Research

Modelling habitat selection at multiple scales with multivariate geostatistics: an application to seabirds in open sea

Edwige Bellier, Grégoire Certain, Benjamin Planque, Pascal Monestiez and Vincent Bretagnolle

E. Bellier (edwige.bellier@nina.no) and P. Monestiez, Unité Biostatistique and Processus Spatiaux, INRA Avignon, Domaine Saint-Paul, Site Agroparc, FR–84914 Avignon Cedex 9, France. Present address for EB and GC: Norwegian Institute for Nature Research-NINA, NO–7485 Trondheim, Norway. – G. Certain, Centre de Recherche sur les Mammifères Marins, Institut du Littoral et de l'Environnement, Univ. de La Rochelle, Avenue du Lazaret, FR–17000 La Rochelle, France. GC and V. Bretagnolle, Centre d'Etudes Biologiques de Chizé, Centre National de la Recherche Scientifique, FR–15 79360 Villiers en Bois, France. – B. Planque, Institute of Marine Research, PO 6404, NO–9294 Tromsø, Norway.

Modelling habitat of species necessitates robust identification of relevant environmental variables linked to species distribution. To achieve this, we connect hierarchical patch theory and habitat modelling at multiple scales. We suggest discriminating between 'circumstancial variables' and 'process variables' on the basis of temporal evolution of the spatial links between species distribution and their environment at different scales. 'Process variables' are informative of the ecological processes driving the distribution of organisms at multiple scales. By opposition, 'circumstantial variable' provide little insight because their relationship with animal spatial distribution is subject to great variations through time. As a real case study, we investigate the relationships between auk distribution (mainly *Uria aalge*) and oceanographic landscapes over two scales (i.e. large ~ 200 km and medium ~ 50 km) during the wintering season in the Bay of Biscay. Surface salinity, mixed layer depth and chlorophyll a are identified as 'process variables' as they are invariably correlated with the spatial distribution of auks, whereas bottom temperature can be viewed as a 'circumstantial variable' since the correlation is non-constant through time at large scale. The process variables at large scale are used to model the potential habitat of auks in the Bay of Biscay during the wintering season. At medium scale, only the chlorophyll a is identified as 'process variable' and used to model preferential habitat of wintering auks in the Bay of Biscay. The analytical approach proposed here (i.e. multivariate factorial kriging in a temporal context) is an effective framework to model the potential and preferential habitat of a species and can be related to the ecological niche concept and by focusing explicitly on scale dependence, the distinction between the variables that can be used as niche descriptors into species distribution models. Then our method lead to the identification of variables that should be used to define the Grinnellian niche which is defined by environmental conditions on broad scales and the Eltonian niche which focuses on biotic interactions and resource–consumer dynamics that can be measured principally at local scales.

The spatial distribution of animals results from the combination of extrinsic and intrinsic processes (Pearson and Dawson 2003, Wagner and Fortin 2005). Extrinsic processes refer to the influence of environmental factors (i.e. either biotic and abiotic), while intrinsic processes refer to population dynamics and intra-specific interactions (e.g. density-dependence, competition: Pacala and Roughgarden 1982, Pickett and Cadenasso 1995, Gaston and Blackburn 2000, Reineking et al. 2006, Schröder and Seppelt 2006). Moreover, these processes act at various spatial and temporal scales, and therefore animal populations often present a spatial distribution with scale-dependent structures (Kotliar and Wiens 1990, Levin 1992, Dungan et al. 2002). In this context, the clear identification of scale dependent structures as well as environmental variables that are correlated to these structures represents a critical step in understanding how and at which scale, extrinsic and intrinsic processes influence the spatial distribution of animals (Lindenmayer et al. 2007). Such identification has been tentatively formalised under the hierarchical patch dynamic theory (Allen and Starr 1982, Kotliar and Wiens 1990, Allen and Hoekstra 1991, Wu and David 2002) which combines the concepts of ecological processes, patterns and scales.

Habitat is at the core of the hierarchical patch dynamic theory (Allen and Starr 1982, Wu 1999) and here defined as a physical place at a particular spatio-temporal scale where and when an organism either actually or potentially lives (Kearney 2006). At the largest scale of the hierarchical patch dynamic system, habitat can be related to environmental conditions and viewed as potential habitat, i.e. the geographical realisation of the ecological niche (Hutchinson 1957, Chase and Leibold 2003) which is defined as the joint

description of the environmental conditions that a given species may find suitable for living. Nested within this potential habitat, smaller patches with higher density of organisms will correspond to preferential habitats. Modelling these different types of habitats requires the identification of a set of environmental variables which are related to the spatial distribution of animals at each scale (i.e. broad scale for potential habitat and finer scales for preferential habitats). A way to detect such relationships is to focus on the temporal variation of the strength and sign of the links (defined below as 'correlation') between the spatial distribution of animals and the environmental variables. We propose that environmental variables may be separated in two categories according to the way they are related to animal distribution: we will refer as a 'process variable' an environmental variable that displays a persistent link (i.e. in terms of strength and sign) through time with animal distribution, and as a 'circumstantial variable' an environmental variable that displays a variable link through time with animal distribution. Our underlying hypothesis is that a 'process variable' results directly from an ecological process, while a 'circumstantial variable' results from a particular event such as the spatial structuring of both the animal population and the environmental variable.

In this study, we bring together hierarchical patch theory and habitat modelling approach, to investigate habitat selection as a multiple scales process (Johnson 1980, Schaefer and Messier 1995, Mayor et al. 2007). The approach is based on a spatially explicit methodology derived from multivariate Geostatistics: the multivariate factorial kriging (MFK) (Goovaerts 1997, Wackernagel 2003). MFK has been used in soil science and environmental science to model multivariate correlation (Castrignano et al. 2000, Pebesma et al. 2005). Bellier et al. (2007) proposed to use MFK to identify and quantify scale-dependent relationships between species and environment. Here we develop an innovative use of multivariate factorial kriging in a spatio-temporal context to 1) identify process variables at multiple scales, 2) model potential and preferential habitat, and 3) perform inference of the processes which drive habitat selection. We will illustrate our approach with an investigation of habitat selection by a marine top predator, the common guillemot *Uria aalge*, in the Bay of Biscay (the French Atlantic continental shelf) during winter. We characterize the oceanographic variables that affect the spatial distribution of this predator and model its potential habitat at large scale as well as its preferential habitats at finer scale. This two-scales approach is justified by the capacity of common guillemot to make hierarchical decisions – first to find broad-scale spatial regions and then to locate small-scale habitat features where preys aggregate (Russel et al. 1992, Weimerkirsch et al. 1997, Pinaud and Weimerskirch 2007). In addition, marine predators are known to present scale-dependent patterns of spatial distribution (Weimerkirsch 2007). In a previous analysis of auks's spatial population structure, Bellier (2007) used a hierarchical geostatistical model that revealed that auk spatial distribution was clearly structured at three scales: 200, 50 and 10 km. The aim of the present study is to identify the environmental variables with which these structures may be associated, to distinguish process variables from circumstantial variables, and finally to model auk's habitat at multiple scale in the Bay of Biscay using only process variables.

