler model, one which excludes the covariate sand, are shown in Table 3 from Appendix B. Unfortunately, it turned out that the Hessian matrix of the full model was not positive definite. Therefore the standard deviations and correlations of the parameters could not be calculated. However, this did not influence the parameter estimates of both random and fixed effects and all futher analysis and results could be based on the full model.
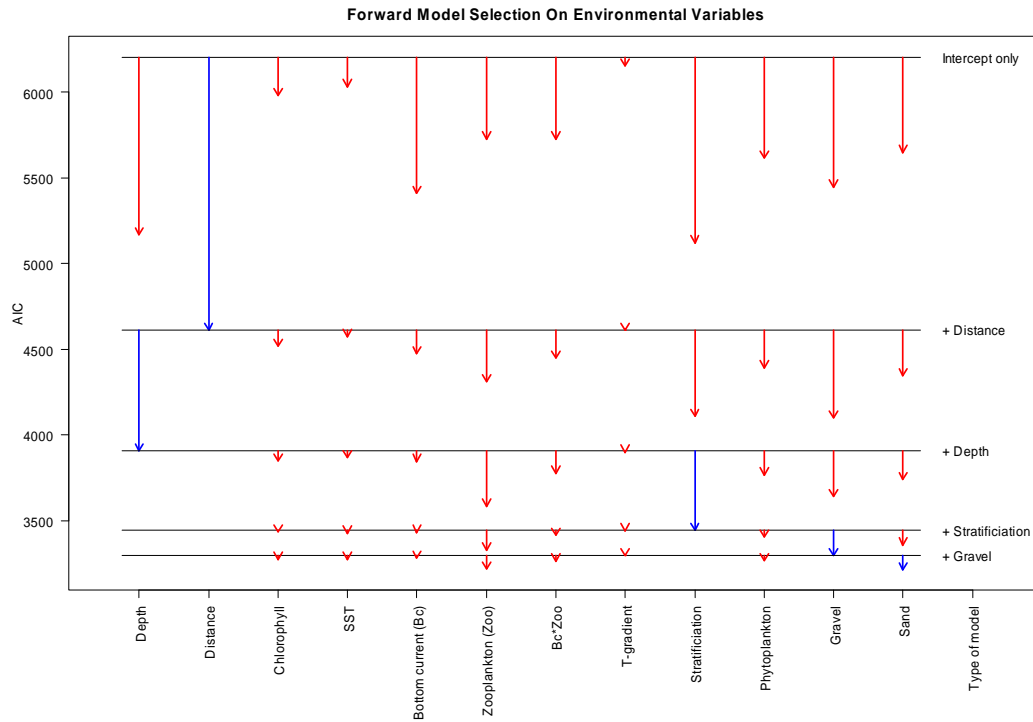
Fig 5.6 Forward model selection on the environmental variables; depth, distance to the haul-out site, satellite-derived Chlorophyll concentrations, Sea Surface Temperature (SST), Bottom current speed, Zooplankton concentrations, an interaction between the Bottom current speed and zooplankton concentrations, temperature gradient, Stratification, Phytoplankton concentration, Gravel and Sand. The arrows indicate the change in AIC between models (horizontal lines) as a result of adding an environmental variable. Arrows for variables that lead to the largest improvement in AIC are coloured in blue.
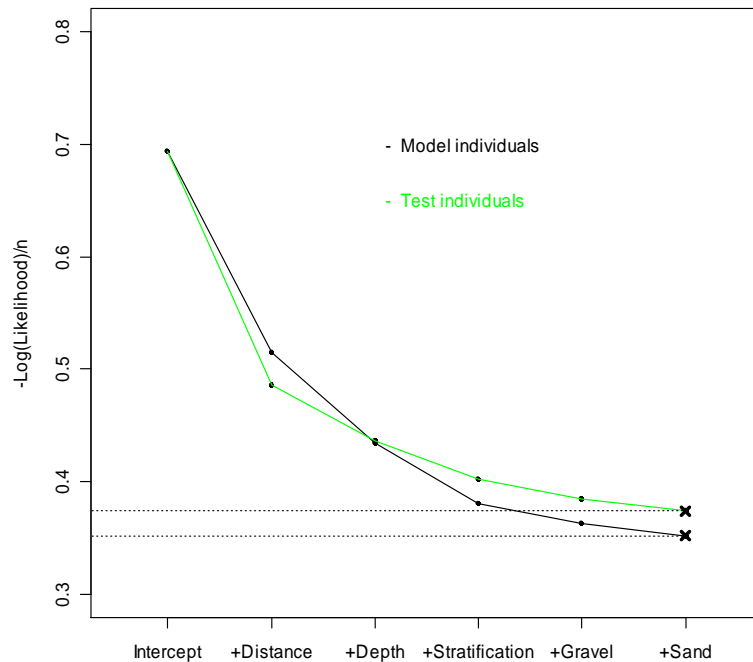
Fig 5.7 Validation of models using the test data set. Note that all parameter estimates used to calculate the likelihood are based on the 14 individuals from 6 haulout sites and are compared with the remaining individuals. Different models are arranged in order of increasing number of environmental variables along the x-axis. The *y*-axis represents the log-likelihood of the data under each model, standardised by the number of data points (*n*) in each of the three datasets. The black crosses indicate the best model for each data set.

It is possible that some important environmental variables were omitted from the analysis. This would reveal itself in the presence of spatial autocorrelation in the residuals (§ 2.2.5). I constructed semi-variograms to investigate the spatial autocorrelation in the response and residuals for large (<40,000 m) and small (<5,000 m) spatial scales (Fig. 5.8). These revealed the presence of spatial autocorrelation in both the small and large spatial scales, meaning that some covariates were missing from the model or the accessibility surface was inappropriately represented by the smooth function of distance.
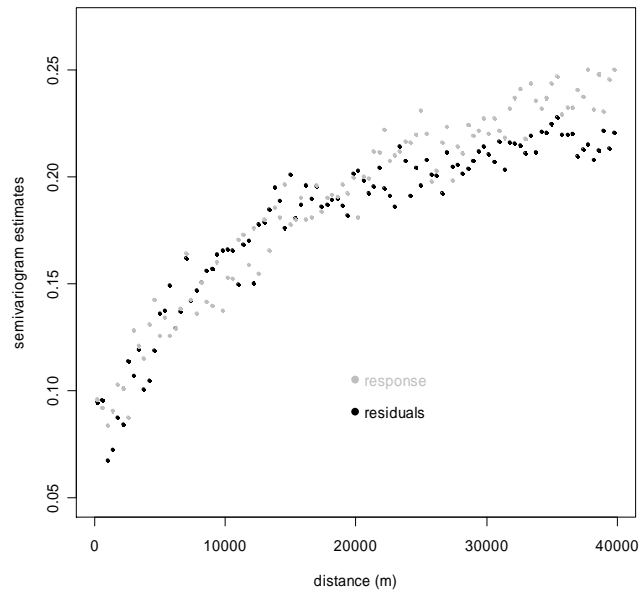
a.
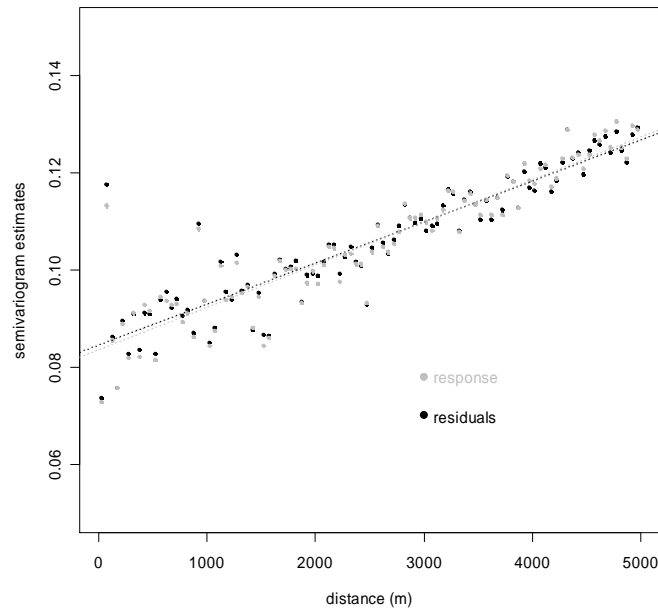
b.

Fig 5.8 Semi-variogram of the response data (0 and 1) and the residuals (observed response - predicted) for up to 40km (**a**) and 5km (**b**) with the range of distance being partitioned into 100 distance classes in both cases. The plotted lines are linear regression models fitted to the data. The trend is significant (*p-value* <<0.001).

5.3.3 The effect of environmental variables on the response.

The most important covariate was distance to the central-place which had a negative relationship with the response, meaning that the observed usage away from the haul-out site is less than predicted by the accessibility model (Fig. 5.9a). The second most important environmental variable was depth. Fig. 5.9b shows that animals avoid deeper areas. Fig. 5.9c shows that grey seals prefer areas that are characterized by high stratification. Finally, the last variables to be included were descriptors of sediment type. Grey seals appear to prefer a mixture of gravely (Fig. 5.9d) and sandy (Fig. 5.9e) areas. Rock, which is treated as a factor, has a positive parameter value of 1.2. However, both sand and gravel concentrations are zero at those places. On average, the gravel and sand concentrations of the model data were 24.9% and 66%, respectively. The cumulative increase of the response on the scale of the link-function at those sand and gravel concentrations (Figs 5.9d and 5.9e) compared to a sediment type with no sand and gravel, is approximately 6.4 (gravel: +2.9 and sand: +3.5). So, even though the parameter value for rock is positive, compared to the mean sediment type, rock is on average, avoided.
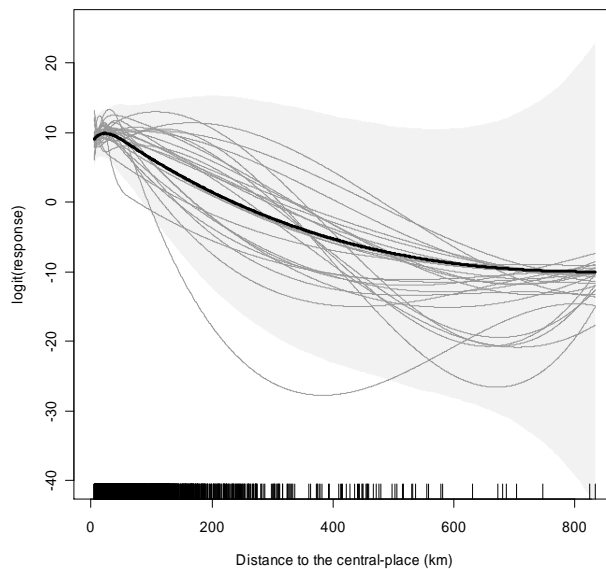
a.
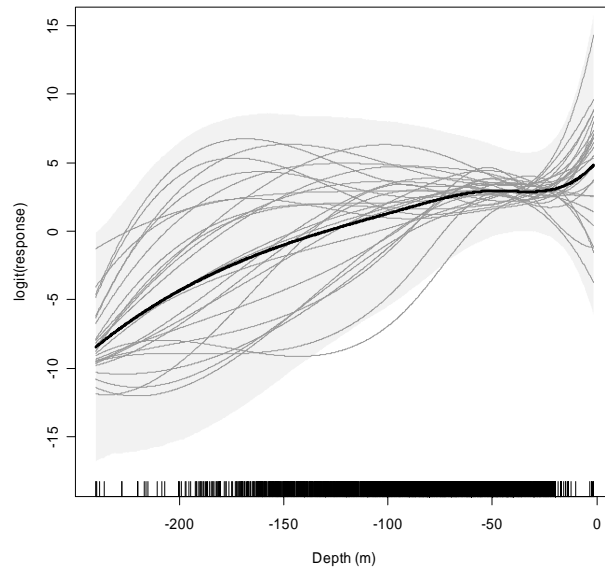
b.
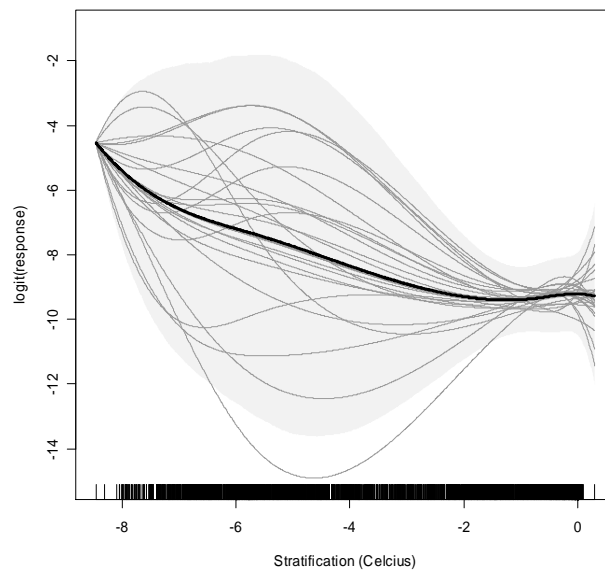
c.
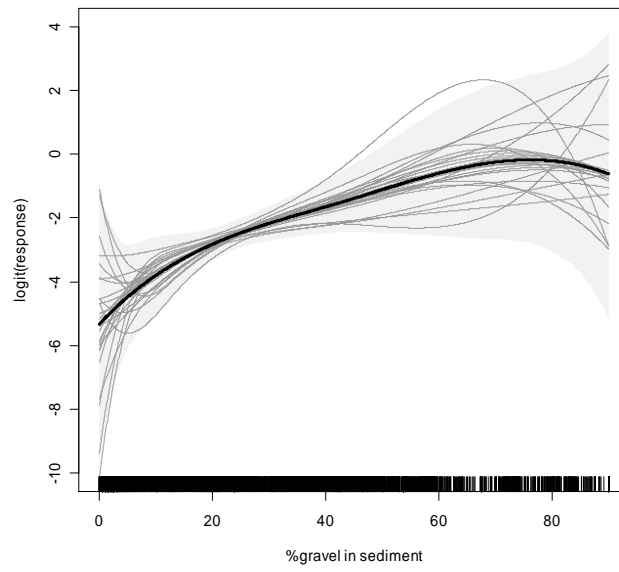
d.
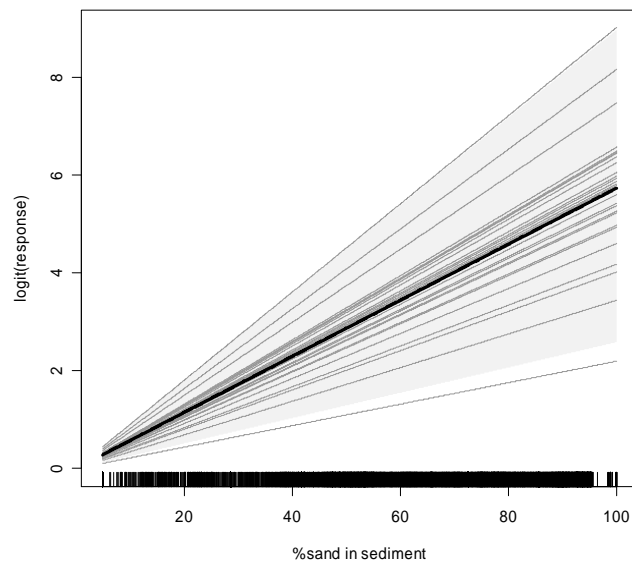
e.
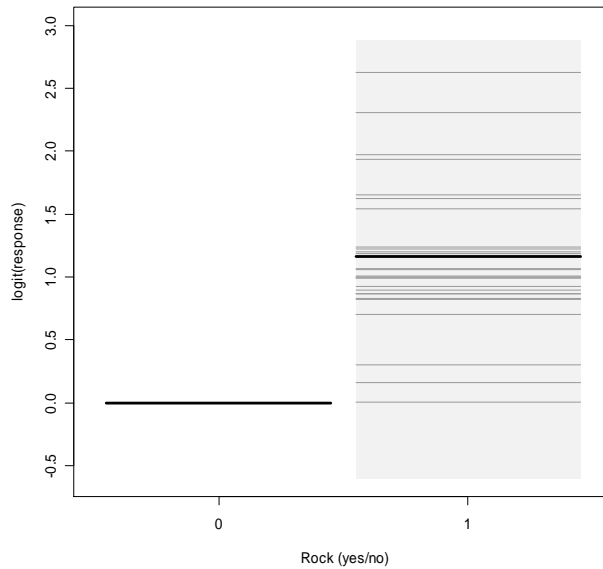
f.