Methods and data

Modelling approaches

Multivariate factorial kriging

The theory of MFK has been described in several publications and has principally been used in soil science (Matheron 1982, Goovaerts 1997, Chilès and Delfiner 1999, Wackernagel 2003). In this paper, we describe the basic computation involved in an MFK analysis as a three step procedure, and then propose an innovative application of MFK through time (Fig. 1). (1) Experimental variograms and cross-variograms computed for each variable (species distribution and environmental variables) in the data set are modelled using a linear model of coregionalisation (LMC). The information contained in the LMC is used for constructing multiple-scale variance-covariance matrices, called 'coregionalisation matrices' (CM). There is one coregionalisation matrix per spatial structure (i.e. scale) considered. (2) A principal component analysis (PCA) is applied to each CM separately. Eigenvectors and eigenvalues are then used as scale-specific transformations coefficients to compute the extracted principal components, called 'regionalized factor' in the framework of MFK. (3) Regionalized factors are estimated (mapped) over the study area using a specific cokriging technique.

The linear model of coregionalisation (LMC): modelling spatial relationships between organisms and their environment at multiple spatial scales

The spatial relationships between organisms and environment were quantified using cross-variograms. Cross-variograms $\gamma^*_{ij}(h)$ describe pairwise difference relationships between the response variables z_i and control variables z_j for different distance classes h gathering N pairs of location x_α, x_β according to their separation vectors $x_\alpha - x_\beta = h$:

$$\gamma^*_{ij}(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} \left(z_i(x_\alpha + h) - z_i(x_\alpha) \right)\left(z_j(x_\alpha + h) - z_j(x_\alpha) \right) \quad (1)$$

The total number of variograms and cross-variograms of the data is p(p + 1)/2 where p is the number of variables in the data set.

In order to link critical spatial scales of organism distribution to environmental variables, the experimental variograms and cross-variogams are modelled by a series of nested variogram models. The resultant set of models is known as the linear model of coregionalisation (Goovaerts 1997, Wackernagel 2003). These models, although common in soil science, are still rarely applied in ecology. Bellier et al. (2007) have demonstrated that LMC are suitable tools to analyse species-environment relationships at multiple scales, because they explicitly account for the spatial structure of both response and control variables while identifying and quantifying species-environment relationships over a wide range of scales. Using LMC, it is possible to model simultaneously all experimental variograms and cross-variograms (Eq. 1) (Fig. 1). The LMC with k = 1,…,q structures may be written as:

$$\gamma_{ij}(h) = b_{ij}^{(1)} g^{(1)} + b_{ij}^{(k)} g^{(k)} + ... + b_{ij}^{(q)} g^{(q)} \quad (2)$$

Where $\gamma_{ij}(h)$ is the variogram model for variables i and j (for i = j, the variogram is obtained, and for i ≠ j the cross-variogram is obtained), the coefficients b_{ij}^k correspond to the partial sill of each variogram and cross-variogram function at each scale k considered and $g^{(k)}$ represents the type of variogram model (i.e. exponential, spherical, ….) for structure k. The selection of the basic models that constitute the LMC (i.e. the model families and range parameters) is realised by examining the main features of the experimental variograms and cross-variograms (Fig.1). This leaves only one parameter that differs between response and control variables at each scale: the partial sill b_{ij}^k, which represents the partial variance in the case of variograms and the partial covariance in the case of cross-variograms. The partial sill coefficients are estimated by the algorithm developed by Goulard and Voltz (1992) for multiple variogram fitting. This algorithm minimizes a least squares criterion under some constraints (see Goulard and Voltz 1992 or Wackernagel 2003 for details). In practice, LMC fitting means adjusting the coef-ficients b_{ij}^k of Eq. 2 in such way that variogram models best fit their experimental counterparts.

Principal component analysis

Coefficients of the LMC were arranged to construct the coregionalisation matrices. A coregionalisation matrix B^k for structure (i.e. scale) k is essentially a positive semi-definite symetric matrix with diagonal and off-diagonal elements being the partial sill b_{ij}^k of the variogram and cross-variograms models, respectively, derived from LMC fitting. In other words, a coregionalisation matrix is an estimated variance-covariance matrix that summarises species-environment relationships at a given spatial scale. Therefore, the total number of matrices to analyse is equal to the number of spatial structures estimated in the first step of the analysis. From theses matrices of coregionalisation at different scales k, a set of classical analyses can be performed. Structural correlation coefficients are computed with a PCA (principal component analysis) on each coregionalisation matrix. These scale-specific
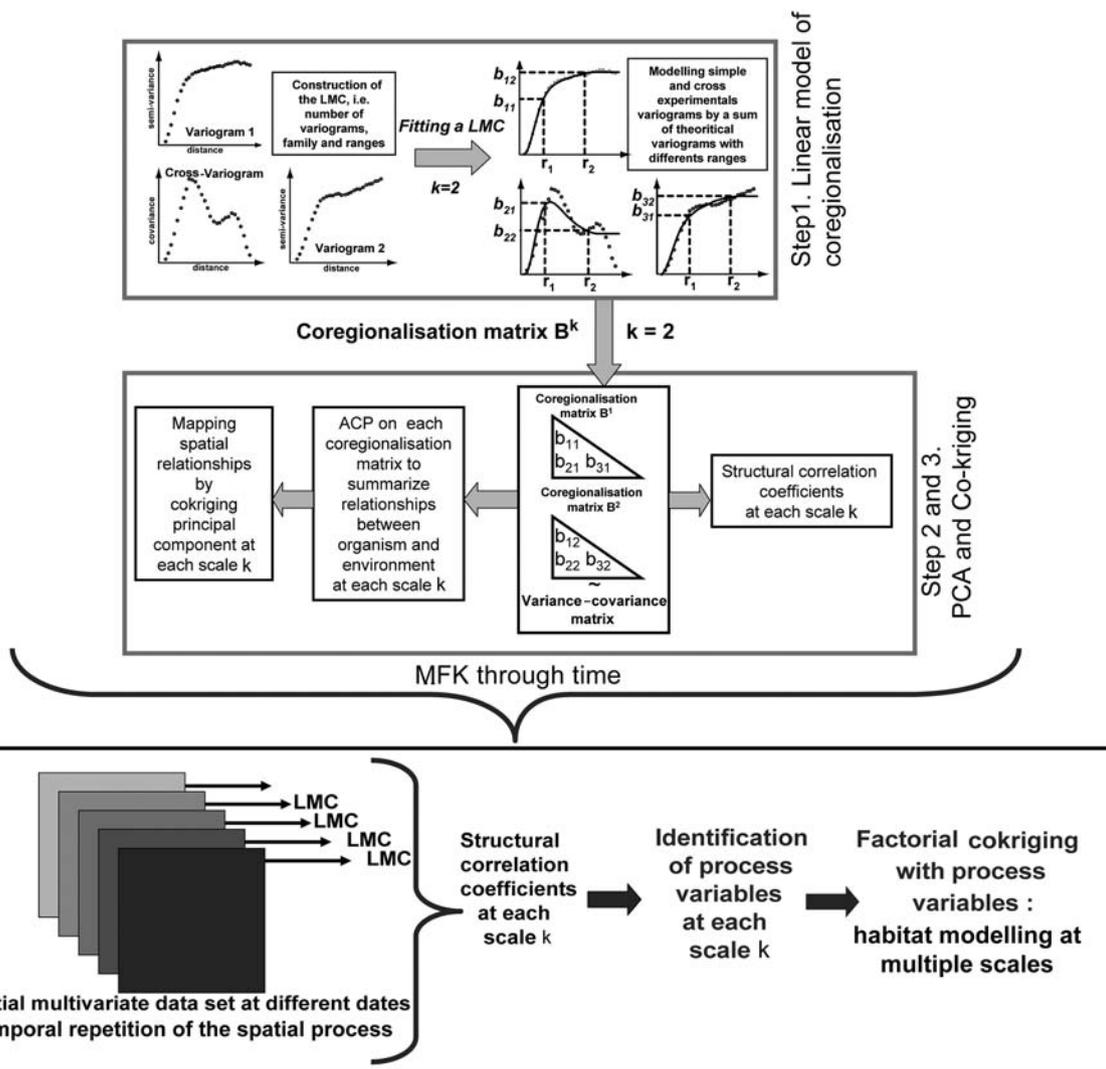

Figure 1. Schematic representation of each step of the multivariate factorial kriging analyses. Step 1: linear model of coregionalisation; step 2: principal component analysis, step 3: cokring regionalized factor. Bottom : description of the used of MFK through time for identifying process variable at each relevant scale and mapping potential habitat and preferential habitat.