Fig. 5.9 The effect of the covariates distance to the haul-out site (**a**), depth (**b**), temperature stratification (**c**), percentage gravel (**d**) and sand (**e**) and whether the substrate is rocky or not (**f**). The 95% confidence limits are represented by the shaded areas, the grey lines represent individual specific responses and the black line the fixed effect population responses.

5.3.4 Spatial prediction of usage and preference

Using the relations between the response and the environmental variables outlined in the previous section, I estimated the usage $v(\mathbf{s})$ generated by animals performing trips from every haulout cluster at every point in space for which environmental data were available. However, environmental data do not exist for all points in space. This means that an unknown proportion of usage is outside the range of the environmental data and therefore $v(\mathbf{s})$ cannot directly be scaled to an estimate of absolute density (eq. 5.6). To overcome this problem, I first predicted $v(\mathbf{s})$ using a simpler model containing distance to the central-place and depth only, values for which are available throughout the study area. I standardised $v(\mathbf{s})$ over space such

that $\int\limits_{All\,\mathbf{s}} u(\mathbf{s})d\mathbf{s} = 1$. I then used this model to obtain a crude estimate of the

proportion of overall usage that fell outside the geographical range of the environmental variables. I found this to be equal to 0.12. Finally, I estimated $\upsilon(\mathbf{s})$ using the full model, weighted this estimate by 1-0.12 and used the predictions of $\upsilon(\mathbf{s})$ generated by the simpler model for those places in space for which environmental data on sediment type and temperature stratification was missing. Using these estimates of $\upsilon(\mathbf{s})$, the number of individuals associated with each haulout site (see eq. 5.3) and the proportion of time spent at sea (61.3%), I calculated the number of individuals expected to be in any given km^2 (see eq. 5.7 and Fig. 5.10a). It is often difficult to interpret the absolute scale of a variable using colour plots alone. I therefore plotted one possible realization of the distribution of individuals that would be obtained from a heterogeneous spatial Poisson process with rate equal to the expected number of individuals per unit area (Fig. 5.10b).
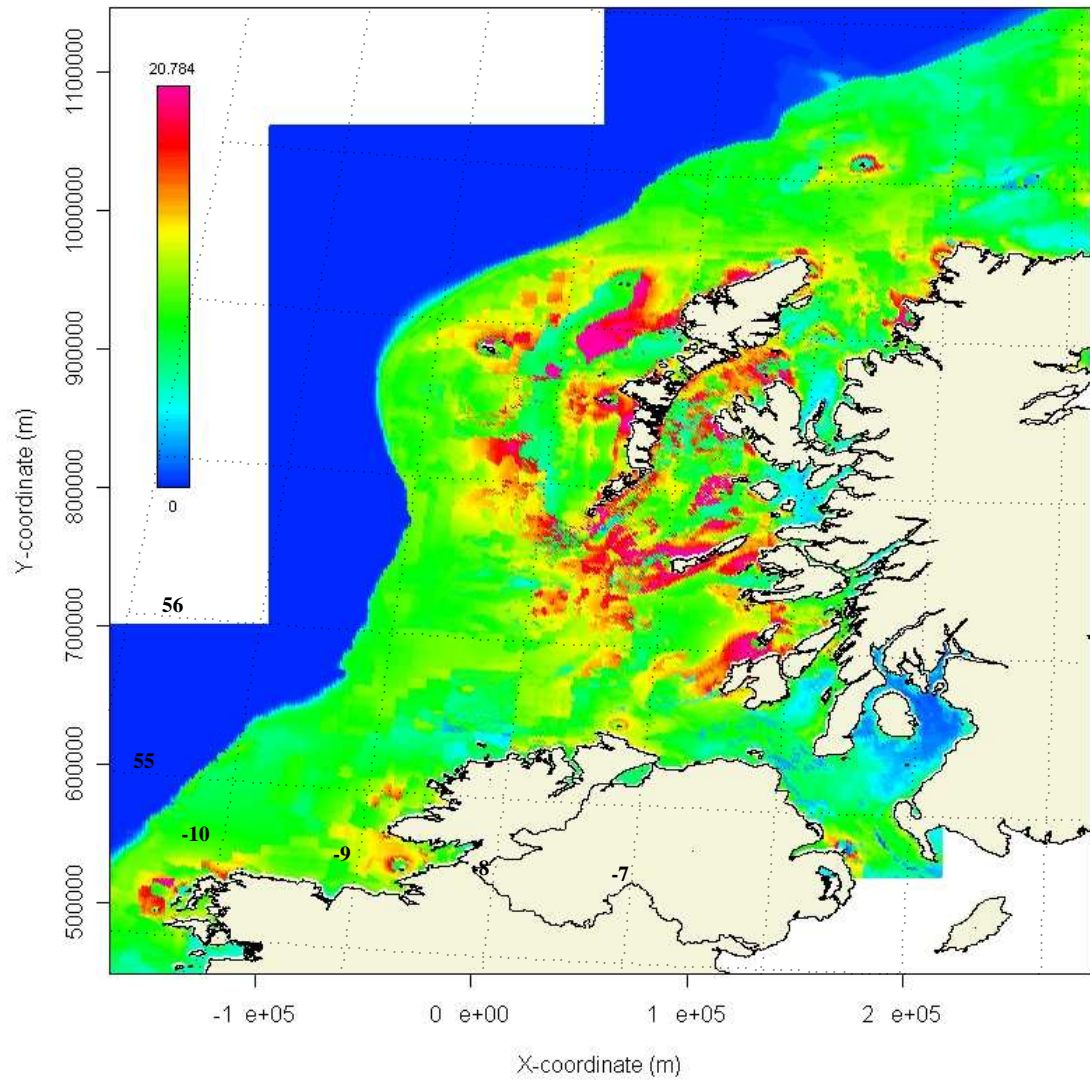
a.
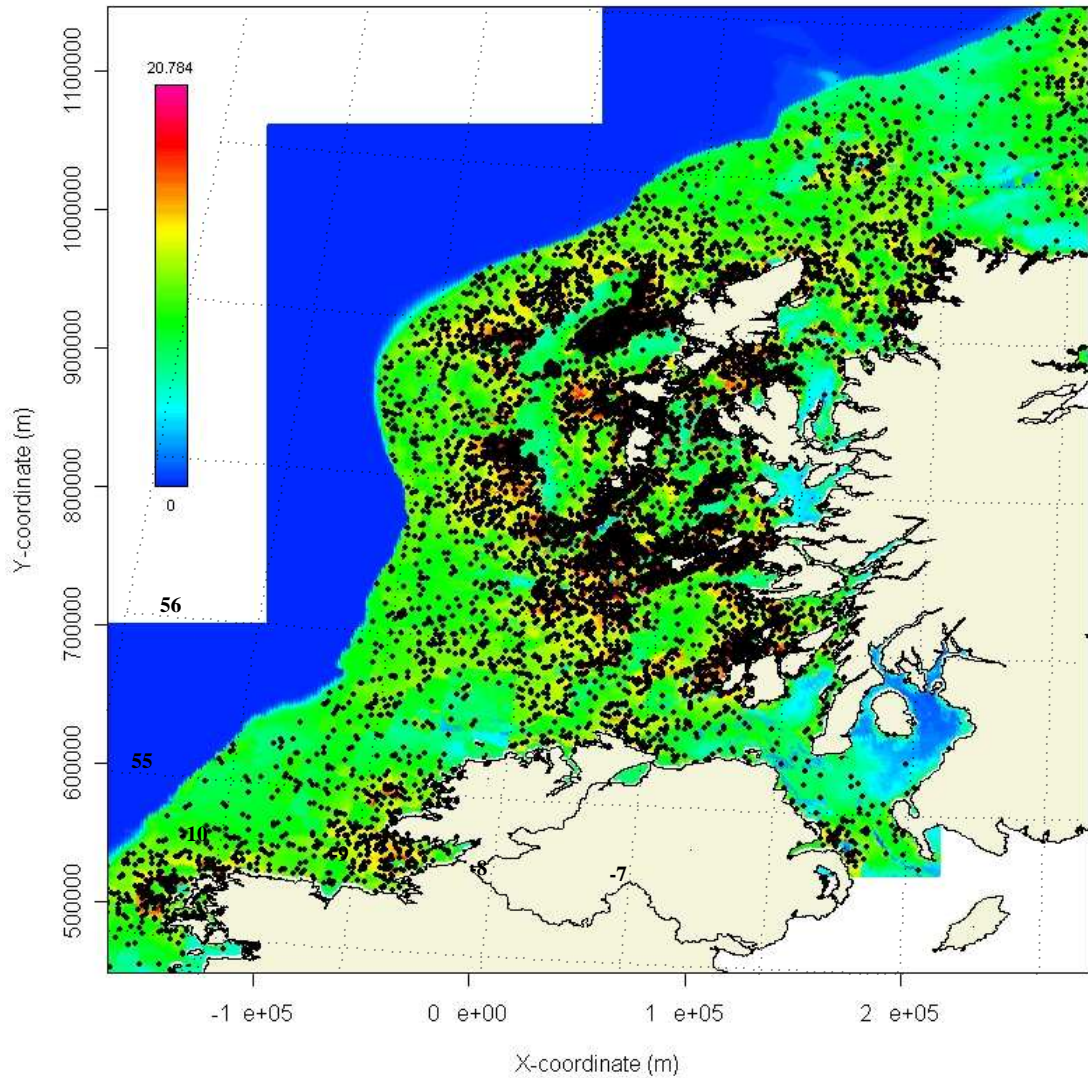
b.

Fig 5.10 Spatial prediction of the average density of individuals (km⁻²) on August 15th 2003 (**a**) and one possible realization of the distribution of all individuals on that day (**b**) plotted on top.

The predictions of usage might be incorrect due to model mis-specification. Fig. 5.8 shows that that there is spatial autocorrelation in the residuals, indicating the model may be mis-specified (e.g. too little flexibility or a missing covariate). To map the geographical regions in which over- or under-prediction occurs, I used the variogram (Fig. 5.8), fitted a spherical model to it (intercept = 0.095, sill = 0.13 and range = 46308) and kriged the residuals (Fig. 5.11)
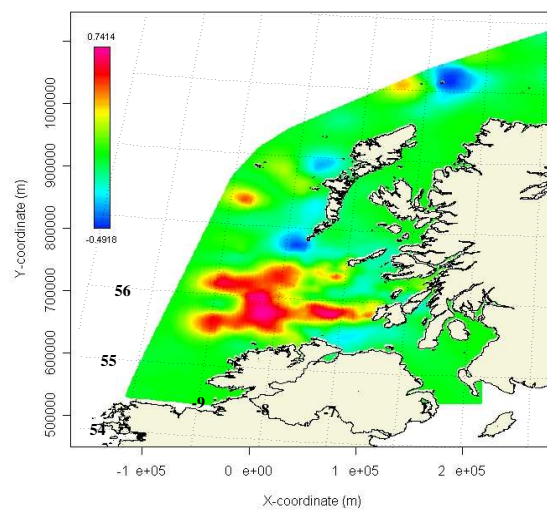
a.
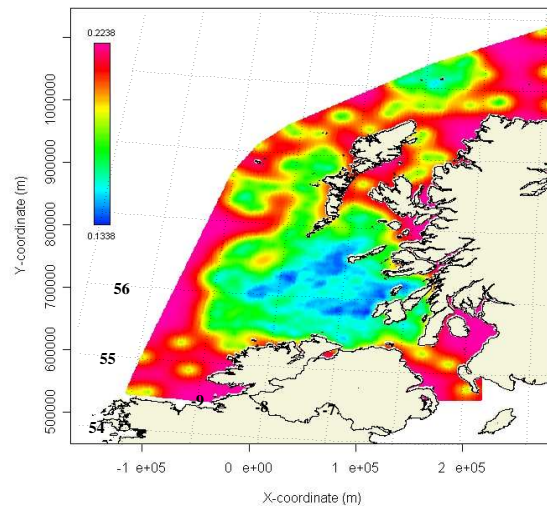
b.

Fig 5.11 Kriged residuals (**a**), red indicating model underprediction and blue indicating model overprediction and variance of kriged residuals (**b**).

To indicate which areas grey seals would use if all points in space were equally accessible, I plotted preference in space (Fig. 5.12).
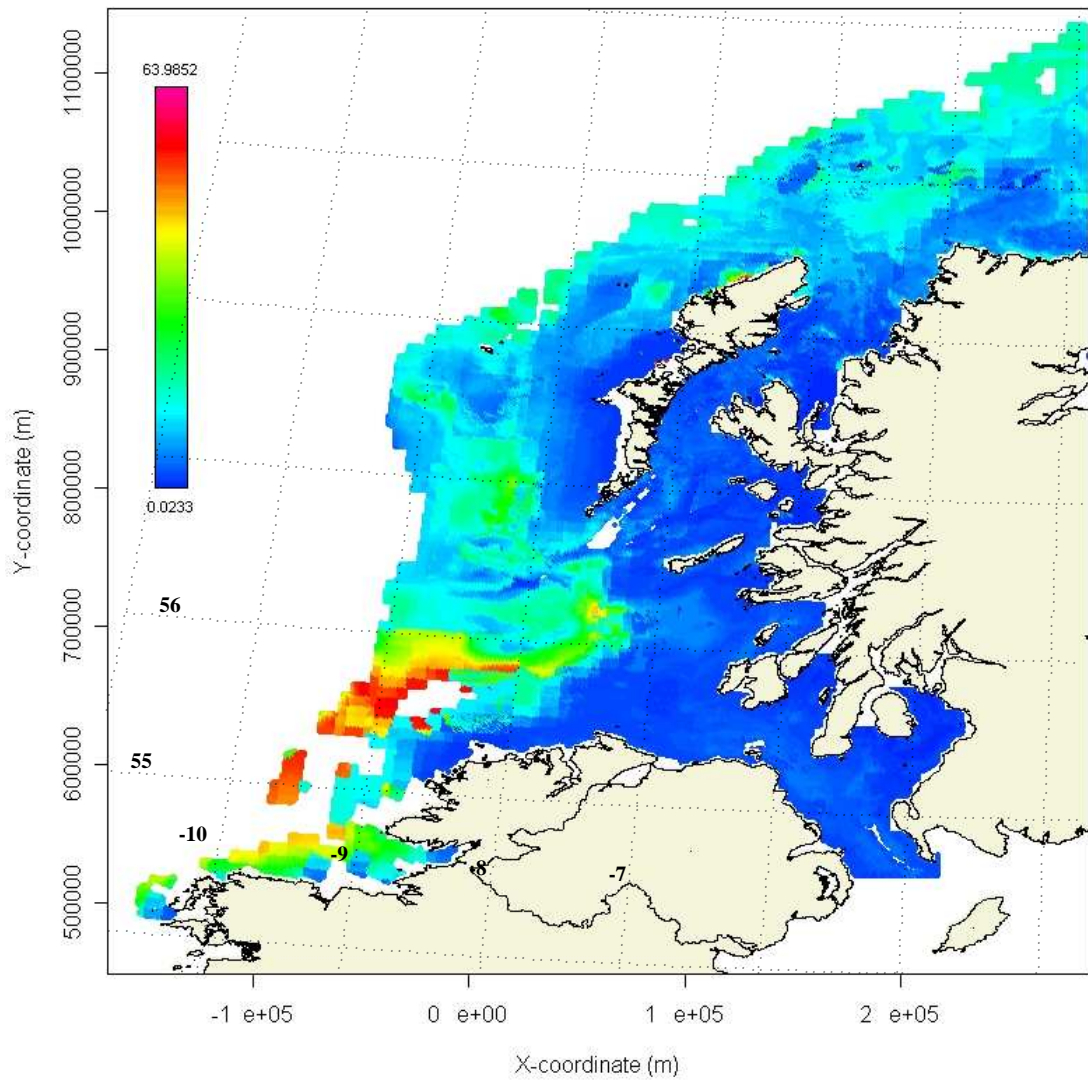
Fig 5.12 Spatial prediction of preference accounted for unequal accessibility and the effect of distance to the haulout site on the response.

5.4 Discussion

5.4.1 Habiat preference; Why are west coast Grey seals where they are?

The candidate environmental covariates used in this study (12 in total) provide a description of the biological and physical processes and characteristics of both the benthic and pelagic components of the shelf-sea west of Scotland. Although the choice of these variables was largely driven by data availability, they were also chosen on the basis of existing scientific information about marine ecological processes that might be important and the biology of the grey seal in particular (see §5.2.2).

The first most important covariate is distance to the central-place, which indicates that grey seals avoid areas far away from the haul-out site more than estimated by the accessibility model.

The relationship between the response variable and depth, the second most important covariate, shows that grey seals tend to avoid deeper areas. Grey seals predominantly feed on or close to the bottom and do have a limited oxygen store capacity. In deeper regions, they spend more time travelling to the bottom which reduces the effective forage duration (Thompson and Fedak 2001), which might explain the negative relationship between the response and depth. In contrast, the study on the east coast of Scotland did not show a strong effect of depth on the response. There, it was postulated (see §3.4), that this was due to the lack of deep areas in close proximity of the haulout such that the physiological restrictions were less likely to have a significant effect.

The third most important covariate was temperature stratification expressed as the difference in temperature between the surface and bottom. Some studies, mostly in oceanic systems, have shown that marine organisms including top marine predators such as mammals and birds are found close to frontal systems (Olson & Brackus 1985, Schneider 1990, Baumgartner et al. 2001, Daunt et al. 2003, Franks 1992). However, Spear et al. (2001) as well as others (Hunt 1990, Turner & Dagg 1983, Hunt et al. 1990, Gould & Piatt 1993, Reilly & Fiedler 1994) have shown that the distribution of higher predators was more strongly related to existence of strong temperature

gradients in the vertical, rather than the horizontal direction. This might suggest that total productivity is less important than the aggregation of phyto- and zooplankton in high densities such that species of fish can efficiently exploit them. Although sandeels (mostly 1+ year-old individuals) do burrow in the sediment, they predominantly forage in the water column. The aggregation of their food sources into tight layers might facilitate feeding and therefore increase their growth, survival and reproduction.

Finally, there is concensus between the results of this chapter and chapter 3 in that grey seals prefer coarse sediment. In addition to the fact that sediment type might influence the grey seals' bioturbating foraging strategy (see §3.2.2), sandeels (comprising 60% of their diet in the west coast) are also known to prefer coarse sediment. The consistent relation between sediment type and grey seal distribution on both the east and west coast of Scotland is an interesting result. Although the sandeel distribution is said to vary considerably in both space and time (Wright et al. 2000) making it almost impossible to use as a covariate in a regression model, surrogates such as sediment type measured at a fine spatial resolution might be a more useful alternative.

In contrast to the environmental variables mentioned above, some covariates explain very little of the variation in the response, even though they are measured at a fine spatial and temporal resolution. This is especially true for the satellite-derived variables such as chlorophyll concentration, sea-surface temperature and spatial gradients in surface temperature, which relate to primary productivity and phytoplankton biomass. This is not unexpected given that seals are 2 trophic levels removed from zooplankton and 3 trophic levels removed from phytoplankton, implying that the correlation between plankton and seal abundance should be weak.

5.4.2 Spatial predictions of usage and preference; Where else are west-coast grey seals likely to be

The final objective of this chapter was to estimate the at-sea distribution of the population of grey seals within the area of interest. Using telemetry data

alone to do so, would strongly bias the estimates to those regions from which animals were tagged. For many regions no telemetry data is available. This problem was recognised by Matthiopoulos et al. (2004) who used model-supervised kernel smoothing; a technique that supports density estimation from the telemetry observations with auxiliary information such as the accessibility of each point in space relative to a particular haulout site as well as estimates on the numbers of individuals associated with that haulout site. A major limitation of this approach as well as all methods that are not based on covariate modelling is that estimates in areas for which there are few or no telemetry observations, are poor.

The major contribution of this chapter to estimating usage in the west coast is that I investigated the relationship between the distribution of grey seal and environmental covariates first, and then used those relationships to predict usage at different points in space. So, although this is technically an extrapolation in geographical space, in practice it is actually interpolation in environmental space. I have shown by means of cross-validation that these models fitted in one region are capable of significantly explaining much of the observed variation in usage elsewhere.

However, under certain circumstances, predictions can be incorrect. This could be the result of measurement error in the environmental data. Generally, the precision of measurements on environmental covariates can be derived from the variability in repeated, local measurements and the spatial distribution of sampling stations. However some of my covariates (such as sediment type and depth) were stored as polygon or poly-line files and one cannot derive the actual underlying sampling distribution. The biggest haulout-cluster on the West coast of Scotland is the Monarch Isles (cluster 12 in Fig 5.3b). According to the BGS sediment survey these areas are predominantly characterized by the presence of rock. However, some finer-scale survey contained within the UKHO data set show that this area shows much more variability than suggested by the BGS dataset. As a consequence, the predictions in these regions might be incorrect.

A second reason why predictions elsewhere can be poor is if habitat preference changes with changes in absolute availability of habitats. For example, this could happen if a particular type of habitat becomes less

available and so does a particular species of fish that depends on it, potentially causing grey seals to switch to alternative prey. Diet studies based on scat samples show strong regional differences in prey consumption (Hammond et al. 1994, SMRU unpublished). However, this might be a consequence of an increased encounter rate and consumption of locally available prey, rather than an active redistribution to different habitats. Studies on the east coast and west coast of Scotland as well as a study of grey seals foraging from Sable island, Canada (Bowen et al. unpublished), have shown a consistent preference for coarse sediment substrates. This suggest that the effects of differences in absolute availability might be relatively small.

Finally, predictions outside the range of the environmental data for which the model was constructed (extrapolation in environmental space) can be extremely unreliable. In this chapter, the areas from which most telemetry data come from, contain relatively little rock. Many seals haulout on the outer Hebrides (clusters 10-14). This area is characterized by the presence of rocky substrate which might be the explanation for the relatively poor predictions in this region.

5.4.4 Future research:

Improving the accessibility model.— Because grey seals are central place foragers, points close to the haul-out site are more likely to be used than those far away. I accounted for the unequal accessibility of different points in space using a pre-specified function of the distance of those points relative to a haul-out site. This model is a simplified version of the underlying individual movement model (Matthiopoulos 2003) and has some limitations. A first shortcoming is that it does not capture some of the variation in null-usage such as the increased usage in a channel leading out to a big stretch of sea. Secondly, the current model considers return trips only. Including transitory trips would require knowledge of the full matrix of transition probabilities between any two haulouts. This may be a complicated function of the distance between, and attractiveness of individual haulout sites. Finally, the accessibility model does not include error in the telemetry

locations. Error, can cause a telemetry observation (or an estimated location generated from track-smoothing) to appear on land and therefore to be removed from the analysis. As this is more likely when the animal is nearer to land, it may incorrectly lead to a data set containing relatively fewer locations close to land.

Although this chapter used a larger number of candidate covariates compared to chapters 3 and 4, distance to the central-place was again the first covariate to be retained in the model. This not only implies that the accessibility model is not describing the true movement processes, it also means that its effect is more important than any of the other covariates used in these studies. Considerable work has been put into collating environmental variables, but these results suggest that perhaps more effort should be put into addressing the inadequacies of the current accessibility model. This could be done by constructing different movement models (perhaps with the aid of GPS telemetry devices) and next to use model information criteria to select among these models in a similar way as is currently done for the environmental covariates. See §6.1.2 and §6.3.1 for more extensive discussion on this.

Correctly defining stratification.— The covariate temperature stratification explained a considerable amount of the variation in spatial usage, even though the variable was estimated at a relatively coarse resolution (9km) and was based on model predictions rather than empirical collected data. Also, the variable was specified as the difference in temperature between the surface and sea-bottom, rather than the temperature gradient at the thermocline (the separation between mixed and stratified water) where most aggregation of phyto- and zooplankton is believed to take place (Turner & Dagg 1983). Correctly specifying the level of prey aggregation, requires oceanographic data collected at a fine spatial and temporal resolution, which is expensive. However, the SRDLs used in this study do measure temperature profiles as well. The major shortcoming of this data is that its collection is not uniformly distributed in space and might be biased towards certain oceanographic features (as suggested by this study). In contrast, a major advantage is that data collection is focussed in those areas that are

important for grey seals. As previously discussed (also see Spear et al. 2001), the underlying process (i.e. aggregation of prey) might influence the spatial distribution of many species of top predators. Therefore, data collected by temperature sensors on seals might not only improve our understanding of the spatial distribution of grey seals, but also of other marine mammals and birds.

5.5 Acknowledgements

5.6 References

Allen, J.I., Blackford, J., Holt, J., Proctor, R., Ashworth, M. and Siddorn, J. 2001. A highly spatially resolved ecosystem model for the North West European Continental Shelf. Sarsia **86**: 423-440

Argos (1989). Guide to the Argos System. Argos CLS, Toulouse.

Baretta, J. W., Ebenhöh, W., Ruardij, P. 1995. The European Regional Seas Ecosystem Model (ERSEM), a complex marine ecosystem model. Netherland Journal of Sea Research, 33 (3/4), 233-246

Baumgartner, M.F., Mullin, K.D., May, L.N., and T.D. Leming. 2001. Cetacean habitats in the northern Gulf of Mexico. Fishery bulletin **99**(2): 219-239.

Burrough, P.A. 1986. Principles of Geographical Information Systems for Land Resources Assessment, Chapter 3. Clarendon Press Oxford.

Daunt, F., Peters, G., Scott, B., Gremillet, D., and S. Wanless. 2003. Rapid-response recorders reveal interplay between marine physics and seabird behaviour. Marine Ecology Progress Series **255**: 283-288.

Franks, P.J.S. 1992. Sink or swim: accumulation of biomass at fronts. Marine Ecology Progress Series **82**: 1-12

Gould, P.J., and J.F. Piatt. 1993. Seabirds of the central North Pacific. In: Vermeer, K., Briggs, K.T., Morgan, K.H., and D. Siegel-Causey (eds). The status, ecology, and conservation of marine birds of the North Pacific. (Spec. Publ.) Canadian wildlife survice, Ottawa, Canada: 27-38.

Fedak, M. A., and B. J. McConnell. 1993. Observing seals by satellite. NERC news:26-28.

Folk, R.L. 1980. Petrology of Sedimentary Rocks. Hemphill, Austin, 187 p.

Goold, J. C. 1996. Acoustic assessment of populations of common dolphin Delphinus delphis in conjunction with seismic surveying. Journal of the Marine Biological Association of the United Kingdom **76**:811-820.

Goold, J. C., and P. J. Fish. 1998. Broadband spectra of seismic survey air-gun emissions, with reference to dolphin auditory thresholds. Journal of the Acoustical Society of America **103**:2177-2184.

Hammond, P. S., A. J. Hall, and J. H. Prime. 1994. The diet of grey seals in the Inner and Outer Hebrides. Journal of Applied Ecology **31**:737-746.

Hammond, P. S., P. Berggren, H. Benke, D. L. Borchers, A. Collet, M. P. Heide-Jorgensen, S. Heimlich, A. R. Hiby, M. F. Leopold, and N. Oien. 2002. Abundance of harbour porpoise and other cetaceans in the North Sea and adjacent waters. Journal of Applied Ecology **39**:361-376.

Harwood, J., and B. Wilson. 2001. The implications of developments on the Atlantic Frontier for marine mammals. Continental Shelf Research **21**:1073-1093.

Holland, G. J., S. P. R. Greenstreet, I. M. Gibb, H. M. Fraser, and M. R. Robertson. 2005. Identifying sandeel Ammodytes marinus sediment habitat preferences in the marine environment. Marine Ecology-Progress Series **303**:269-282.

Holt, J. T., and I. D. James. 2001. An s coordinate density evolving model of the northwest European continental shelf - 1, Model description and

density structure. Journal of Geophysical Research-Oceans **106**:14015-14034.

Hunt, G.L Jr. 1990. The pelagic distribution of maine birds in a heterogeneous environment. Polar Research **8**: 43-54

Hunt, G.L. Jr, Harrison, N.M., and T. Cooney. 1990. The influence of hydrographic structure and prey abundance on foraging of least auklets. Studies in Avian Biolog **14**: 7-22

Jepson, P. D., M. Arbelo, R. Deaville, I. A. P. Patterson, P. Castro, J. R. Baker, E. Degollada, H. M. Ross, P. Herraez, A. M. Pocknell, F. Rodriguez, F. E. Howie, A. Espinosa, R. J. Reid, J. R. Jaber, V. Martin, A. A. Cunningham, and A. Fernandez. 2003. Gas-bubble lesions in stranded cetaceans - Was sonar responsible for a spate of whale deaths after an Atlantic military exercise? Nature **425**:575-576.

King, J. E. 1983. Seals of the world, 2 edition. Oxford University Press, Oxford.

Lunneryd, S. G. 2001. Fish preference by the harbour seal (Phoca vitulina), with implications for the control of damage to fishing gear. Ices Journal of Marine Science **58**:824-829.

Macleod, K., M. P. Simmonds, and E. Murray. 2003. Summer distribution and relative abundance of cetacean populations off north-west Scotland. Journal of the Marine Biological Association of the United Kingdom **83**:1187-1192

Matthiopoulos, J. 2003a. Model-supervised kernel smoothing for the estimation of spatial usage. Oikos **102**:367-377.

Matthiopoulos, J. 2003b. The use of space by animals as a function of accessibility and preference. Ecological Modelling **159**:239-268.

Matthiopoulos, J. 2004. The use of space by animals as a function of accessibility and preference. Ecological Modelling **159**:239-268.

McConnell, B. J., C. Chambers, K. S. Nicholas, and M. A. Fedak. 1992. Satellite Tracking of Gray Seals (Halichoerus-Grypus). Journal of Zoology **226**:271-282.

McConnell, B. J., M. A. Fedak, P. Lovell, and P. S. Hammond. 1999. Movements and foraging areas of grey seals in the North Sea. Journal of Applied Ecology **36**:573-590.

Olson, D.B., R.H. Backus. 1985. The concentration of organisms at fronts: a cold-water fish and a warm-core Gulf Stream ring. Journal of Marine Research **43**: 113-137.

Otter Research Ltd. 2004a. An introduction to AD Model Builder Version 7.1.1. For use in Non-linear Modelling and statistics. Otter Research Ltd, Sidney B.C.

Otter Research Ltd. 2004b. Random effects in AD Model Builder. ADMB-RE user guide. Otter Research Ltd, Sidney B.C.

Pinheiro, J. C., and D. M. Bates. 2000. Mixed-Effects Moedels in S and S-PLUS. Springer, New York.

Proctor, R., and I. D. James. 1996. A fine-resolution 3D model of the Southern North Sea. Journal of Marine Systems **8**:285-295.

R Development Core Team 2004. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, URL http://www.R-project.org.

Read, A. J., and C. R. Brownstein. 2003. Considering other consumers: Fisheries, predators, and Atlantic herring in the Gulf of Maine. Conservation Ecology **7**:art. no.-2.