PCA provide a set of transformation coefficients that can be used to compute new composite regionalized variables called regionalized factors (Appendix 1, part A). A regionalized factor is a model-derived unobserved spatial variable specific to the kth spatial scale resulting from a decomposition of the original variables (for details see Appendix 1, part A). In this way, an ecological variable structured at several scales can be decomposed in a set of factors structured at a single, specific scale.

Mapping the regionalized factors

The final step of MFK consists in mapping the extracted regionalized factors. The spatial estimation is realized with modified cokriging techniques (Goovaerts 1997) which allow for the estimation of the spatial regionalized components from the LMC properties (for details see Appendix 1, Part B).

Structural correlation coefficients: identifying process and circumstantial variables

Structural correlation coefficients can be derived from the partial sill coefficients (i.e. b_{ij}^k) to quantify the spatial links between organisms and environment (Fig. 1):

$$r_{ij}^k = \frac{b_{ij}^k}{\sqrt{b_{ii}^k b_{jj}^k}} \qquad (3)$$

We thus computed structural correlation coefficient from Eq. 3 for each environmental variable and for each temporal repetition of the survey. For each individual survey (i.e. temporal repetition of the process under study) carried, this coefficient provides information on the strength and sign of the spatial link between species distribution and the environmental variable considered, thus allowing for discrimination between process variables and circumstantial variables.

Mapping potential and preferential habitat

Here we extend the use of MFK to the modelling of potential and preferential habitat, based on the kriging analyses and on the previous identification of process variables (Fig. 1). First a LMC (step 1 of MFK) is used to identify and model the relationships between process variables (i.e. identifying previously by structural correlation coefficients and auks). As a second step, the spatial relationship between response variables and environmental factors identified as process variables is mapped by extracting the first regionalized component of each scale k. The mapping of potential habitat is performed in step 3 of MFK. The resulting spatial distribution of the regionalized component at the broadest scale corresponds to the potential habitat whilst the spatial distribution of regionalized component at the finest scales corresponds to preferential habitats.

Geostatistical analyses and the LMC were performed in R (R Development Core Team 2005) with personal code development. The algorithm proposed by Goulard and Voltz (1992) was used to estimate the coefficients b_{ij}^k (i.e. partial sill) in order to fit the LMC to the experimental variograms and cross-variograms. This iterative algorithm is based on least square optimisation and ensures the positive semi-definiteness of all coregionalisation matrices B^k.

Data on species spatial distribution

Auks are small (40 cm length) and cryptic seabirds. They are dark pursuit divers who spend most of their time at sea swimming on the sea surface and diving to search and catch their food (Gaston and Jones 1998). Three species occur in the Bay of Biscay, and though it was not possible to distinguish between them, guillemot *Uria aalge* is by far much more common than razorbill *Alca torda* and puffin *Fratercula arctica* (Bretagnolle et al. 2004).

The spatial distribution of auks was sampled by aerial surveys during the winter 2001–2002 (Certain and Bretagnolle 2008). The sampling scheme (Fig. 2) consisted in 5000 km of aerial transects (24 transects perpendicular to the coast and spaced every 20 km) on the continental shelf of the Bay of Biscay (100 000 km² on the French Atlantic coast). Six surveys were conducted monthly between October 2001 and March 2002, totalling 30 000 km of aerial transects. During the wintering period, seabird distribution is not influenced by colony location or breeding constraints. Data were collected at 150 m above the sea level and at 150 km h⁻¹ using strip-transect aerial survey methods (Eberhardt 1978, Briggs et al. 1985). The strip was 230-m wide, and was materialized by permanent marks on the wings. Only data from surveys carried out between November 2001 and March 2002 were analysed here, because too few auks were sighted in the October 2001 survey.

Before statistical analyses, data were pre-processed as follows: transects were sliced into a succession of rectangular 5-km long cells of known area. The 5 km length was chosen as a compromise between larger cells in which fine-scale structures would have been lost and smaller cells which would have resulted into zero-inflated statistical distribution. The cells were geographically localized and for each cell the number of sightings and the number of individuals were recorded (the two may differ when a sighting contains several individuals). The spatial distribution of auks in the Bay of Biscay was shown to be spatially structured at three scales (Bellier 2007): a determinist trend at large scale (200 km), a medium (50 km) and a fine (10 km) scale. At the finest scale auk's spatial structure presumably results from biotic processes (either the distribution of prey or auk behaviour: (Fauchald 1999, Fauchald et al. 2000, Fauchald and Tveraa 2006). However, available oceanographic variables are not spatially structured at theses scales. We therefore chose to keep only the large and medium scale structures, and filtered out the fine-scale structure using filter kriging (see Wackernagel 2003 for technical details and Bellier et al. 2007 for an application on filter kriging).

Environmental variables

Hydrological data from hydro-dynamic simulations

Environmental variables we derived from simulations generated by the three-dimensional hydrodynamical model MARS 3D. This model simulates the hydrological structure of the continental shelf of the Bay of Biscay and is calibrated with field data (Lazure and Jégou 1998, Lazure and Dumas 2008). The model simulations were performed with a 5-km horizontal resolution and 10 vertical layers in sigma
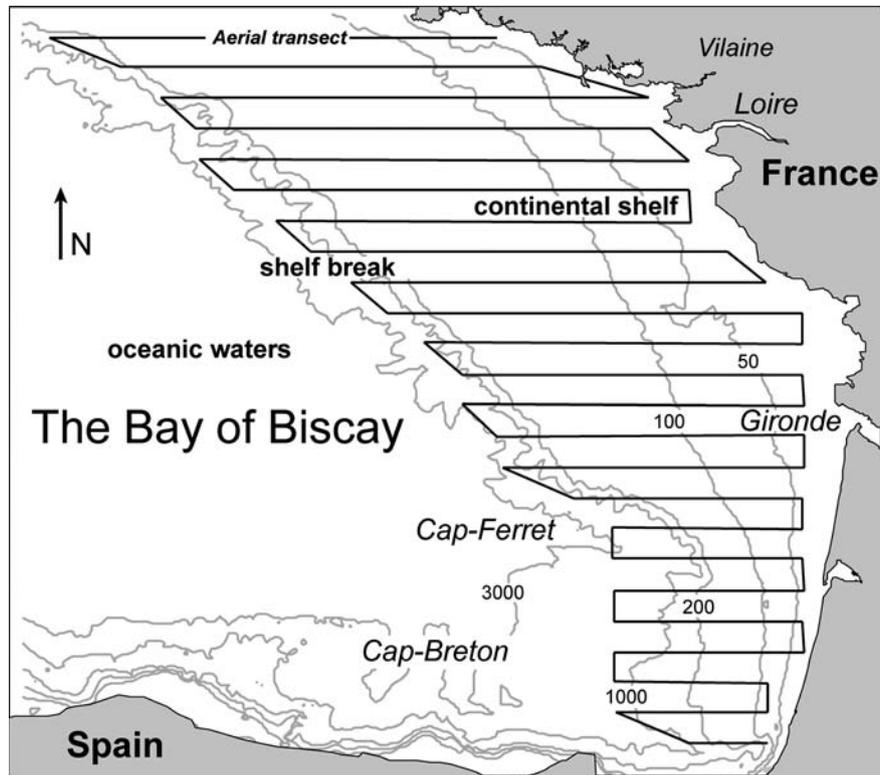
Figure 2. Sampling scheme used during the study, composed of 24 transects (totalling ca 5000 km) over the continental shelf (100 000 km²) of the Bay of Biscay.

coordinates (i.e. the thickness of the layers is proportional to bottom depth). Hydrological data were extracted for every month surveyed, (i.e. November 2001 to March 2002). To reduce the amount of data and simplify the interpretation of model outputs, only a few key variables were retained. The selection of hydrological variables was made by following the four criteria described in Planque et al.(2004). These criteria are as follows: (1) the modelled variables can also be observed in the field; (2) they are relevant to hydrology and biology (3) they are less variable in time than in space; and (4) the number of retained variables retained should be small in order to avoid difficulty in the estimation of the model parameters if the number of covariables is too large. The four variables retained were: surface salinity (SS), bottom temperature (BT), mixed layer depth (MLD) and potential energy deficit (PED: a measure of the amount of energy needed to vertically re-homogenize the water column). SS is primarily related to the volume of river outflow over the shelf, while BT reflects a combination of previous seasonal situations and heat exchanges through the water column. MLD provides an indication of the depth of vertical stratification of the water column and PED reflects the intensity of the stratification. SS and BT were directly extracted from the model outputs, while MLD and PED were calculated from the vertical density profiles (Planque et al. 2006).

Chlorophyll a data

We used ocean colour data produced by the seaWIFS Project at Goddard Space flight Center. The estimation of the concentration of chlorophyll a was performed using the empirical algorithm OC5 (Gohin et al. 2002). Chlorophyll a concentrations were used as a proxy of the abundance of phytoplankton in the sea surface layer.

Results

Modelling relationships between auk distribution and hydrological variables at multiple spatial scales

The relationships between auk spatial distribution and their environment can be found in the experimental cross-variogram for each environmental variable (Fig. 3). According to the spatial structure identified in Bellier et al. (unpubl.), the linear model of coregionalization (Eq. 2) is constructed by adding a spherical model with a range around 80 km to a gaussian model with a range around 150 km (see Table 1 for the exact values of range variables for each month). The LMC was fitted by means of the iterative algorithm and the best model was chosen, as suggested by Goulard and Voltz (1992), by comparing the goodness of fit in terms of the weigthed residual sum of squares of several combinations of functions with different ranges.

As an example of LMC fitting, Fig. 3 presents the variograms of auk distribution and of the five environmental covariates, and the associated cross-variograms for the month of December. The spatial covariance between auk spatial distribution and bottom temperature (BT) and also
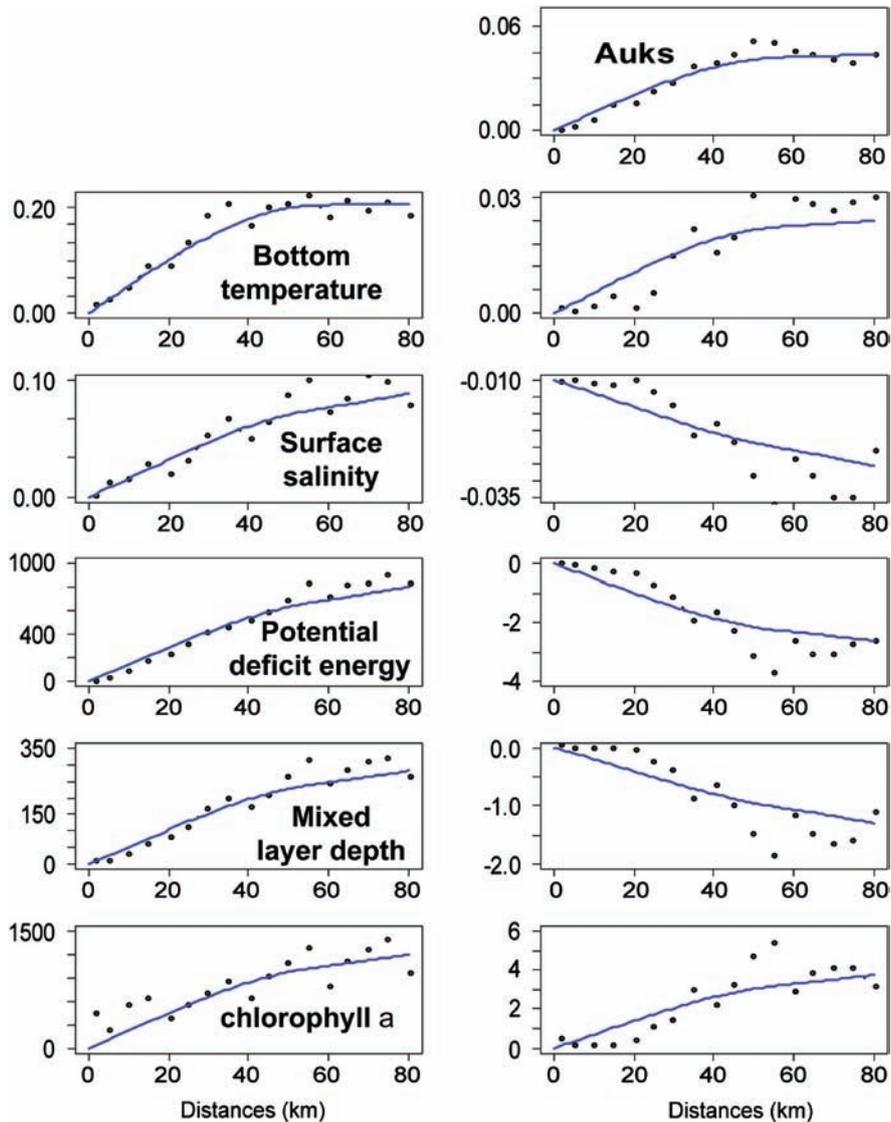
Figure 3. Experimental variograms and cross-variograms of the six variables for survey 3 (December) modelled by linear model of coregionalisation. Dotted line: empirical variogram and cross-variogram; solid line: the fitted linear model of coregionalisation. The analyses focus on the cross-variograms of the left-hand column. Distance unit on X-axis is in km.

chlorophyll a (chl a) increases from 0 to 40 km, highlighting a positive relationship between auks and these two covariates. Conversely, the cross-variogams combining auk distribution and the other three hydrological variables (i.e. SS, PED and MLD) present a negative covariance which decreases from 0 to 60 km.

Process variables at broad and medium scales

At a large-scale (i.e. ~200 km), we found that the relationships between auk distribution and hydrological variables were marked and persistent (Fig. 4a): for surface salinity and mixed-layer depth, the sign of the correlation was always negative (> –0.7), while for chl a it was always positive with a small dispersion of the coefficients. The three latter variables therefore appeared as possible 'processes variables' for controlling auk distribution at this scale. In contrast, the dispersion of the coefficients of correlation of Bottom temperature (BT) and potential energy deficit (PED) was

high and moreover, either positive or negative. These latter two variables were thus likely 'circumstantial variables'.

At medium-scale (i.e. ~50 km), no stable relationships between auk distribution and the hydrological variables could be detected (Fig. 4b). Except for chlorophyll a, the dispersion of the correlation coefficients were high and their signs were either negative or positive depending on month, indicating therefore that the relationship between auk distribution and environmental variables were not persistent in time. For chlorophyll a, all correlation coefficients were positive, but all but one values was relatively low; chlorophyll a could be the only environmental variables which being considered as a process variable at medium scale.