Reilly, S.B., and P.C. Fiedler. 1994. The interannual variability of dolphin habitats in the eastern tropical pacific 1. Reseach vessel surveys, 1986-1990. Fishery Bulletin **92**: 434.

Schneider, D.C. 1990. Seabirds and fronts: a brief overview. Polar Research **8**: 17-21

Sea Mammal Research Unit. 2002. Scientific Advice on Matters Related to the Management of Seal Populations. http://smub.st-and.ac.uk/CurrentResearch.thm/02.pdf. Sea mammal Research Unit, St Andrews, UK

Spear, L. B., L. T. Ballance, and D. G. Ainley. 2001. Response of seabirds to thermal boundaries in the tropical Pacific: the thermocline versus the Equatorial Front. Marine Ecology-Progress Series **219**:275-289.

TerBraak, J.F., Hoijtink, H., Akkermans, W., and P.F.M. Verdonschot. 2003., model-based cluster analysis for predicting macrofaunal communities. Ecological Modelling, Volume **160 (3)**: 235-248

Thompson, D., and M. A. Fedak. 2001. How long should a dive last? A simple model of foraging decisions by breath-hold divers in a patchy environment. Animal Behaviour: 287-296.

Turner, J.M., and Dagg M.J. 1983. Vertical distributions of continental shelf zooplankton in stratified and isothermal waters. Biological Oceanography **3**: 1-40

Wickens, P. A., D. W. Japp, P. A. Shelton, F. Kriel, P. C. Goosen, B. Rose, C. J. Augustyn, C. A. R. Bross, A. J. Penney, and R. G. Krohn. 1992. Seals and Fisheries in South-Africa - Competition and Conflict. South African Journal of Marine Science-Suid-Afrikaanse Tydskrif Vir Seewetenskap **12**:773-789.

Wood, S. N. 2001. mgcv: GAMs and Generalized Ridge Regression for R. R News **Vol 1**:20-25.

Wursig, B., and C. R. Greene. 2002. Underwater sounds near a fuel receiving facility in western Hong Kong: relevance to dolphins. Marine Environmental Research **54**:129-145.

Wright, P. J., H. Jensen, and I. Tuck. 2000. The influence of sediment type on the distribution of the lesser sandeel, Ammodytes marinus. Journal of Sea Research **44**:243-256.

Appendix 5.A

Counts of grey seals on haulout sites are available for August 1996 and 1997, including a total of 483 haul-out sites (Fig 5.3). These haul-out sites are divided into clusters on the basis of their spatial position and number of individuals using the following algorithm:

1) Designate every haulout site as a unique cluster, (i.e. create the cluster data set)

2) Remove one haulout from the cluster data set.

3) Calculate for each individual from each haulout the squared Euclidean distance to the nearest cluster

4) Calculate the sum of squared distances for all individuals

5) Place this haulout back in the cluster data set

6) Repeat steps 2-5 for all other haul-out sites.

7) Select that combination of *n-1* clusters with the lowest squared distance value. In practice, this means that a haulout with both a low number of individuals and in close proximity to another haulout site is most likely to be incorporated into a neighbouring cluster.

8) Repeat steps 2-7 until the required number of clusters are left.

6.   GENERAL DISCUSSION

The recognition of the importance of space in shaping population dynamics (Tilman and Kareiva 1997, Bolker 2004) and ecological interactions (Hilborn 1975, McLaughlin and Roughgarden 1992, Holmes et al. 1994, Jansen 1995, Farnsworth and Beecham 1997) has greatly advanced the development of techniques for the collection of data on species distribution. For transect data, this increase in activity has been matched by the development of analytical techniques (Buckland et al. 2001, Buckland et al. 2004).

Modelling of telemetry data has been a less active area of research mainly for two reasons: first, work has been dogged by multiple, conflicting interpretations of model components as crucial as the response variable and second, none of the off-the-shelf statistical frameworks can address the large number of problems encountered when working with telemetry data.

Many applied studies proceed with the analysis of telemetry data without acknowledging these problems and ambiguities. Although this may not always affect their results, there is nevertheless a distinct risk that conservation and management of some populations is being based on false premises. Therefore, the four most important contributions of this thesis are 1) to highlight these conceptual ambiguities and propose a functional and consistent framework of definitions for terms old and new, 2) to trace the evolution of statistical analysis of space-use and habitat preference, 3) to enumerate, describe and propose solutions for eight fundamental challenges of analysing telemetry data and 4) provide illustrations for the practitioner using real case studies.

The framework brings together case-control, mixed-effects and generalized-additive modeling. I believe that this combination best addresses the problems listed in chapter 2 and is also flexible enough to benefit from forthcoming advances in computer speed, estimation software and statistical methodology.

6.1 Empirical models of usage and preference

The presentation of the framework was structured around the problems encountered when analyzing telemetry data. I opted for this less conventional presentation because I assumed that most practitioners would be familiar with the fundamental components of empirical modeling (response variable, explanatory variables, model structure, parameter estimation, model selection, validation and prediction). However, at this point, it is useful to collect my insights of each component in the context of telemetry studies:

6.1.1 Response variable

Obtaining a response variable from wildlife telemetry data is challenging because, by definition, they only provide information about the presence of animals. Some studies (Drake et al. 2006) have regressed the density of telemetry observations against environmental covariates, a technique known as niche-based analysis. This method makes the strict assumption that sampling effort is uniformly distributed in space and ignores variations in the relative availability of different habitats. In practice, even if an animal showed no preference, it would still be observed more frequently in more abundant habitats. This has long been recognized by the habitat-preference literature and has been resolved by classifying environmental space into habitats and correcting usage for their relative availability in geographical space (Manly et al. 1993).

My work and other, related, studies (Boyce et al. 1999), have replaced habitat classification by a case-control design, allowing the model to capture the finest variations in usage, within the limits of computational power. This causes some ambiguity in the interpretation of the response variable in the

resulting models. Keating & Cherry (2004) have shown that the response variable in a case-control design is not proportional to usage. Here, I have shown how usage can be estimated from it.

Confusion also exists about the total number and position of controls that are required under the case-control design. Anecdotal advice is to use a number of controls equal to the number of telemetry observations. However, the objective of the controls is to reflect the relative availability of habitats in space. A larger number of zeros leads to a more accurate representation of the environment. Indeed, Prentice and Pyke (1979) showed that all model parameters except the intercept remain unaffected by the number of controls as long as a sufficiently large sample is taken.

The positions of the controls can be selected uniformly randomly from within the postulated range of the animals. It may also be possible to use biological knowledge to account for the unequal accessibility of points in space, as was done in this study. Different sampling designs ($a(\mathbf{s})$) will produce comparable estimates of usage $v(\mathbf{s})$ provided a sufficiently large sample of control data is included in the analysis (eq. 2.7). Using a model of accessibility is preferable for large data sets because it can direct more computational power to those areas that are more likely to be visited by the animals.

6.1.2 Covariates

Environmental variables.—— Including too many covariates in a regression model can lead to 'a subjective and iterative search for data patterns and significance' (Burnham and Anderson 2002). Statistical inference might, instead, be conducted among models with a small number of covariates that are believed to be proximately related to the response variable. However, excluding particular combinations of covariates *a priori*, is no less subjective. If, contrary to biological intuition, a covariate is retained by out-competing others in the model selection process, this would hint at an ecological process that was not previously considered and thus enhance our understanding of the mechanisms underlying the distribution of the species. Given the

exploratory nature of most telemetry studies it is perhaps better to examine as many candidate covariates as permitted by sample size and computer power, and allow issues of parsimony to be dealt with entirely by model selection.

Accessibility.—— Points in space are not necessarily equally accessible to all animals and relating wildlife telemetry observations with environmental variables may incorrectly suggest a preference for those habitats that happen to be more accessible. Limited effort has been directed at accounting for this effect in habitat preference studies (Arthur 1996, Hjermann 2000, Matthiopoulos 2003).

Correctly defining accessibility is complicated, because it depends on the individual's physiological restrictions, life-history and spatial perspective of its environment. Matthiopoulos et al. (2004) simulated individual movement of grey seals in the absence of preference and showed that, for their data, the resulting surface of null-usage was closely approximated by a simple function of at-sea distance from the haulouts. For the case study on the east coast of Scotland (Chapter 3), an area characterized by relatively few obstacles to seal movement, this approximation was good. In contrast, the west coast of Scotland is characterized by a more complex topography comprising many islands and peninsulas. Because grey seals appear to memorize their environment and follow specific short-distance routes (Thompson et al. 1991), spend a large proportion of time in proximity to haul-out sites (an obstacle in itself) and frequently move between them, the simple function of distance which quantifies the accessibility of points in space, might be insufficient. This can cause biases in the parameters of the model. One solution to the problem might be to simulate movement under multiple scenarios (e.g. movement of individuals with or without spatial memory), and to use all maps of null-usage as candidate covariates in the model. This idea and extensions are discussed further in § 6.3.1

6.1.3 Model structure.

The binary, case-control data were modeled using a logit link with a Bernoulli likelihood. In wildlife telemetry only a few animals are sampled and observations within individuals are more likely to occur at similar environmental conditions than between individuals. To capture this hierarchical structure in the data, I've used a mixed-effects model (Pinheiro and Bates 2000). This type of model does not only estimate the amount of between-individual variability, a necessity for making population level inferences, it can also model the effect of individual specific characteristics (such as sex and length) on the observed individual variability. If there is a bias in catching effort towards certain types of individuals (e.g. young males), independent information on population structure can be used to correct these biases and generate balanced predictions of population distribution.

Due to variations in the availability of, and preference for different habitats, response data will always form clusters in particular regions of environmental space. When using linear models with such data sets, these data-rich regions will tend to dominate model estimates and result in biased or imprecise predictions in the rest of environmental space. For example, use of a GAM in my case study indicated that the preference of grey seals is highest for depths of about 80m. Consequently, the final model comprises a positive trend down to depths of 80m and a negative trend below that depth. In contrast, a GLM would have been dominated by data close to shore, at shallower depths, and would have unrealistically predicted a continued positive trend. An advantage of using GAMs together with mixed models is that differences in the functional form of the responses of different individuals (e.g. Figs. 3.10, 4.7 and 5.9) are highlighted.

6.1.4 Model selection and model validation

Detection of preference in early, test-based analyses required a high average number of observations per habitat to maintain sufficient power, i.e. reduce type II errors. Therefore, for a telemetry data set of a given size, it was necessary to divide the observations between a small number of habitats.

This could be achieved either by using a coarse habitat classification scheme, or by a-priori pruning of the dimensionality of environmental space, at the risk of losing important covariates.

The case-control design which makes it unnecessary to discretize environmental space, and the use of model selection, leads to a fundamental different way of drawing inferences. Not only does it provide an answer to whether a covariate significantly contributes in explaining the response (taking model parsimony into account), but more importantly, it provides an objective way of ranking their explanatory power, a feature particularly well illustrated in Figs. 3.5, 4.4 and 5.6.

Model selection is often implemented using information criteria (IC) which penalize the likelihood of a candidate model by the number of parameters it contains. Although there are theoretical justifications for the severity of the penalty (Burnham & Anderson 2002), their validity is sensitive to mis-specification of the likelihood function. For computational reasons, the likelihood of most non-linear models assumes that the data are independent. When, as with telemetry studies, the data are spatially and temporally autocorrelated, model selection by IC leads to over-fitted models. This can be overcome either by modeling autocorrelation as part of the likelihood, or by using alternative approaches, such as cross-validation, for model selection. Cross-validation prevents over-fitting by using one data set for fitting the model and another for assessing its predictive power. In my case-study, cross-validation led to a reduction of the number of environmental and individual-specific covariates, compared to the models suggested by the IC.

6.1.5 Predictions of usage and preference

Spatial predictions can be classified into three categories in order of decreasing reliability; i) predictions for the area and time in which the telemetry data were collected (interpolation), ii) predictions from a different place or time, but for similar environmental conditions as those used to construct the model (geographical extrapolation) and iii) predictions outside

the region of environmental space for which data exist (environmental extrapolation).

Despite being the most reliable, interpolation is still subject to biases in catching effort towards certain types of animals, large individual variation and a small sample size. Mixed-effects models with individual-specific covariates can account for some of these biases and the use of cross-validation ensures that a model fitted to one group of individuals is applicable to other individuals from that same sub-population.

I have also shown that it is possible to predict the distribution of individuals from a different, albeit neighbouring, sub-population (see §3.3.2). This is particularly useful because differences in absolute habitat availability between different geographical regions can weaken predictions of usage (Mysterud and Ims 1998). My ability to predict the distribution of Abertay animals using data from the Farnes was probably the result of similarity in the conditions experienced by these two sub-populations, implying that my predictions were extrapolations in geographical, but not environmental space.

Extrapolations in environmental space are likely to be less reliable. For example, the current grey seal data collected on the east coast of Scotland provide no hint as to the maximum depth that grey seals can dive to. Applied to the west coast of Scotland, this model might predict seal usage beyond the continental shelf, while the case-study carried out in this area indicates that this is not the case. Therefore, extrapolation in environmental space is generally less reliable and is best avoided. However, this may also restrict the geographical coverage of predictions. Because an unknown proportion of population usage lies outside the prediction area, predictions of usage cannot be scaled up to reflect absolute population densities. An approximate solution is to use a simpler model (e.g. a model with distance to the central-place as only covariates) to calculate the proportion of usage that is inside environmental space of the full model and then rescale predictions of usage from the full model by that proportion (see also §5.3.4).

6.2 Habitat preference of two marine top-predators: a comparison accross species and regions

In addition to serving as showcases and validation data sets for the statistical framework, the three case studies in this thesis also permitted a comparison of UK marine usage and preference across species and regions. Data availability varied between regions and only the covariates "distance to the central-place", "depth" and "sediment type" were in common to all three case studies. Hence, although ideally the same set of environmental covariates would be used for all case studies to facilitate comparisons, I placed the emphasis on getting the best model for each data set. Therefore, I did not restrict the west coast analysis to this small set of covariates.

6.2.1 A comparison between grey seals and gannets on the east coast

The results of this study indicate that the distribution of grey seals is both more heterogeneous and predictable than that of gannets. Predictability of usage depends on the relevance and accuracy of the covariates included in the analysis. Diet studies show that grey seals predominantly feed on benthic species of fish such as sandeel and gadoids (Prime & Hammond 1990, Hammond et al. 1994), which often prefer particular types of sediment. Because this environmental characteristic is more static compared to some of the physical oceanographic variables such as temperature, the distribution of these benthic fish species can be reasonably well-predicted (Wright et al. 2000, Holland et al. 2005). In contrast, gannets predominantly feed on pelagic species such as mackerel, herring and 0-group sandeel (Hamer et al. 2000) whose distributions are patchy and dynamic and consequently, gannets might spend more time moving randomly in space in search of their food resources.

Another reason why the gannet distribution is less spatially heterogeneous than the grey seals, could be the result of density dependent effects in consumption. Lewis et al. (2001), showed evidence for density dependent effects in foraging from larger colonies. Gannets aggregate on land in sufficiently large numbers to induce depletion of the 'good' foraging areas near the colony. As a consequence, the distribution of prey which is normally heterogeneous in space, might start to approximate uniformity.

However, this explanation relies on the untested assumption that the rate of regeneration of prey, migration and growth isn't also spatially heterogeneous.

There are some interesting, independent observations of the interactions between the top-predators that could be used in the search for insights. Thompson et al. (1991) observed an individually-tracked grey seal repeatedly moving directly towards the location of feeding seabird aggregations that included gannets. Others have also observed aggregations of grey seals and gannets off the east coast of Scotland (Simon Greenstreet pers. comm.). It is possible that grey seals and gannets both respond to environmental covariates not included in this study. The case-study conducted on the west coast of Scotland shows that grey seals have a strong preference for stratified water. Studies on the distribution of oceanic sea birds (Hunt 1990, Turner & Dagg 1983, Hunt et al. 1990, Reilly & Fiedler 1994, Spear et al. 2001) have also found that areas of stratified water were preferred to areas of well-mixed water. So, perhaps a measure of stratification that was missing from the east coast case studies in this thesis is an important potential covariate for future work.

6.2.2 A comparison between east and west coast seals

It has been postulated (e.g. see §6.1.5), that changes in the absolute availability of habitats may lead to changes in estimated preference. Understanding the impact of variability in habitat availability is important because it determines the validity of predictions in other regions of space. In this thesis, I investigated environmental preference and space use of grey seals from both the east and west coast of Scotland. Although these regions differ substantially in terms of the common covariates "sediment type" (Fig. 3.4a and Fig. 5.5k) and "depth" (Fig 3.4b and Fig. 5.5l), seals showed similar patterns of preference.

In particular, grey seals were shown to prefer coarse substrates in both regions. This finding isn't unexpected given that sandeels prefer coarse substrates (Wright et al. 2000, Holland et al. 2005) and they are the seals' most important prey species in both regions (Hammond & Prime 1990,

Hammond et al. 1994, SMRU unpublished). In contrast, while depth is the second most important covariate in shaping the distribution of grey seals on the west coast, it is the last variable to enter the east coast model. This might be because grey seals on the east coast are not exposed to sufficiently deep areas for depth to reveal itself as an important covariate. So, this result is primarily the result of sparse data to support this trend (i.e. there is very large variability in this region of the covariate, see Fig. 3.10d).

6.3 Wider context

6.3.1 Empirical v mechanistic modelling

Science is the organized body of knowledge accumulated through observation, experimentation, analysis and modelling (verbal or quantitative). This is an iterative process of proposals, refutations and improvements. In this thesis, I relied extensively on this knowledge by using recent technical advances in data collection and statistical analysis, but also used natural history information about which covariates could affect the spatial distribution of the species. My approach to modelling was primarily empirical because it relied on regression techniques. I chose empirical modelling because there was not sufficient scientific knowledge to enable me to adopt a mechanistic approach. Hence, I used flexible models (GAMMs) because they are ideal for modelling unknown non-linear relationships.

However, using such flexible models also carries two penalties: First, the outcomes of model fitting and selection can be sensitive to stochasticity. Especially when multi-collinearity is present, the explained deviance of two or more explanatory variables can be similar and which one gets selected will be largely driven by stochasticity. Therefore, given scientific publications alone, the use of flexible empirical models hampers comparisons between published inferences from different studies. Second, environmental extrapolation is based on pattern instead of process and is therefore unreliable.

For some species it may be possible to construct more mechanistic models describing some of the causal relationships between space-use and its covariates. For example, in grey seals the results of physiological experiments might be used to construct energetic models of the cost of traveling to particular depths or distances from the haulout (Thompson et al. 1993, Thompson and Fedak 2001). This information could than be translated into mechanistic models.

There are different ways to increase the mechanistic content of a model. The traditional approach is to construct a theoretical model that merges old and new experimental data (collected for that or a similar species) with well-founded ecological first-principles. Generally the model includes all processes that are assumed to be important a priori. The model can then be used to make spatial predictions of usage and those predictions can be validated with actual observations (e.g. wildlife telemetry data). One limitation of this approach is that it can't help identify the importance of other processes (i.e. those covariates that are not included *a priori*), a property which is the distinctive quality of inferential modelling, traditionally associated with regression.

An alternative route, that can be viewed as a convergence between empirical and mechanistic models, is to construct functions describing the anticipated relationship between a covariate or group of covariates and the response. For example, one could construct different movement models based on the isotropic random walk, levy flight or correlated random walk, and use simulations to make spatial predictions of usage, that can then enter the empirical model as candidate covariates. If physiological experiments suggest a specific relationship with depth this could be used as a candidate covariate. Covariates for which a priori suppositions about the structure of the functional form are absent, can enter the model as unspecified smooth functions. The advantage of this approach is that the covariates (e.g. those that are based on different movement models) can naturally compete with one another in model selection. Importantly however, the *a priori* specification of the mechanistic functions means that their parameters are not estimated from the telemetry data.

The reasons why statistical estimation and inference were not traditionally associated with mechanistic models were a) that the error in the observation process could not easily be modelled along with process stochasticity and b) that fitting such models was computationally prohibitive (Harwood & Stokes 2003). These restrictions have been alleviated, but not removed, by the introduction of state-space models (Durbin and Koopman 2001, Johnsen et al. 2003), more powerful computers and new model fitting algorithms such as the Kalman Filter (Brown & Hwang 1992), Monte Carlo Markov Chain (Gilks et al.1996) and Sequential Importance Sampling (Doucet et al. 2001). Adopting the Bayesian approach to state-space modelling has the additional advantage that parameter estimates and uncertainties derived from previous studies can enter the model as parameter priors. Although this is a promising area of research, currently only simple models with few covariates can be fitted, model selection is computationally expensive and population-level predictions rely on individual-based simulation.

6.3.2 Eulerian or Lagrangian models

Eulerian models focus on the density of animals in the neighbourhood of a point in space. Eulerian movement is phrased as the flux of density between neighbouring points in space. In contrast, the Lagrangian approach focuses on individuals as they move across different points in space. Lagrangian movement is generally characterized by velocity, direction and acceleration (Turchin 1998). There is a direct correspondence between Eulerian density and transect data just as there is congruence between Lagrangian trajectories and wildlife telemetry data. Trying to shoe-horn telemetry data into the Eulerian approach, is the cause of many of the problems addressed in this thesis, such as the unequal accessibility of points in space and spatio-temporal correlation of telemetry observations. This prompts the question of why I chose to use a Eulerian model in the first place. I did so firstly because the Eulerian approach considers the density of individuals at both used and un-used locations, while most Lagrangian models consider changes in

movement as a function of *local* environmental characteristics. Lagrangian models consequently, use no information about habitats not visited by the individual. Also, Eulerian approaches are better suited to population-level inferences than the more behaviorally-orientated Lagrangian modeling. They are therefore ideal for large-scale, exploratory studies with an applied agenda.

6.3.3 Hierarchical decisions in space use and habitat preference

The results of my analyses were predicated on the animals' capture locations, but the very fact that an animal was caught at a particular location, was already a consequence of historical processes and decisions that were to some extent, driven by environmental variables. Generally, animals are born at places characterized by favourable environmental conditions that are intended by their parents to increase their fitness. If environmental conditions are more favourable elsewhere, individuals might decide to relocate to a different colony or establish a new territory. At some point in their life, some animals were caught and fitted with a telemetry device, after which they were observed for a relatively short duration. It is only during this period of its life, the telemetry data provide an insight into the individuals behaviour and decisions.

The existence of multiple layers of decisions in space use and preference has been recognized in the literature. Johnson (1980) identifies the existence of a natural ordering of selection processes. "*First-order selection* can be defined as the selection of the physical or geographical range of a species. Within that range, *second-order selection* determines the home range of an individual or social group. *Third-order selection* pertains to the usage made of various habitat components within the home range. Finally, if third-order selection determines a feeding site, the actual procurement of food items from those available at that site can be termed *fourth-order selection*."

Specifically for telemetry data, Aebischer et al. (1993) recommend the investigation of both *second* and *third order selection*. Both investigations rely on a correct quantification of the position and size of individual home ranges, which is not only driven by the distribution of all individuals from

the population during the capture event and the movement of individuals after the capture event, but also by the probability of catching a specific individual from that population. For the position of home-ranges to be a representative sample of the population, the capture probability has to be identical for all individuals. In practice, the capture probability is often driven by the environmental conditions surrounding an animal. For example, it might be easier to catch a particular species in more exposed habitats. Also the distance to urban features (e.g. towns or roads) might increase the probability of capture. As a consequence, the distribution of home-ranges and therefore conclusions about *second* and *third order* habitat preference, will be biased towards those habitats that facilitate capture. Due to this inappropriateness, I suggest a different approach to modelling the hierarchical decisions in space use and environmental preference.

The first expression of environmental preference is not made by the individual in question, but by its parents. Parents decide to produce offspring at a location characterised by particular, most often, favourable environmental conditions. Especially for short-lived, slow-moving species this decision can impose considerable restrictions on the future distribution of that individual. To capture this first habitat selection process, one could investigate the spatial distribution of pup production (e.g. in grey seals) as a function of environmental variables. This somewhat corresponds to the *first order selection* type of investigation proposed by Johnson (1980).

If the offspring survives, it may decide to relocate to a different colony, territory or region of space. These decisions will also largely be driven by local environmental conditions and can be modelled as such. For grey seals the spatial distribution of the number of individuals at haulout-sites could be modelled as function of environmental conditions. I define this as *second order selection*.

From these colonies (or regions of space), some individuals are caught and tagged. The data can then be used to investigate space use and environmental preference, conditional on the individual starting at that capture location. This is the type of investigation carried out in this thesis and closely resembles the *third order selection* defined by Johnson (1980). The only, though major difference is that the approach used in this thesis

does not rely on an arbitrary definition of home-ranges to decide what is accessible and therefore available to the animal.

Finally, as a result of recent technical advances in remote sensing of individual behaviour and physiology (Cooke et al. 2004), it has been possible to remotely measure the rate of food acquisition in some species. For example, Beringer et al. (2004) measure plant consumption by white-tailed deer using real-time video recordings and Bowen et al. (2002) investigate prey indigestion by grey seals. Studies that could not directly measure food acquisition, have recorded behavioural and physiological changes instead. For example, Xaviers et al. (2006) investigated changes in stomach temperature in Wandering Albatrosses, Biuw et al. (2003) measured changes in body composition based on drift dives in Southern Elephant seals and Miller et al. (2004) used hydrophones to measure prey capture attempts. All of these can be considered as studies of *fourth order* selection and can be particularly well formulated as state-space models (Morales et al. 2002, Johsen et al. 2003 and see also §6.3.1).
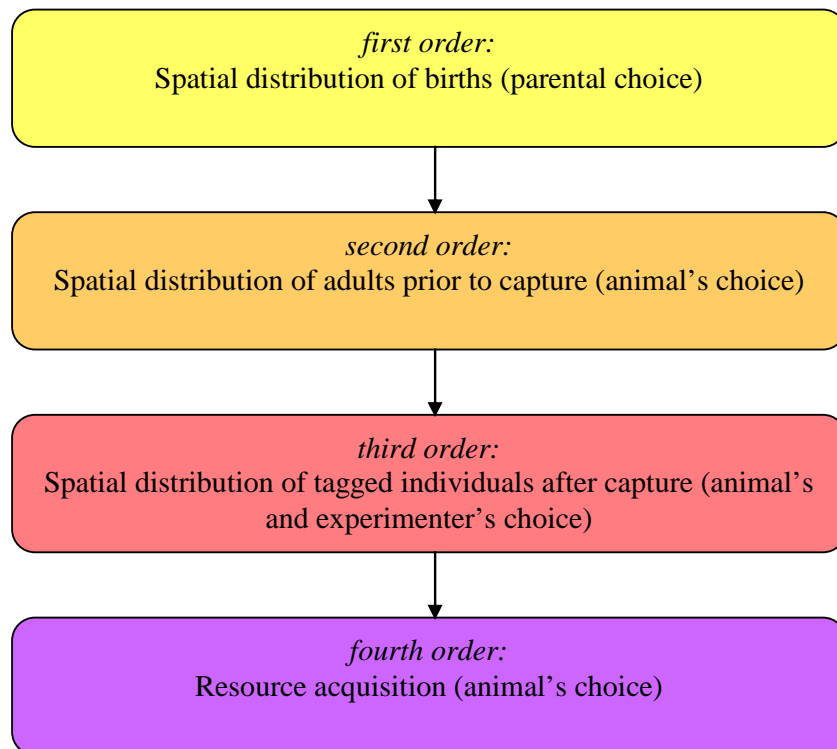
Fig. 6.1 Diagram of the hierarchical decision on space use and environmental preference.

When setting out to investigate space use and environmental preference it is important to choose the appropriate sampling design, which can be line or point transects (for 1st and 2nd order), wildlife telemetry (for 3rd order) possibly in combination with physiological sensors (for 4th order). The appropriateness of a sampling design depends on the study species. Some central-place foragers such as rabbits might carefully choose to make burrows that are completely surrounded by good foraging areas (except those living on the SMRU car park). After this decision has been made, usage within the rabbit's territory may appear uniform to a wildlife telemetry study. In that case, a second order type of study, modelling the spatial distribution of burrows, might be more appropriate. Similarly, species with no spatial memory or those relying on uniformly random, ephemeral resources, might search randomly through space, but the actual occasional procurement of food might not be uniformly distributed. In that case, a fourth order type of study, modelling food selection, might be more appropriate. Equally, the objective of the study is an important consideration. For local interactions with fisheries one might want to know the multispecies functional response of a generalist predators such as seals (fourth order selection). However, if spatial prediction of usage is of the essence and estimates of total population size on land are known (as was the case in this study), wildlife telemetry is probably most valuable.

6.4 Future Directions

6.4.1 Multilevel modelling

Based on the classification of section 6.3.3, the work presented in this thesis is a study in *third order selection*, an investigation of the space use and environmental preference of individuals making foraging trips from a given central-place. In this thesis, I only examined the effect of individual characteristics on usage and preference but the mixed-effect approach can be extended to account for all the variance components outlined in Fig. 2.2. I consider the hierarchy of these extensions below:

Sub-populations.—— The sub-population can also be treated as a random-effect. Similar to the individual level, variation across sub-populations can be explicitly modelled as a function of sub-population characteristics, such as the number of individuals using the haul-out on average, its geographical position (i.e. latitude and longitude) or a measure of the relative availability of particular environmental conditions in proximity to the haulout site. Including sub-population characteristics might be particularly advantageous when making predictions for haul-out sites for which no wildlife telemetry data is available.

Foraging trips.—— A second extension is to treat the variability between trips as a random effect. Animals may perform different functions in trips of different duration and this could be included as a trip characteristic to model its effect on spatial usage.

Locations.—— Finally, at the lowest level, variability between single observations could be treated as a random effect. Most model approaches assume that the observation process for both the response as well as the environmental covariates is error free but this is not generally the case. For example, ARGOS provides a Location Quality (LQ) index for each observation which can be recast into an estimate of precision. With some additional work, this can also be obtained for spatio-temporally smoothed and regularized data. Similarly, the estimates of the environmental covariates at a point in space are characterised by variable, often known, precision. This information can be included by explicitly modelling the random effect errors between locations as a function of LC or the local standard errors in the estimates environmental covariates. This is particularly beneficial, because it allows some residual noise (due to the above errors) to be accounted for by the model.

All of these extensions to the mixed-effect model are conceptually simple, but difficult to fit because the computation required for the approximation of the model's likelihood increases exponentially with the number of random effects included. This is particularly relevant when

treating individual telemetry observations as random effects because of their large number.

6.4.2 Spatio-temporal models

Spatial and temporal autocorrelation generally leads to an over-fitted model (i.e. loss in model parsimony) and underestimation of the standard errors (§2.2.6 and §2.2.7). These effects depend on the scale of spatial autocorrelation in both the response (i.e. the distribution of the species) and the explanatory variables. In a highly heterogeneous landscape, the values for the explanatory variables are likely to be different for neighbouring telemetry observations, and as a consequence, the detrimental effect of spatial autocorrelation will be small (see also Fig. 2.3). In practice though, environmental variables are almost always spatially autocorrelated.

A solution to the problem as suggested in §2.2.6 and §2.2.7, is to explicitly model the spatial autocorrelation in the response (using an auto-logistic models (Augustin et al. 1996)) or the residual errors (using a geostatistical models model (Diggle et al. 1998)). Doing this, results in an appropriate likelihood function, and therefore allows for the use of standard model selection criteria such as AIC, but most importantly, it also provides information about potential causes of spatial autocorrelation in the species distribution. These can be intrinsic (e.g. inter- or intra-specific competition or coexistence) or extrinsic, due to a missing environmental variable.