From the above results, we can thus expect the presence of auks in area with high concentration of chl a, weak salinity and a low depth of the mixed layer depth, oceanographic conditions that are prevalent in river plumes (Planque et al. 2006).

Table 1. Modelling parameters of nested variograms. The partial sill only concerns the variogram of auk distribution.

Surveys	Scales	Range (km)	Model type	Partial sill
Survey 2	preferencial habitat	45	spherical	0.014
	potential habitat	75	gaussian	0.027
Survey 3	preferencial habitat	48	spherical	0.046
	potential habitat	80	gaussian	0.019
Survey 4	preferencial habitat	35	spherical	0.015
	potential habitat	80	gaussian	0.0038
Survey 5	preferencial habitat	40	spherical	0.030
	potential habitat	120	gaussian	0.024
Survey 6	preferencial habitat	30	spherical	0.011
	potential habitat	80	gaussian	0.019

The relationships between auk spatial distribution and process variables were mapped with multivariate kriging methods. Based on its relationships with the three process variables (SS, MLD and Chl a), the potential distribution of auks at large-scale is shown in Fig. 5 (middle graphs), which may be viewed as a map of its potential habitat in the Bay of Biscay during the wintering season. For all surveys, the maps reveal a relatively similar potential habitat located along the coast, with high values of probability of presence in front of the main rivers (Gironde and Loire), at least for surveys 2, 3 and 6.

A similar procedure was used to map the preferential habitats at medium scale: from the process variable identified at this scale (i.e. chl a), a LMC was fitted to a set of variables including auks and chl a. Map of preferential habitat are shown for each survey (Fig. 5, lower part), though the structural correlation coefficients are low for survey 2 (0.138) and survey 6 (0.009), so the result for these two survey should be considered with caution. The maps of preferential habitat we produced also allow a finest description of auk's habitat selection in the Bay of Biscay. It reveals that preferential habitats are composed of a succession of patches along the coast; for survey 3, patch of high probability of presence are located inside the potential habitat and in front of Gironde estuaries. For survey 4 and 5, some patches of preferential habitat are located in the south of the Bay of Biscay out of the potential habitat for these surveys. Localization of preferential habitats varies through time within the potential habitat, with some patches out of the potential habitat.
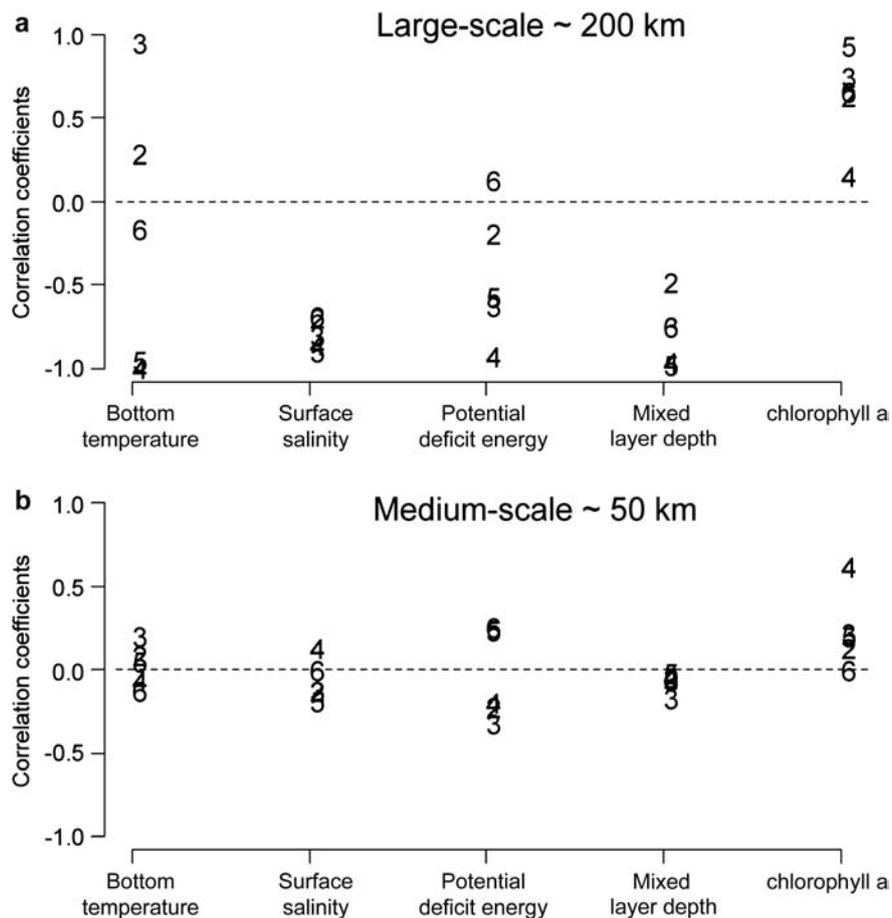

Figure 4. Values of the structural correlation coefficients between the auks and each oceanographic variables for each survey at each relevant scale. Top: correlation coefficients for each variable and auk distribution at large scale. Bottom: correlation coefficients for each variable and auk distribution at medium scale. The number correspond to the survey number: 2: survey 2, November 2001; 3: survey 3, December 2001; 4: survey 4, January 2002; 5: survey 5, February 2002; 6: survey 6 : March 2002.
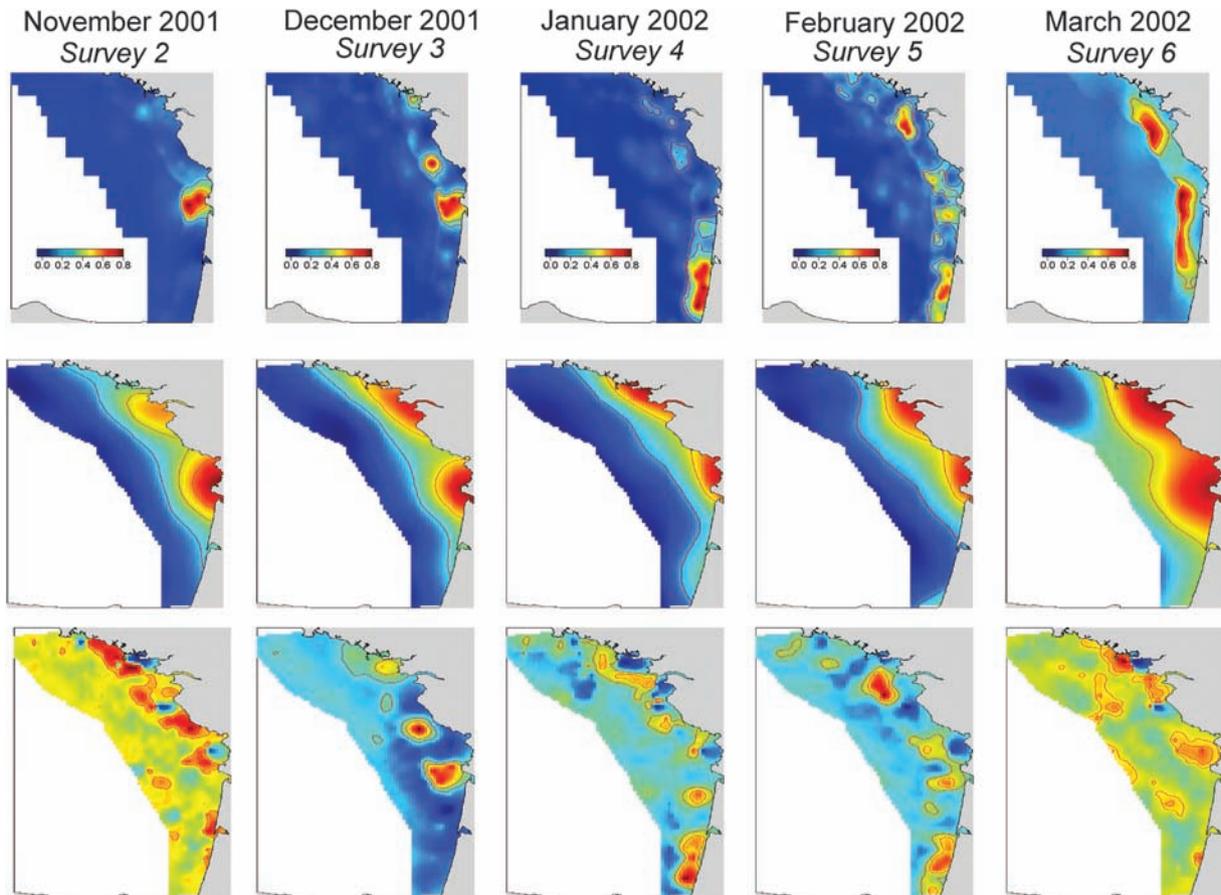
Figure 5. Top: maps of large-scale distribution of auk sightings obtained by filter kriging for each survey (i.e. the fine scale distribution has been filtered out). Middle: cokriged map of the first factor of the large scale structure from the auks (density of sightings) and process variables (i.e. SS, MLD, chl a) which represent potential habitat for each survey. Down: cokriged map of the first factor of the medium scale structure from the auks (density of sightings) and process variables (i.e. chl a) which represent preferential habitats for each survey.

Discussion

Auks and the oceanic environment

Our analyses indicate that the relationships between auk distribution and several oceanographic variables are scale-dependent. At large scale, the distribution of auks is (negatively) related to three variables: surface salinity, mixed layer depth and chlorophyll a. The former two are directly involved in the haline front in the Bay of Biscay (Puillat et al. 2004), in front of Loire and Gironde river plumes, two areas where auks are concentrated (Bretagnolle et al. 2004, Certain et al. 2007). Furthermore, haline stratification due to fresh water runoff from estuaries during winter was shown to enhance early phytoplankton production in the Gironde plume and adjacent oceanic waters (Puillat et al. 2004). Such phytoplankton production mainly occurs in near-surface waters, and can be detected by remote sensing and was therefore presumably taken into account in the chl a index used in the present study. Estuaries and river plumes may constitute primary habitats for pelagic fish as they present the properties of the 'ocean triad': enrichment, concentration, and retention (Bakun 1996, Agostini and Bakun 2003). Enrichment is provided by the nutrient-rich freshwater input, the process of concentration occurs in haline fronts and retention is known to exist over the continental shelf in front of the river plume (Allain et al. 2003). With such characteristics, waters in, and surrounding river plumes can provide areas suitable for high concentration of small pelagic fish (principally anchovies), the main prey of auks in the Bay of Biscay (Pasquet 1986). Therefore, these areas are likely very attractive foraging areas for auks at medium scale. At this scale, from one month to the other, auks moved around three main areas: the Loire and Gironde river plumes and another further south (Fig. 5). However, it should be noted that although remote sensing measurements of chlorophyll a may provide a proxy for auk prey distribution (small pelagic fishes in this case), our analyses failed to relate chlorophyll a and auk distribution at fine-scale. These results support previous observations that auk habitat selection may involve a set of multiple scales decisions by which animals first identify coarse-scale regions where prey availability is likely to be greater than average, and then locate small-scale habitat patterns where prey potentially aggregate in higher densities (Russel et al. 1992, Fauchald et al. 2000).

Identifying relevant environmental variables at multiple scales

This analytical work relies on the principle that the scale and nature of the relationships between a population and its environment have to be carefully checked before building any predictive model. In the present study, we have identified key process variables that are essential in the understanding of ecological processes controlling animal distribution at specific and various spatial scales. This finding has important implications in the field of habitat modelling and predictions. In the case of auks, the use of bottom temperature to predict auk distribution was not appropriate because the spatial correlation between the bottom temperature and auks varied from one month to the other in an unpredictable way, leading to wrong predictions most times of the year. Conversely, surface salinity, mixed layer depth and chlorophyll a were better predictors and more likely to be involved in the processes that govern auk distribution in the Bay of Biscay, but only at large scale. As a consequence, the spatial distribution of auks in the Bay of Biscay may be predicted from environmental variables only at large scale, and such predictions can be viewed as the potential habitat or the upper patch of the hierarchical patch dynamic system (sensu Wu and Loucks 1995). In contrast, modelling the preferential habitat (i.e. at fine scale) of auks from the available data could not be achieved for the whole period as the only variable displaying a temporally stable correlation coefficient with auk distribution at coarse scale was chlorophyll a. Further work is needed to predict the preferential habitat at coarse scale, in particular to identify environmental factors which are more strongly correlated to auk distribution at this scale (e.g. prey distribution). However, the linear model of coregionlisation assumes that each spatial structure is independent from the others. Therefore, we are not able to quantify trans-scale interactions, such as how the larger scale processes impose some constraints to the finer scales or how small scale processes interact to generate complex spatial patterns detected on a larger scale (Urban et al. 1987, George and Zack 2001). One way to explicitly account for trans-scale interactions (e.g use the large scale structure as a prior of the medium scale structure) could be the use of hierarchical statistical modelling (Cressie et al. 2009). This statistical framework based on conditional probabilities might allow for the definition of large scale structure given the medium one or inversely.

Conclusion

The method presented here is an effective framework to model the potential and preferential habitat of a species and can be related to the ecological niche concept re-emphasized in a number of recent studies (Pulliam 2000, Kearney 2006, Soberon 2007, Kearney and Porter 2009).

The classical Hutchinson's niche (a n-dimensional hypervolume, Hutchinson 1957) can be separated into two classes: the Grinnellian class which is defined by environmental conditions on broad scales (Grinnell 1917, Whittaker et al. 1973, Peterson 2003); and the Eltonian class which focuses on biotic interactions and resource–consumer dynamics that can be measured principally at local scales (Elton 1927, Leibold 1996). Kearney (2006) emphasized a clear distinction between niche and habitat, and respectively between mechanistic and correlative modelling. A niche would be defined mechanistically using for example physiological experiments, while habitat would be defined by more classical correlative methods such as the analysis of field survey data. Soberon (2007) further proposed a modelling framework to transpose Grinnellian and Eltonian niche into maps, i.e. habitats sensu Kearney (2006). Carrying repeated field surveys, where each corresponds to a 'natural' experiments, is an alternative to laboratory experiments for defining niche descriptors (i.e. as axes in the n-dimensional hypervolume), useful for any study models for which experiments can't be carried out. However, since it is relying on field data, it will hardly account for the fundamental niche since field data are drawn from realised niches. But the fundamental niche may be approximated by a large collection of temporal repetitions.

When applied to these repeated surveys, the statistical developments presented in this study constitutes a powerful method to achieve both a statistically robust and an ecologically meaningful classification of the variables that are required to model habitat; together with an explicit consideration of the spatial scale parameters. Focusing explicitly on temporal variability, our method allows for the clear identification of process variables that can be used as niche descriptors into species distribution models (Kearney and Porter 2009). One of the important questions involved in niche modelling is whether including variables (such as biological variables reflecting trophic or competitive interactions) in the predictive models or not. This latter question also relates to the need of the explicit considerations of the spatial scale parameters. Focusing explicitly on scale dependence, our method allows for the distinction between the variables that should be used to define the Grinnellian and the Eltonian niche. Indeed, the identification of the variables composing the Eltonian niche is a crucial element of accurate niche modelling. Several studies have shown that considering biotic interactions when designing empirical models of species distributions increased the explanatory power of these models at a local to regional scale (Anderson et al. 2002, Leathwick 2002). Furthermore, Araújo and Luoto (2007) provide evidence that biotic interactions might also matter at macroecological scales, in particular by affecting models of species' distributions under climate change scenarios.