Autoregressive models use response values of neighbouring observations as a candidate covariate. However, there is a potential risk that if the response data are measured very precisely relative to environmental data, the auto-covariate may outcompete many other covariates. This would be an interesting future research topic.

Geostatistical models are essentially random effect models, similar to the multi-level mixed model that describe the variability in the residuals as random effects. They express random variability in terms of the value and distance of neighbouring residual errors. Predictably, they are computationally very difficult to fit due to the high-dimensional integration over all random effects.

6. 5 Modelling space use and habitat preference from wildlife telemetry data

Observing how individuals move through space and use the resources that are available to them, generally broadens our understanding of the biology of the species and is therefore of great academic interest. However, if the research exercise does not lead to population predictions, this information is of limited scientific value and of no practical use to conservationists and wildlife managers. In contrast, measuring the distribution of non-identifiable individuals from a population (e.g. using transect surveys) allows population-level predictions, but cannot quantify individual variability in behaviour. For example, such studies cannot tell whether an animal observed at extreme environmental conditions (e.g. a grey seal off the continental shelf) reflects a one-off event for that individual or a one-off individual for that population. Identifying the ability of certain members of the population to exist in marginal habitats is important in trying to gauge the ability of the species to adapt in changing conditions. Wildlife telemetry combined with the present statistical framework can produce population estimates and measures of individual variability aimed at addressing the needs of practitioners at a time of pressing issues in conservation and management.

6.6 References

Aebischer, N. J., P. A. Robertson, and R. E. Kenward. 1993. Compositional Analysis of Habitat Use from Animal Radio- Tracking Data. Ecology **74**:1313-1325.

Arthur, S. M., Manly, B.F.J., McDonald, L.L. & Garner, G.W. 1996. Assessing habitat selection when availability changes. Ecology **77**:215-227.

Augustin, N. H., Mugglestone, M.A. & Buckland, S.T. 1996. An autologistic model for the spatial distribution of wildlife. Journal of Applied Ecology **33**:339-347.

Beringer, J., J. J. Millspaugh, J. Sartwell, and R. Woeck. 2004. Real-time video recording of food selection by captive white-tailed deer. Wildlife Society Bulletin **32**:648-654.

Biuw, M., McConnell, B., Bradshaw, C.J.A., Burton, H., Fedak, M. 2003. Blubber and buoyancy: monitoring the body condition of free-ranging seals using simple dive characteristics. Journal of experimental biology **206** (19): 3405-3423

Bolker, B.M. 2004. Continuous space models for population dynamics. In: Ecology, genetics, and evolution of metapopulations (Hanski, I., and O. E. Gaggriotti, eds). Elsevier academic press, London.

Bowen, W. D., D. Tully, D. J. Boness, B. M. Bulheier, and G. J. Marshall. 2002. Prey-dependent foraging tactics and prey profitability in a marine mammal. Marine Ecology-Progress Series **244**:235-245.

Boyce, M. S., and L. L. McDonald. 1999. Relating populations to habitats using resource selection functions. TREE **14**:268-272.

Brown, R.G., and P.Y.C. Hwang. 1992. Introduction to random signals and applied Kalman Filtering. John Wiley & Sons, New York.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., and L. Thomas. 2001. Introduction to Distance Sampling. Oxford University Press, Oxford.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., and L. Thomas. 2004. Advanced Distance Sampling. Oxford University Press, Oxford.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference. A practical Information-Theoretic Approach, 2nd edition edition. Springer-Verlag, New York.

Butler. 2004. Biotelemetry: a mechanistic approach to ecology. Trends in Ecology & Evolution **19** (6): 334-343

Cooke, S. J., S. G. Hinch, M. Wikelski, R. D. Andrews, L. J. Kuchel, T. G. Wolcott, and P. J. Butler. 2004. Biotelemetry: a mechanistic approach to ecology. Trends in Ecology & Evolution **19**:334-343.

Diggle, P. J., J. A. Tawn, and R. A. Moyeed. 1998. Model-based geostatistics. Journal of the Royal Statistical Society Series C-Applied Statistics **47**:299-326.

Doucet, A., de Freitas, N., and N. Gordon, 2001. Sequential Monte Carlo Methods in Practice. Springer-Verlag, New York.

Drake, J. M., C. Randin, and A. Guisan. 2006. Modelling ecological niches with support vector machines. Journal of Applied Ecology **43**:424-432.

Durban, J., and S. J. Koopman. 2001. Time series analysis by state-space methods. Oxford University Press, New York.

Farnsworth, K. D., and J. A. Beecham. 1997. Beyond the ideal free distribution: More general models of predator distribution. Journal of Theoretical Biology **187**:389-396.

Gilks, W.R., Richardson, S., and D.J. Spielgelhalter.1996. Markov Chain Monte Carlo in practice. Chapman & Hall Boca Raton.

Hamer, K. C., R. A. Phillips, S. Wanless, M. P. Harris, and A. G. Wood. 2000. Foraging ranges, diets and feeding locations of gannets Morus bassanus in the North Sea: evidence from satellite telemetry. Marine Ecology-progress Series **200**:257-264.

Hammond, P. S., A. J. Hall, and J. H. Prime. 1994. The diet of grey seals in the Inner and Outer Hebrides. Journal of Applied Ecology **31**:737-746.

Harwood, J., and K. Stokes. 2003. Coping with uncertainty in ecological advice: lessons from fisheries. Trends in Ecology & Evolution **18**:617-622.

Hastie, T. J., and R. J. Tibshirani. 1990. Generalized Additive Models. Chapman & Hall, London.

Hilborn, R. 1975. Effect of Spatial Heterogeneity on Persistence of Predator-Prey Interactions. Theoretical Population Biology **8**:346-355.

Hjermann, D. O. 2000. Analyzing habitat selection in animals without well-defined home ranges. Ecology **81**:1462-1468.

Holland, G. J., S. P. R. Greenstreet, I. M. Gibb, H. M. Fraser, and M. R. Robertson. 2005. Identifying sandeel Ammodytes marinus sediment habitat preferences in the marine environment. Marine Ecology-Progress Series **303**:269-282.

Holmes, E. E., M. A. Lewis, J. E. Banks, and R. R. Veit. 1994. Partial-Differential Equations in Ecology - Spatial Interactions and Population-Dynamics. Ecology **75**:17-29.

Hunt, G.L Jr. 1990. The pelagic distribution of maine birds in a heterogeneous environment. Polar Research **8**: 43-54

Hunt, G.L. Jr, Harrison, N.M., and T. Cooney. 1990. The influence of hydrographic structure and prey abundance on foraging of least auklets. Studies in Avian Biolog **14**: 7-22

Jansen, V. A. A. 1995. Regulation of predator-prey systems through spatial interactions: A possible solution to the paradox of enrichment. Oikos **74**:384-390.

Johnsen, I. D., Myers, R. A., and J. M. Flemming. 2003. Meta-analysis of animal movement using state-space models. Ecology **84** (11): 3055-3063

Johnson, D. H. 1980. The comparison of usage and availability measurements for evaluating resource preference. Ecology 6:65-71.

Keating, K. A., and S. Cherry. 2004. Use and interpretation of logistic regression in habitat selection studies. Journal of Wildlife Management **68**:774-789.

Lewis, S., Sherratt, T.N., Hamer, K.C., and S. Wanless. 2001. Evidence of intra-specific competition for food in a pelagic seabird. Nature **412**: 816-819.

Manly, B. F. J., L. L. McDonald, and D. L. Thomas. 1993. Resource selection by animals: statistical design and analysis for field studies. Chapman and Hall, London

Matthiopoulos, J. 2003. The use of space by animals as a function of accessibility and preference. Ecological Modelling **159**:239-268.

Matthiopoulos, J., B. McConnell, C. Duck, and M. Fedak. 2004. Using satellite telemetry and aerial counts to estimate space-use by grey seals around the British Isles. Journal of Applied Ecology **41**:476-491.

McLaughlin, J. F., and J. Roughgarden. 1992. Predation across Spatial Scales in Heterogeneous Environments. Theoretical Population Biology **41**:277-299.

Miller, P., Johnson. M., and P. Tyack. 2004. Sperm Whale Behavior Indicates the Use of Ecolocation Click Buzzes 'Creaks' in Prey Capture. Proc. R. Soc. Lond. B **271**: 2239-2247.

Morales, J. M., and S. P. Ellner. 2002. Scaling up animal movements in heterogeneous landscapes: The importance of behavior. Ecology **83**:2240-2247.

Mysterud, A., and R. A. Ims. 1998. Functional responses in habitat use: Availability influences relative use in trade-off situations. Ecology **79**:1435-1441.

Pinheiro, J. C., and D. M. Bates. 2000. Mixed-effects models in S and S-plus. Springer-Verlag, New York.

Prentice, R. L., and R. Pyke. 1979. Logistic Disease Incidence models and case-control studies. Biometrika **66**:403-411.

Prime, J. H., and P. S. Hammond. 1990. The Diet of Gray Seals from the South-Western North-Sea Assessed from Analyses of Hard Parts Found in Feces. Journal of Applied Ecology **27**:435-447.

Reilly, S.B., and P.C. Fiedler. 1994. The interannual variability of dolphin habitats in the eastern tropical pacific 1. Reseach vessel surveys, 1986-1990. Fishery Bulletin **92**: 434.

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

Spear, L. B., L. T. Ballance, and D. G. Ainley. 2001. Response of seabirds to thermal boundaries in the tropical Pacific: the thermocline versus the Equatorial Front. Marine Ecology-Progress Series **219**:275-289.

Thompson, D., and M. A. Fedak. 2001. How long should a dive last? A simple model of foraging decisions by breath-hold divers in a patchy environment. Animal Behaviour **61**:287-296.

Thompson, D., Hiby, A.R., and M. A. Fedak. 1993. How fast should I swim? Behavioural implications of diving physiology. Symposia of the Zoological Society of London **66**:349-368.

Thompson, D., P. S. Hammond, K. S. Nicholas, and M. A. Fedak. 1991. Movements, diving and foraging behaviour of grey seals {IHalichoerus grypus}. Journal of Zoology **224**:223-232.

Tilman, D., and P. Kareiva. 1997. Spatial Ecology, the role of space in population dynamics and interspecific interactions, Princeton, New Jersey.

Turchin, P. 1998. Quantitative analysis of movement; measuring and modeling population redistribution in animals and plants. Sinauer, Sunderland, Massachusetts.

Turner, J.M., and Dagg M.J. 1983. Vertical distributions of continental shelf zooplankton in stratified and isothermal waters. Biological Oceanography **3**: 1-40

Wright, P. J., H. Jensen, and I. Tuck. 2000. The influence of sediment type on the distribution of the lesser sandeel, Ammodytes marinus. Journal of Sea Research **44**:243-256.

Xavier, J.C., Tarling, G.A., and J.P. Croxall. 2006. Determining prey distribution patterns from stomach-contents of satellite-tracked high-predators of the Southern Ocean. Ecography **29** (2): 260-272.

APPENDIX A: ADMB CODE

```
// Modelling environmental preference

DATA_SECTION
  init_int n                                          // Number of observations
  init_vector y(1,n)                                  // Response vector; 0 or 1
  init_int p                                          // Total number of fixed effects
  init_matrix X(1,n,1,p)                              // Covariate matrix for fixed effects
  init_int r                                          // Total number of random effects
  init_matrix Z(1,n,1,r)                              // Covariate matrix for random effects
  init_int M                                          // Total number of individuals
  init_vector nobs_i(1,M)                             // The total number of observations for each individual

PARAMETER_SECTION
  init_bounded_vector beta(1,p,-50,50,1)

                                                      // Fixed effects parameters

  init_bounded_vector sd_b(1,r,0.00000000000001,50,2)

                                                      // standard devations of random effects, covariances between random effects are here assumed to be
                                     0

  random_effects_matrix b(1,M,1,r,2)

                                                      // Unscaled individual random effects matrix;see nested4.tpl
  objective_function_value g                          // g will become the log-likelihood function to be minimised

PRELIMINARY_CALCS_SECTION
  cout << setprecision(4);

GLOBALS_SECTION
  #include <df1b2fun.h>

PROCEDURE_SECTION
  int i,ii;
  g = 0.0;
  ii = 0;

  for(i=1;i<=M;i++)
    fit_individual(beta,b(i),sd_b,i,ii);

SEPARABLE_FUNCTION void fit_individual(const dvar_vector& beta, const dvar_vector& b_i, const dvar_vector& sd_b,int i, int& ii)

    int j, q;

    dvariable eta;
```

```
    for(q=1;q<=r;q++)
    {
      g -= -log(sd_b(q)) - .5*square(b_i(q)/sd_b(q));        //LogL(b|sd_b)
    }


    for(j=1;j<=nobs_i(i);j++)                               //so j is the j'th observation of the i'th individual
    {
      ii++;                                                 //ii is the unique identifier for an observation (ii++ means ii+1)

      eta = X(ii)*beta + Z(ii)*b_i;                         //eta is the predictor on the scale of the link(ie logit)-function
      g -= y(ii)*eta - log(1+mfexp(eta));                   //LogL(y|b_i)
    }
REPORT_SECTION
  report << beta << endl;
  report << sd_b << endl;
  report << b << endl;

RUNTIME_SECTION
  maximum_function_evaluations 20000
  convergence_criteria 1.e-5

TOP_OF_MAIN_SECTION
  arrmblsize = 950000;
  gradient_structure::set_GRADSTACK_BUFFER_SIZE(300000);
  gradient_structure::set_CMPDIF_BUFFER_SIZE(310000);
  gradient_structure::set_MAX_NVAR_OFFSET(1000000);
```

Table 1. Parameter estimates of the final model fitted to grey seal data from the Farnes Isles. The table contains mean estimates, standard deviations and parameter correlations of the fixed-effects parameters β and the variances v of the random-effects b.