References

Agostini, V. N. and Bakun, A. 2003. 'Ocean triads' in the Mediterranean Sea : physicals mechanism potentially structuring reproductuve habitat suitability (with example application to European Anchovy, *Engraulis encrasicolus*). – Fish. Oceanogr. 11: 129–142.

Allain, G. et al. 2003. The selection process from larval to juveniles stages of anchovy (*Engraulis encrasicolus*) recruitment in the Bay of Biscay invistigated by Lagrangian simulations and comparative otolith growth. – Fish. Oceanogr. 12: 407–418.

Allen, T. F. and Starr, T. B. (eds) 1982. Hierarchy perspectives for ecological complexity. –Univ. of Chicago Press.

Allen, T. and Hoekstra, T. 1991. Role of heterogeneity in scaling of ecological systems under analysis. – In: Kolas, A. and Pickett, T. (eds), Ecological heterogeneity. Springer, pp. 47–48.

Anderson, R. P. et al. 2002. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. – Oikos 98: 3–16.

Araújo, M. B. and Luoto, M. 2007. The importance of biotic interactions for modelling species distributions under climate change. – Global Ecol. Biogeogr. 16: 743–753.

Bakun, A. 1996. Patterns in the ocean. Ocean processes and marine population dynamics. – California Sea Grant College System, La Jolla, CA.

Bellier, E. 2007. Spatial distribution within hierachical patchs dynamics systems: identifying species-environment spatial relationships at multiscales. PhD thesis. – Univ. de la Mediterranée Marseille.

Bellier, E. et al. 2007. Identifying spatial relationships at multiple scales: principal coordinates of neighbour matrices (PCNM) and geostatistical approaches. – Ecography 30: 385–399.

Bretagnolle, V. et al. 2004. Distribution maps and minimum abundance estimates for wintering auks in the Bay of Biscay, based on aerial surveys. – Aquat. Living Resour. 17: 353–360.

Briggs, K. T. et al. 1985. Comparison of ships and aerial surveys of birds at sea. – J. Wildlife Manage. 49: 405–411.

Castrignano, A. et al. 2000. A geostatistical approach to estimate the probability of occurence of *Tuber melanosporum* in relation to some soil properties. – Geoderma 98: 95–113.

Certain, G. and Bretagnolle, V. 2008. Monitoring seabirds in marine ecosystem: the use of strip-transect aerial surveys. – Remote Sensing Environ. 112: 3314–3322.

Certain, G. et al. 2007. Characterising the temporal variability of the spatial distribution of animals: an application to seabirds at sea. – Ecography 30: 695–708.

Chase, J. M. and Leibold, M. A. 2003. Ecological niches: linking classical and contemporary approaches. – Univ. of Chicago Press.

Chilès, J. P. and Delfiner, P. 1999. Geostatistics: modeling spatial uncertainty. – Wiley.

Cressie, N. et al. 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. – Ecol. Appl. 19: 553–570.

Dungan, J. L. et al. 2002. A balanced view of scale in spatial statistical analysis. – Ecography 25: 626–640.

Eberhardt, L. L. 1978. Appraising variability in population studies. – J. Wildlife Manage. 42: 207–238.

Elton, C. 1927. Animal ecology. – Sedgwick and Jackson, London.

Fauchald, P. 1999. Foraging in a hierarchical patch system. – Am. Nat. 153: 603–613.

Fauchald, P. and Tveraa, T. 2006. Hierarchical patch dynamics and animal movement pattern. – Oecologia 149: 1432–1939.

Fauchald, P. et al. 2000. Scale-dependent predator–prey interactions: the hierarchical spatial distribution of seabirds and prey. – Ecology 81: 773–783.

Gaston, A. J. and Jones, I. L. 1998. The auks. – Oxford Univ. Press.

Gaston, K. J. and Blackburn, T. M. 2000. Pattern and process in macroecology. – Blackwell.

George, T. L. and Zack, S. 2001. Spatial and temporal considerations in restoring habitat for wildlife. – Restor. Ecol. 9: 272–279.

Gohin, F. et al. 2002. A five channel chlorophyll algorithm applied to SeaWIFS data processes by SeaDAS in coastal waters. – Int. J. Remote Sensing 23: 1639–1661.

Goovaerts, P. 1997. Geostatistics for natural resources evaluation. – Oxford Univ. Press.

Goulard, M. and Voltz, M. 1992. Linear coregionalisation model: tools for estimation and choice of cross-variogram matrix. – Math. Geol. 24: 269–286.

Grinnell, J. 1917. The niche-relationships of the California thrasher. – Auk 34: 427–433.

Hutchinson, G. E. 1957. Concluding remarks. –Cold Spring Harbor Symp. Quant. Biol. 22: 415–427. Reprinted in 1991: Classics in theoretical biology. Bull. Math. Biol. 53: 193–213.

Johnson, D. H. 1980. The comparison of usage and availability measurements for evaluating resource preference. – Ecology 61: 65–70.

Kearney, M. 2006. Habitat, environment and niche: what are we modelling? – Oikos 115: 186–191.

Kearney, M. and Porter, W. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. – Ecol. Lett. 12: 334–350.

Kotliar, N. and Wiens, J. 1990. Multiples scales of patchiness and patch structure: a hierarchical framework for the study of heterogeneity. – Oikos 59: 253–260.

Lazure, P. and Jégou, A. M. 1998. 3D modelling of seasonal evolution of Loire and Gironde plumes in the Bay of Biscay continental shelf. – Oceanol. Acta 21: 165–177.

Lazure, P. and Dumas, F. 2008. An external-internal mode coupling 3D hydrodynamical model for application at regional scale (MARS). – Adv. Water Resour. 31: 233–250.

Leathwick, J. R. 2002. Intra-generic competition among *Nothofagus* in New Zealand's primary indigenous forests. – Biodiv. Conserv. 11: 2117–2187.

Leibold, M. A. 1996. The niche concept revisited: mechanistic models and community context. – Ecology 76: 1371–1382.

Levin, S. 1992. The problem of pattern in ecology. – Ecology 73: 1943–1967.

Lindenmayer, D. B. et al. 2007. The need for pluralism in landscape models: a reply to Dunn and Majer. – Oikos 116: 1419–1421.

Matheron, G. 1982. Pour une analyse krigeante de donnés régionalisées. – Centre de Géostatistique, Fontainebleau, France.

Mayor, S. J. et al. 2007. Spectrum of selectio: new approaches to detecting the scale-dependent response to habitat. – Ecology 88: 1634–1640.

Pacala, S. and Roughgarden, J. 1982. Spatial heterogeneity and interspecific competition. – Theor. Popul. Biol. 21: 92–113.

Pasquet, E. 1986. Etude du regime alimentaire et de la distribution des populations d'oiseaux arins: cas des alcidés hivernants le long des côtes atlantiques françaises. – Ministère de l'environnement, France, p. 50.

Pearson, R. and Dawson, T. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimatic envelop models useful? – Global Ecol. Biogeogr. 12: 361–371.

Pebesma, E. J. et al. 2005. Mapping sea bird densities over the North Sea: spatially aggregated estimates and temporal changes. – Environmetrics 16: 573–587.

Peterson, A. T. 2003. Predicting the geography of species invasions via ecological niche modeling. – Q. Rev. Biol. 78:419–433.

Pickett, S. T. A. and Cadenasso, M. L. 1995. Landscape ecology: spatial heterogeneity in ecological systems. – Science 269: 331–334.

Pinaud, D. and Weimerskirch, H. 2007. At-sea distribution and scale-dependent foraging behaviour of petrels and albatrosses :a comparative study. – J. Anim. Ecol. 76: 9–19.

Planque, B. et al. 2004. Detecting hydrological landscapes over the Bay of Biscay continental shelf in spring. – Climate Res. 28: 41–52.

Planque, B. et al. 2006. Typology of hydrological structures modelled and observed in the Bay of Biscay. – Sci. Mar. 70S1: 43–50.

Puillat, I. et al. 2004. Hydrological variability on the French continental shelf in the Bay of Biscay, during the 1990s. – Continental Shelf Res. 24: 1143–1163.

Pulliam, H. R. 2000. On the relationship between niche and distribution. – Ecol. Lett. 3: 349–361.

Reineking, B. M. et al. 2006. Environmental variability and allocations tradeoffs maintain species diversity in a process-based model of succulent plant communities. – Ecol. Modell. 199: 486–504.

Russel, R. W. et al. 1992. Foraging in a fractal environment: spatial patterns in a marine predator–prey systems. – Landscape Ecol. 7: 195–209.

Schaefer, J. A. and Messier, F. 1995. Habitat selection as a hierarchy: the spatial scales of winter foraging by muskoxen. – Ecography 19: 333–344.

Schröder, B. and Seppelt, R. 2006. Analysis of pattern-process interactions based on landscape models – overview, general and concepts, and methodological issue. – Ecol. Modell. 199: 505–516.

Soberon, J. 2007. Grinnelian and Eltonian niches and geographic distributions of species. – Ecol. Lett. 10: 1115–1123.

Urban, D. et al. 1987. A hierarchical perspective can help scientist understand spatial patterns. – Biosciences 37: 119–127.

Wackernagel, H. 2003. Multivariate geostatistics (3rd ed.). – Springer.

Wagner, H. and Fortin, M. 2005. Spatial analysis of landscapes: concept and statistics. – Ecology 86: 1975–1987.

Weimerkirsch, H. 2007. Are seabirds foraging for unpredictable resources? –Deep-Sea Res. II 54: 211–223.

Weimerkirsch, H. et al. 1997. Activity pattern of foraging in the wandering albatross : a marine predator with two modes of prey searching. – Mar. Ecol. Progr. Ser. 151: 245–254.

Whittaker, R. H. et al. 1973. Niche, habitat, ecotope. – Am. Nat. 955: 321–338.

Wu, J. 1999. Hierarchy of scaling: extrapolation along a scaling ladder. – Can. J. Remote Sensing 25: 367–380.

Wu, J. and Loucks, O. L. 1995. From balance of nature to hierarchical patch nature dynamics: a paradigm shift in ecology. – Q. Rev. Biol. 70: 439–466.

Wu, J. and David, J. L. 2002. A spatially explicit hierarchical approach to modelling complex ecological systems: theory and applications. – Ecol. Modell. 153: 7–26.

Appendix 1

Part A. Details of step 2 of MFK – decomposition of a multivariate random function

The linear model of coregionalization (LMC) permits the decomposition of the original random functions $Z_i(x_\alpha)$ into a linear combination of p x q mutually uncorrelated random functions $Y_{lk}(x_\alpha)$ called regionalized factors:

$$z_i(x_\alpha) = m_i(x_\alpha) + \sum_{k=1}^{q} \sum_{l=1}^{p} a_{il}^k Y_{lk}(x_\alpha) \quad i=1, \ldots\ldots p \tag{A1}$$

where mi is the varying mean of $Z_i(x_\alpha)$ and a_{il}^k are an unknown transformation coefficients corresponding to the lth regionalized factor of the ith variable for the kth structure.

Equation A1 may be written in matrix form:

$$z = m + \sum_{k=1}^{q} A_i Y_k \tag{A2}$$

In fact, only the coregionalization matrix for the kth structure can be estimated:

$$B_k = A_k A_k' \tag{A3}$$

However, PCA can provide a natural determination of matrix A_k (Chilés and Delfiner 1999). More specifically, PCA, when applied to each coregionalization matrix B_k separately, provides a set of p eigenvalues λ and their corresponding eigenvectors x_l:

$$B_k = Q_k V_k Q_k' \tag{A4}$$

where Q_k is the orthogonal matrix of eigenvectors and V_k is the diagonal matrix of eigenvalues of the coregionalization matrix B_k. The matrix A_k can then be estimated as

$$A_k = Q_k \wedge_k^{1/2} \tag{A5}$$

The regionalized factors $Y_{lk}(x_\alpha)$ defined in Eq. A1 are the equivalent principal components of a classical PCA.

Part B. Details of step 3 of MFK – mapping the regionalised factors

Usually, cokriging provides a spatial estimation of the primary variable using data for the primary variable and one or more secondary variables. Typically, the primary variable is sampled over a limited number of points, while secondary data are more densely sampled. The estimation of the regionalized factors is a special case of cokriging where no measurements for the primary variable are available. The absence of primary data implies that the cross-covariance between regionalized factors and the original regionalized variables cannot be inferred. However, the spatial estimation of a regionalized factor is possible owing to the LMC properties: based on decomposition Eq. A1, the cross-covariance between $Z_i(x_\alpha)$ and $Y_{lk}(x_\alpha)$ is determined as (Goovaerts 1997):

$$Cov\{Z_i(x_\alpha), Y_{lk}(x+h)\} = a_{il}^k c^k(h) \tag{A6}$$

where $c^k(h)$ is the covariance (for the kth structure) for lag h, while a_{il}^k is the element of the ith row and lth column of matrix A_k.

In that sense, the LMC is very convenient because it allows inference of the cross-covariance between the original variables and unavailable regionalized factors using only the direct and cross-covariance models of the original variables (Goovaerts 1997). The optimal cokriging weights $\lambda_{\beta_i l}^k$ assigned to each data location β_i for the lth regionalized factor are given by the solution of the following cokriging system of equations:

$$\sum_{j=1}^{p} \sum_{\beta_j=1}^{n_j} \lambda_{\beta_j l}^k C_{ij}(x_{\alpha_i} - x_{\beta_i}) + \mu_{il}^k = a_{il}^k c^k (x_{\alpha_i} - x_{\beta_i}) \qquad (A7)$$

$$\sum_{\beta_i}^{n} \lambda_{\beta_i l}^k = 0$$

where n_i and n_j are the number of sample locations for the original variables i and j, respectively, and $C_{ij}(x_{\alpha_i} - x_{\beta_i})$ is the auto- or cross-covariance for variables i and j between locations x_{α_i} and x_{β_i}.

Finally, μ_{il}^k is the Lagrange multiplier for the ith variable and the lth regionalized factor. System differs from the classical cokriging system of equations in the way the cross-covariance between primary and secondary data is computed (i.e. variable $Z_i(x_\alpha)$ and the lth regionalized factor of the kth spatial structure, $Y_{lk}(x_\alpha)$, as well as in the unbiasedness constraints (Goovaerts 1997).

References

Chilés, J.P. and Delfiner, P. 1999. Geostatistics: modelling spatial uncertainty. – Wiley.

Goovaerts, P. 1997. Geostatistics for natural resources evaluation. Applied geostatistics series. – Oxford Univ. Press.