index	name	estimate	std dev	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
1	β-intercept	-1.56E+01	5.16E+00	1																																											
2	β-distance1	5.94E+00	2.97E+00	-0.5796	1																																										
3	β-distance2	4.33E+00	2.98E+00	-0.5802	0.9882	1																																								
4	β-distance3	6.51E+00	2.96E+00	-0.5831	0.9916	0.9924	1																																						
5	β-distance4	3.66E-01	4.10E+00	-0.4561	0.7456	0.7565	0.7483	1																																				
6	β-distance5	-1.64E+01	5.41E+00	-0.2233	0.389	0.3821	0.3952	0.1963	1																																		
7	β-distance6	-1.19E+01	4.80E+00	-0.095	0.1429	0.1427	0.1418	0.1188	-0.0061	1																																	
8	β-mud1	-6.11E+00	6.59E+00	-0.2169	-0.0134	-0.0129	-0.0109	-0.0246	0.0067	0.0068	1																																
9	β-mud2	2.83E+00	3.63E+00	-0.4326	-0.0092	-0.0061	-0.0079	-0.0029	0.0022	0.0074	0.2787	1																															
10	β-mud3	6.53E-01	3.30E+00	-0.4731	-0.0142	-0.0133	-0.0109	-0.0264	0.0056	0.013	0.4649	0.8425	1																													
11	β-mud4	1.41E+01	6.71E+00	-0.2718	0.0056	0.009	0.003	-0.0083	0.013	0.0039	0.2045	0.5114	0.4226	1																												
12	β-mud5	1.62E+01	2.58E+01	-0.061	-0.0165	-0.0187	-0.014	-0.0341	0.0051	-0.0054	0.1011	0.0909	0.198	-0.2261	1																										
13	β-mud6	-1.36E+01	2.09E+01	0.0875	-0.0119	-0.0101	-0.0121	-0.0132	0.0122	0.0278	-0.0572	-0.1118	-0.1186	0.0074	-0.1798	1																								
14	β-gravel	3.60E-02	6.94E-03	-0.0389	-0.0153	-0.0128	0.0309	0.0741	-0.037	0.0032	0.0022	0.0016	-0.0115	-0.0049	-0.0174	-0.006	1																						
15	β-depth1	3.45E+00	4.29E+00	-0.0417	-0.0146	-0.012	-0.0105	-0.0072	-0.0091	0.0042	0.0162	0.0295	0.0301	0.0131	0.01	0.0359	0.0221	1																					
16	β-depth2	-1.20E+00	4.72E+00	-0.2703	-0.0009	0.0004	0.0031	-0.0238	0.0117	0.0089	-0.0058	0.002	-0.0023	-0.0038	0.0204	-0.0021	-0.0134	-0.0188	1																			
17	β-depth3	6.08E+00	3.15E+00	-0.572	0.0153	0.012	0.0117	0.0255	-0.0065	0.0045	-0.0263	-0.031	-0.042	-0.0031	-0.0372	-0.0331	0.0479	0.0503	0.3906	1																		
18	β-depth4	3.40E+00	3.03E+00	-0.5776	0.0155	0.0115	0.0116	0.0235	-0.0028	0.0083	-0.0273	-0.0327	-0.0427	-0.0078	-0.0277	-0.0284	0.0023	0.0462	0.4741	0.9728	1																
19	β-depth5	2.56E+00	3.13E+00	-0.5653	0.0116	0.0135	0.014	0.0231	-0.0007	0.0046	-0.0243	-0.027	-0.0378	-0.0057	-0.032	-0.0327	0.0181	0.0494	0.4161	0.9742	0.9635	1															
20	β-depth6	3.26E+00	3.06E+00	-0.5746	0.0157	0.013	0.0136	0.0271	-0.0036	0.0094	-0.0259	-0.0297	-0.0395	-0.0059	-0.0295	-0.0289	0.0227	0.0496	0.4562	0.9735	0.9844	0.9578	1														
21	β-mud1*length	4.93E-02	3.82E-02	-0.0216	0.0165	0.0159	0.0155	0.0259	-0.0019	-0.0031	-0.8812	0.0654	-0.1008	0.0033	-0.0406	0.0147	0.0063	-0.0018	0.0106	0.0108	0.0134	0.0088	0.0114	1													
22	β-mud2*length	1.88E-02	1.47E-02	-0.0365	0.0092	0.0083	0.0071	-0.0052	0.0189	0.0075	0.0953	-0.5613	-0.1977	-0.1996	0.0578	0.0344	-0.0155	-0.0014	0.0093	0.0074	0.0151	0.0006	0.0081	-0.1035	1												
23	β-mud3*length	2.85E-02	1.04E-02	-0.0586	0.0339	0.0351	0.0294	0.0701	0.0043	0.0006	-0.1905	-0.2769	-0.4377	0.0043	-0.1159	0.0445	0.0033	-0.0025	0.0203	0.0161	0.0341	0.0127	0.0225	0.2296	0.5007	1											
24	β-mud4*length	-1.32E-01	3.94E-02	0.0208	-0.0107	-0.014	-0.0078	0.0111	-0.0125	0.0086	-0.0146	-0.1698	-0.0281	-0.8681	0.3148	-0.0671	0.0399	0.0074	-0.008	-0.0028	-0.0012	-0.0075	-0.003	-0.0077	0.2535	0.0035	1										
25	β-mud5*length	-9.85E-02	1.78E-01	0.0232	0.019	0.0215	0.0158	0.0336	-0.004	0.0041	-0.067	-0.0323	-0.1305	0.2415	-0.9669	0.1659	0.0012	-0.0162	-0.0225	0.034	0.0245	0.0305	0.0259	0.0376	-0.0668	0.1059	-0.3145	1									
26	β-mud6*length	1.22E-01	1.27E-01	-0.1203	0.0226	0.0211	0.0233	0.0244	-0.0081	-0.0324	0.0741	0.1409	0.1497	0.0068	0.1931	-0.9792	0.0125	-0.0442	-0.0009	0.044	0.037	0.0425	0.0379	-0.0173	-0.0342	-0.0429	0.0727	-0.1828	1								
27	b-distance1	1.00E+00	3.85E-01	-0.0183	-0.0187	-0.0001	-0.0024	0.0012	-0.014	0.0053	0.0213	0.0009	0.0135	0.006	0.0077	-0.008	0.0323	0.0146	0.0043	0.0024	0.0019	0.0054	-0.0031	-0.0063	0.0375	0.0285	0.0046	-0.0021	0.0065	1							
28	b-distance2	9.42E-01	4.21E-01	-0.0345	0.034	0.0452	0.0264	0.0635	0.013	-0.012	-0.0585	-0.0164	-0.0851	0.0315	-0.028	0.0276	0.0104	0.0016	0.0167	0.0397	0.0315	0.0244	0.0353	0.0635	0.0179	0.1504	-0.0322	0.0193	-0.0243	0.0329	1						
29	b-distance3	3.30E-05	4.50E-01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1					
30	b-distance4	1.11E+01	2.38E+00	0.0357	-0.0229	-0.0328	-0.0175	-0.2483	0.0042	0.0095	0.0248	-0.0062	0.0398	0.0216	0.0303	-0.0038	-0.0205	0.0071	0.0295	-0.025	-0.0305	-0.0237	-0.0304	-0.0221	0.0042	-0.0906	-0.0282	-0.0299	0.0014	0.0053	-0.0776	0	1				
31	b-distance5	1.71E+01	3.59E+00	-0.0893	0.1412	0.1447	0.1411	0.1484	-0.2044	-0.0269	0.0027	-0.0123	-0.0021	-0.0043	0.0064	-0.0003	0.0283	0.0045	-0.0117	0.0065	0.0031	0.0025	0.004	0.0049	0.0148	0.0142	0.0055	-0.0065	-0.0006	0.0103	0.0225	0	0.0218	1			
32	b-distance6	1.27E+01	4.28E+00	-0.0532	0.0918	0.0907	0.0938	0.0665	0.0538	-0.1985	0.0024	0.0142	0.0069	-0.0006	0.0174	0.0022	0.0055	-0.0006	-0.0002	-0.0092	-0.0102	-0.0091	-0.0109	0.0023	-0.0134	0.0048	-0.001	-0.0146	-0.0008	0.0007	0.009	0	-0.0123	0.0772	1		
33	b-mud1	3.98E+00	1.23E+00	-0.0599	0.0174	0.0182	0.0156	0.0191	-0.0064	0.0093	-0.0669	0.0488	0.0452	0.0083	0.0165	-0.017	-0.0152	0.0003	0.0004	0.0112	0.0149	0.0111	0.0129	0.0094	0.0587	0.0439	0.0349	-0.0159	0.0219	0.0001	0.0112	0	-0.0224	-0.0157	0.015	1	
34	b-mud2	1.04E-04	1.70E+00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
35	b-mud3	6.10E-01	2.92E-01	-0.04	0.0045	-0.0037	0.0097	-0.035	0.0004	0.005	0.0325	-0.0338	0.1031	-0.0146	0.076	0.0119	-0.0169	-0.0043	0.0138	-0.0091	-0.0074	-0.0082	-0.0052	-0.0063	0.0997	-0.0626	0.0265	-0.0847	-0.0124	0.0032	-0.0666	0	0.1437	0.022	-0.0047	-0.0463	0
36	b-mud4	3.96E+00	1.40E+00	-0.0411	0.0239	0.0243	0.0243	-0.0055	0.0162	-0.023	-0.0045	0.0252	0.012	0.0753	0.0364	-0.0054	-0.0305	-0.0033	0.0169	0.0209	0.0242	0.0283	0.0219	0.0275	0.0341	-0.1327	-0.047	0.0042	0.0198	-0.0208	0	0.0498	-0.0064	-0.0101	-0.0514	-0.0001	0.0768
37	b-mud5	1.49E+01	5.77E+00	-0.092	-0.0232	-0.0263	-0.0225	-0.0313	-0.0025	0.0001	0.0594	0.092	0.146	-0.0125	0.3448	0.0087	-0.0082	0.0116	-0.0015	0.0056	0.011	0.0055	0.0089	0.0123	0.0855	0.0263	0.1067	-0.4306	-0.012	-0.0056	0.0062	0	0.0227	0.0169	0.0244	0.0242	0
38	b-mud6	1.63E+00	4.35E+00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
39	b-gravel	2.11E-02	5.67E-03	-0.0152	0.0028	0.0005	0.0093	-0.0054	-0.0197	-0.003	-0.0122	-0.0481	-0.0436	0.0037	0.0077	-0.0119	0.0894	0.0041	0.0245	0.0137	0.0078	0.0093	0.0096	0.0198	0.0708	0.099	-0.0104	-0.0145	0.0137	0.1509	0.0798	0	0.0879	0.0334	0.0144	-0.0318	0
40	b-depth1	1.94E-05	4.57E+00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
41	b-depth2	1.23E+01	4.37E+00	-0.1364	0.026	0.0256	0.0248	0.039	0.0397	-0.0137	-0.0237	-0.0612	-0.0206	0.0184	0.0388	0.0884	-0.0056	-0.0205	-0.1004	0.1273	0.1375	0.1202	0.1333	0.0484	0.2028	0.1892	0.0041	-0.0503	-0.1034	-0.0042	-0.0109	0	-0.0789	-0.0204	0.0084	0.0964	0
42	b-depth3	9.85E-01	7.45E-01	0.0841	-0.0436	-0.0432	-0.0437	-0.0876	-0.012	0.0008	0.1324	0.2582	0.266	0.0994	0.0061	-0.0668	0.0044	0.0053	-0.0008	-0.0368	-0.0615	-0.0315	-0.0496	-0.16	-0.4715	-0.6591	-0.1226	0.0015	0.0693	-0.0004	0.0001	0	0.1302	-0.0141	-0.0194	-0.0563	0
43	b-depth4	1.20E+00	3.89E-01	0.0481	-0.0265	-0.0265	-0.0237	-0.0733	-0.031	0.0165	0.0653	0.1222	0.0969	0.054	0.011	-0.0357	0.0491	0.0196	0.0088	-0.0122	-0.025	-0.0099	-0.0192	-0.0752	-0.2447	-0.289	-0.0803	-0.0094	0.0362	0.0497	0.0555	0	0.1956	0.0034	-0.0087	-0.0996	-0.0001
44	b-depth5	2.34E+00	7.47E-01	-0.037	0.0229	0.0106	0.0112	0.0425	-0.0077	-0.0102	-0.0865	-0.1687	-0.1749	-0.1323	0.0024	0.0184	0.0028	0.0201	0.0339	0.01	0.0339	-0.0247	0.0276	0.1024	0.306	0.4192	0.1559	-0.0062	-0.0192	0.0239	0.0069	0	-0.0675	0.0128	0.0173	0.0342	0
45	b-depth6	1.77E+00	4.93E-01	-0.0141	0.0003	-0.0082	-0.0046	0.0091	-0.0216	-0.0074	-0.086	-0.1413	-0.1724	-0.0966	-0.0183	-0.0151	-0.0044	0	0.0034	0.0346	0.0345	0.043	-0.0032	0.0911	0.2288	0.3512	0.1015	0.0181	0.0163	0.0585	-0.0966	0	-0.0254	0.0191	0.0259	0.0353	0

Continued columns (36–45) for rows 35–45:

index	name	36	37	38	39	40	41	42	43	44	45
35	b-mud3	1									
36	b-mud4	1									
37	b-mud5	0.1266	-0.0549	1							
38	b-mud6	0	0	1							
39	b-gravel	0.2216	0.0202	0.047	0	1					
40	b-depth1	0	0	0	0	1					
41	b-depth2	0.0707	-0.0348	0.1433	0	-0.0122	0	1			
42	b-depth3	-0.0361	-0.0166	-0.0708	0	-0.029	0	-0.3837	1		
43	b-depth4	-0.0371	0.0687	-0.0303	0	0.2318	0	-0.2278	0.3683	1	
44	b-depth5	-0.0087	0.0196	0.0396	0	0.0539	0	0.1744	-0.4941	-0.1861	1
45	b-depth6	-0.133	0.0274	0.0026	0	0.0462	0	0.0465	-0.2653	-0.1244	0.1516	1

Table 2. Parameter estimates of the final model fitted to gannet data from the Bass Rock. The table contains mean estimates, standard deviations and parameter correlations of the fixed-effects parameters β and the variances ν of the random-effects b.

index	name	value	std dev	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
1	β-intercept	-1.10E+01	5.68E+00	1																																				
2	β-coastdistance1	6.51E+00	3.24E+00	-0.5682	1																																			
3	β-coastdistance2	5.97E+00	3.21E+00	-0.5668	0.9827	1																																	
4	β-coastdistance3	7.12E+00	3.23E+00	-0.5641	0.9863	0.9918	1																																
5	β-coastdistance4	6.85E+00	3.29E+00	-0.5508	0.9525	0.9697	0.9578	1																															
6	β-coastdistance5	6.31E+00	3.59E+00	-0.5162	0.9027	0.902	0.9173	0.8478	1																														
7	β-coastdistance6	5.71E+00	3.62E+00	-0.4652	0.8009	0.8187	0.8066	0.8205	0.6707	1																													
8	β-depth1	4.20E+00	3.68E+00	-0.4482	-0.0079	-0.0106	-0.0096	-0.0171	-0.0033	-0.0079	1																												
9	β-depth2	2.70E-01	4.20E+00	-0.4435	-0.0016	-0.0071	-0.0123	-0.0065	-0.0001	0.0028	0.5491	1																											
10	β-depth3	6.76E+00	4.41E+00	-0.5649	0.0136	0.0079	0.0086	0.0106	0.0074	-0.0023	0.7853	0.7211	1																										
11	β-depth4	3.24E+00	3.34E+00	-0.5679	0.0107	0.0018	0.0009	0	0.0104	-0.0034	0.7786	0.7957	0.9627	1																									
12	β-depth5	4.55E+00	3.39E+00	-0.5627	0.0053	0.0087	0.0076	0.0095	0.0098	0.0062	0.7795	0.7481	0.9714	0.968	1																								
13	β-depth6	2.38E+00	3.36E+00	-0.5521	-0.0004	0.0083	0.0031	0.007	0.0089	0.0052	0.7639	0.7679	0.9408	0.9708	0.9359	1																							
14	β-distance1	-5.44E-01	3.40E+00	-0.5731	-0.0004	-0.0025	-0.0035	0.0002	-0.0067	0.0047	-0.0153	-0.0218	-0.0237	-0.0346	-0.0319	-0.033	1																						
15	β-distance2	7.38E-01	3.38E+00	-0.5798	0.0095	0.0033	0.0017	0.0074	-0.0037	0.0131	-0.0082	-0.0144	-0.0172	-0.0265	-0.0238	-0.0298	0.991	1																					
16	β-distance3	-1.04E+00	3.44E+00	-0.5744	0.0021	-0.0013	-0.0078	0.0005	-0.0106	0.0049	-0.0142	-0.0134	-0.0233	-0.0298	-0.0275	-0.0321	0.995	0.9884	1																				
17	β-distance4	-2.35E+00	3.35E+00	-0.5668	0.0121	0.0064	0.0052	-0.0034	0.0075	0.0151	-0.0061	0.0011	-0.0361	-0.0197	-0.0272	-0.0271	0.9643	0.9713	0.9589	1																			
18	β-distance5	5.29E-02	3.96E+00	-0.5316	0.0004	-0.0033	-0.0123	0.0064	-0.0402	-0.0064	-0.015	-0.0421	-0.0074	-0.0339	-0.0284	-0.0332	0.9281	0.9152	0.938	0.8509	1																		
19	β-distance6	-5.58E+00	3.69E+00	-0.4132	0.0505	0.0386	0.0469	0.0241	0.069	0.0166	-0.0313	0.0186	-0.0211	-0.0012	-0.0155	-0.0161	0.656	0.6685	0.64	0.717	0.4523	1																	
20	b-coastdistance1	7.94E-01	4.05E-01	-0.0505	0.0427	0.0707	0.0682	0.0646	0.0553	0.0696	0.0782	0.043	0.059	0.0667	0.0577	0.0459	-0.0531	-0.0343	-0.0531	-0.0182	-0.0641	-0.0052	1																
21	b-coastdistance2	2.18E-06	2.75E-01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1															
22	b-coastdistance3	2.98E-01	4.19E-01	-0.0641	0.0469	0.0519	0.0375	0.0564	0.0325	0.0462	0.0397	0.0474	0.0644	0.0731	0.0707	0.0778	-0.0162	-0.005	-0.0155	0.0139	-0.0222	-0.0499	0.0202	0	1														
23	b-coastdistance4	2.04E+00	6.57E-01	0.0179	-0.0114	-0.02	-0.0131	-0.0027	0.0011	-0.0146	-0.016	-0.0129	0.0023	-0.0061	-0.0024	-0.0044	-0.0103	-0.0086	-0.0166	-0.0291	-0.0157	-0.0323	-0.0789	0	0.1012	1													
24	b-coastdistance5	3.60E+00	1.15E+00	0.0007	-0.0292	-0.0185	-0.0316	0.0061	-0.0822	0.0912	-0.0044	0.0152	0.014	0.0104	0.018	0.0146	0.0096	0.0084	0.0136	-0.0065	0.0054	0.0075	-0.0077	0	-0.0267	-0.1003	1												
25	b-coastdistance6	4.47E+00	2.34E+00	-0.089	0.1211	0.1139	0.127	0.0863	0.2079	-0.1722	0.0112	0.0314	0.0134	0.024	0.0209	0.0168	0.0135	0.0188	0.0063	0.0408	-0.0458	0.0082	-0.0003	0	0.0885	0.0834	-0.2925	1											
26	b-depth1	1.26E-04	2.12E+00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1											
27	b-depth2	6.97E+00	1.98E+00	-0.0339	0.0135	0.0127	0.0174	0.0056	0.0238	0.017	0.0282	-0.0674	0.0705	0.0426	0.0624	0.0456	-0.0083	-0.004	-0.0136	0.0037	-0.0387	-0.0188	0.026	0	0.0987	0.015	0.004	0.0474	0	1									
28	b-depth3	1.34E+00	6.50E-01	-0.0559	0.0904	0.0802	0.0828	0.0806	0.0792	0.0756	0.062	0.0564	0.0596	0.0964	0.0713	0.0859	-0.0863	-0.0573	-0.0923	-0.0104	-0.126	-0.0077	0.1468	0	0.4456	0.1413	-0.0681	0.129	0	0.1066	1								
29	b-depth4	3.51E-07	1.85E-01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1								
30	b-depth5	7.97E-01	8.59E-01	-0.0268	0.0578	0.0393	0.0469	0.0347	0.0314	0.0469	0.082	0.0329	0.0361	0.0617	0.0056	0.0307	-0.0599	-0.0359	-0.0619	-0.009	-0.0722	0.0093	0.3261	0	0.0482	-0.0447	-0.0239	-0.0078	0	0.0156	0.3166	0	1						
31	b-depth6	2.14E-05	8.84E-01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1						
32	b-distance1	1.22E-04	2.71E-01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0002	0	0	0	0	0	0	0	0	0	0	1					
33	b-distance2	6.83E-01	4.63E-01	-0.0826	0.0975	0.0919	0.0901	0.0853	0.0781	0.0863	0.1047	0.0708	0.0889	0.1212	0.0887	0.0978	-0.0787	-0.0448	-0.0779	-0.0013	-0.1024	-0.0025	0.3928	0	0.361	-0.0141	-0.0508	0.0737	0	0.1089	0.6045	0	0.5822	0	0.0002	1			
34	b-distance3	4.30E-04	3.36E-01	-0.0001	0	0.0001	0.0001	0.0001	0.0001	0	0.0001	0	0.0001	0	0.0001	0	0.0002	-0.0001	0.0001	0	0	-0.0002	0	0	-0.0001	0	0.0001	0	0	-0.0002	0	-0.0001	0	0	0.0001	1			
35	b-distance4	1.31E+00	8.78E-01	0.0458	-0.0804	-0.0743	-0.0748	-0.05	-0.0685	-0.0534	-0.065	-0.0611	-0.0302	-0.0848	-0.0485	-0.0713	0.0791	0.0535	0.0786	-0.0234	0.1152	-0.039	-0.2209	0	-0.3006	0.2203	0.0786	-0.0982	0	-0.0757	-0.5461	0	-0.3503	0	-0.0001	-0.5776	-0.0001	1		
36	b-distance5	2.36E+00	9.86E-01	-0.0217	0.0024	0.0012	0.0064	-0.0075	0.0584	-0.0203	0.0037	0.062	0.0267	0.0328	0.0375	0.0319	0.0009	0.0064	-0.0079	0.032	-0.1256	0.0789	-0.0232	0	0.0619	-0.0042	0.0708	0.294	0	0.0754	0.0186	0	-0.1277	0	0	-0.0304	0.0002	-0.0103	1	
37	b-distance6	2.09E+00	2.87E+00	-0.1254	-0.0485	-0.0458	-0.0582	-0.041	-0.0846	-0.0027	0.032	-0.0245	-0.0087	-0.0231	-0.0205	-0.0151	0.2794	0.2681	0.2949	0.2195	0.398	-0.3379	-0.0278	0	0.0879	-0.0467	-0.0089	-0.0018	0	0.0224	-0.0393	0	-0.0686	0	0	-0.0267	-0.0001	0.0166	-0.1033	1

Table 3. Parameter estimates of the model containing the covariates distance to the haul-out site, depth, temperature stratification, gravel and rock. This model is fitted to grey seals data from the west coast of Scotland. The table contains mean estimates, standard deviations and parameter correlations of the fixed-effects parameters β and the variances v of the random-effects b.

index	name	value	std dev	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
1	β-intercept	-30.00	0.01	1.00																																																		
2	β-distance1	26.20	1.29	0.00	1.00																																																		
3	β-distance2	30.00	0.00	0.00	0.00	1.00																																																	
4	β-distance3	27.50	1.84	0.00	0.50	0.00	1.00																																																
5	β-distance4	3.38	16.16	0.00	-0.02	0.00	-0.08	1.00																																															
6	β-distance5	-30.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00																																														
7	β-distance6	-30.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00																																												
8	β-depth1	-30.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00																																												
9	β-depth2	10.76	20.73	0.00	-0.02	0.00	-0.01	-0.01	0.00	0.00	0.00	1.00																																											
10	β-depth3	10.15	17.42	0.00	-0.02	0.00	-0.01	-0.01	0.00	0.00	0.00	0.96	1.00																																									
11	β-depth4	14.02	17.88	0.00	-0.02	0.00	-0.01	-0.01	0.00	0.00	0.00	0.98	0.99	1.00																																							
12	β-depth5	13.43	17.55	0.00	-0.03	0.00	-0.01	-0.01	0.00	0.00	0.00	0.98	0.99	1.00	1.00																																						
13	β-depth6	16.99	18.08	0.00	-0.03	0.00	-0.01	-0.01	0.00	0.00	0.00	0.97	0.98	0.99	0.99	1.00																																				
14	β-T-stratification1	-6.44	17.71	0.00	-0.02	0.00	-0.02	0.01	0.00	0.00	0.00	-0.97	-0.99	-0.99	-0.99	-0.99	1.00																																			
15	β-T-stratification2	-10.37	17.73	0.00	-0.02	0.00	-0.02	0.01	0.00	0.00	0.00	-0.97	-0.99	-0.99	-1.00	-0.99	0.99	1.00																																	
16	β-T-stratification3	-11.61	18.00	0.00	-0.02	0.00	-0.02	0.00	0.00	0.00	0.00	-0.96	-0.98	-0.98	-0.98	-0.98	0.98	0.98	1.00																															
17	β-T-stratification4	-12.90	17.70	0.00	-0.02	0.00	-0.02	0.01	0.00	0.00	0.00	-0.97	-0.99	-1.00	-1.00	-0.99	1.00	1.00	0.98	1.00																														
18	β-T-stratification5	-12.87	17.72	0.00	-0.02	0.00	-0.02	0.01	0.00	0.00	0.00	-0.98	-0.99	-1.00	-1.00	-0.99	1.00	1.00	0.99	1.00	1.00																												
19	β-T-stratification6	-11.96	17.74	0.00	-0.02	0.00	-0.02	0.01	0.00	0.00	0.00	-0.97	-0.99	-0.99	-1.00	-0.99	0.99	1.00	0.98	1.00	1.00	1.00																											
20	β-gravel1	0.78	36.00	0.00	-0.01	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00																										
21	β-gravel2	1.47	35.98	0.00	-0.01	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00																									
22	β-gravel3	2.52	35.99	0.00	-0.01	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00																								
23	β-gravel4	1.30	35.98	0.00	-0.01	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00																							
24	β-gravel5	2.39	36.02	0.00	-0.01	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00																						
25	β-gravel6	1.24	36.01	0.00	-0.01	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00																					
26	β-rock	1.80	35.99	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	1.00																				
27	b-distance1	1.78	0.98	0.00	-0.04	0.00	-0.01	0.01	0.00	0.00	0.00	-0.02	0.02	0.01	0.01	0.00	-0.01	-0.01	0.00	-0.01	-0.01	-0.01	0.01	0.01	0.01	0.01	0.01	0.01	-0.01	1.00																			
28	b-distance2	3.14	1.10	0.00	-0.15	0.00	-0.11	0.03	0.00	0.00	0.00	0.04	0.00	0.01	0.01	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.01	-0.49	1.00																		
29	b-distance3	4.99	1.27	0.00	-0.05	0.00	-0.12	-0.01	0.00	0.00	0.00	-0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.31	-0.18	1.00																	
30	b-distance4	47.15	14.40	0.00	0.08	0.00	0.13	-0.41	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.01	0.01	0.01	0.01	0.01	0.01	-0.01	0.00	-0.05	0.06	1.00																
31	b-distance5	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00															
32	b-distance6	0.02	68.87	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00													
33	b-depth1	19.98	14.10	0.00	-0.02	0.00	-0.01	-0.01	0.00	0.00	0.00	0.74	0.75	0.75	0.75	0.75	-0.75	-0.75	-0.74	-0.75	-0.75	-0.75	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.01	0.03	-0.01	0.02	0.01	0.00	0.00	1.00													
34	b-depth2	10.44	3.93	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	-0.30	-0.23	-0.24	-0.23	-0.24	0.23	0.23	0.23	0.23	0.23	0.23	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	-0.16	0.09	0.01	0.00	0.00	-0.21	1.00												
35	b-depth3	6.71	2.27	0.00	0.02	0.00	-0.02	0.01	0.00	0.00	0.00	-0.07	-0.11	-0.09	-0.10	-0.08	0.09	0.09	0.09	0.09	0.09	0.10	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	0.02	-0.31	0.27	-0.17	-0.02	0.00	0.00	-0.11	-0.10	1.00											
36	b-depth4	3.40	1.08	0.00	-0.03	0.00	-0.02	0.00	0.00	0.00	0.00	-0.04	-0.02	-0.02	-0.02	-0.03	0.02	0.02	0.03	0.02	0.02	0.02	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.01	0.48	-0.30	0.32	0.02	0.00	0.00	0.00	0.13	-0.19	1.00										
37	b-depth5	0.85	2.31	0.00	0.00	0.00	-0.01	0.01	0.00	0.00	0.00	-0.04	-0.09	-0.08	-0.08	-0.07	0.08	0.08	0.07	0.08	0.08	0.08	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.01	-0.60	0.53	-0.36	-0.03	0.00	0.00	-0.11	-0.17	0.54	-0.48	1.00									
38	b-depth6	3.96	3.23	0.00	-0.06	0.00	-0.03	0.01	0.00	0.00	0.00	0.16	0.13	0.14	0.13	0.15	-0.13	-0.14	-0.14	-0.14	-0.14	-0.13	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.00	-0.07	0.19	-0.12	-0.01	0.00	0.00	0.11	-0.16	0.08	-0.09	0.12	1.00								
39	b-T-stratification1	0.00	4.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00							
40	b-T-stratification2	2.76	1.10	0.00	0.00	0.00	-0.02	0.03	0.00	0.00	0.00	-0.02	-0.01	-0.02	-0.02	-0.01	0.01	0.02	0.00	0.02	0.02	0.02	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.01	0.06	-0.02	-0.04	0.00	0.00	0.00	-0.02	0.01	0.08	-0.06	0.06	0.03	0.00	1.00						
41	b-T-stratification3	8.46	3.14	0.00	-0.02	0.00	-0.03	0.03	0.00	0.00	0.00	0.01	0.02	0.01	0.01	0.02	-0.02	-0.01	-0.06	0.00	-0.02	-0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	-0.02	-0.05	0.02	0.03	0.00	0.09	0.00	1.00						
42	b-T-stratification4	1.08	1.01	0.00	0.03	0.00	0.04	-0.04	0.00	0.00	0.00	-0.04	-0.03	-0.03	-0.03	-0.04	0.03	0.03	0.06	0.01	0.03	0.02	0.02	0.02	0.02	0.02	0.03	0.02	-0.02	0.31	-0.21	0.06	0.01	0.00	0.00	-0.02	0.06	-0.07	0.18	-0.17	-0.11	0.00	-0.27	-0.32	1.00				
43	b-T-stratification5	0.00	0.68	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00			
44	b-T-stratification6	3.17	1.69	0.00	0.00	0.00	0.04	-0.04	0.00	0.00	0.00	-0.02	-0.02	-0.01	-0.02	-0.02	0.02	0.01	0.03	0.01	0.02	0.00	0.03	0.03	0.03	0.03	0.03	0.03	-0.03	0.12	-0.01	0.03	0.02	0.00	0.00	-0.01	0.00	-0.02	0.08	-0.06	0.00	0.00	-0.15	-0.25	0.39	0.00	1.00			
45	b-gravel1	2.51	0.66	0.00	-0.02	0.00	-0.02	0.00	0.00	0.00	0.00	0.02	0.02	0.02	0.02	0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	0.01	0.01	0.00	0.01	0.00	0.00	-0.01	-0.02	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.03	0.01	0.00	0.03	0.02	0.00	0.00	0.00	0.00	1.00		
46	b-gravel2	0.00	0.36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	
47	b-gravel3	0.00	0.65	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
48	b-gravel4	1.25	0.55	0.00	-0.04	0.00	-0.02	0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.01	-0.01	-0.01	-0.01	0.00	-0.01	0.00	-0.03	-0.03	-0.03	-0.03	-0.03	-0.02	0.03	-0.05	0.03	-0.01	-0.01	0.00	0.00	0.00	-0.02	0.04	-0.03	0.04	0.02	0.00	0.01	0.02	-0.07	0.00	-0.09	-0.01	0.00	0.00	1.00
49	b-gravel5	0.00	0.59	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
50	b-gravel6	3.88	1.54	0.00	-0.01	0.00	-0.01	-0.07	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.02	0.01	-0.04	0.04	0.02	0.05	0.00	0.00	0.01	-0.01	0.04	0.00	0.04	0.01	0.00	0.01	0.01	-0.04	0.00	-0.04	0.01	0.00	0.00	-0.01	0.00
51	b-rock	1.37	0.48	0.00	0.01	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	-0.01	-0.01	-0.01	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	-0.01	-0.13	0.16	0.00	0.01	0.00	0.00	-0.01	0.02	0.06	0.02	0.09	-0.07	0.00	0.00	-0.04	0.05	0.00	0.07	0.06	0.00	0.00	-0.16

Note: rows 49, 50 and 51 continue with diagonal value 1.00 in columns 49, 50 and 51 respectively.