Biodiversity and Conservation (2005) 14:3235–3253
DOI 10.1007/s10531-004-0444-2

Predicting distribution and density of European badger (*Meles meles*) setts in Denmark

J.U. JEPSEN[1,2,3,*], A.B. MADSEN[1], M. KARLSSON[1] and D. GROTH[1]
*[1]National Environmental Research Institute, Department of Wildlife Ecology and Biodiversity, Grenaavej 14, DK-8410 Rønde, Denmark; [2]University of Copenhagen, Department of Population Ecology,Universitetsparken 15, DK-2100 Copenhagen O, Denmark; [3]UFZ-Centre for Environmental Research, Leipzig-Halle, Department of Ecological Modelling, Permoserstrasse 15, D-04318 Leipzig, Germany; *Author for correspondence (e-mail: juj@dmu.dk; phone: + 45-89201700; fax: +45-89201515)*

Received 22 September 2003; accepted in revised form 18 May 2004

Key words: Badger, Denmark, Habitat modelling, Logistic regression, Meles meles, Prediction, Presence–absence, Sett density, Validation

Abstract. Predictive models of the spatial distribution and abundance of species based on habitat characteristics are finding increasing use in management and conservation. The European badger attracts interest as a model species both for conservation reasons and because of the important role the species is playing in understanding carnivore sociality. We developed a statistical habitat model based on presence/absence data on badger setts. To maximise the utility of the model in management, we limited the choice of model variables to those that had a clear basis in badger ecology and that could be obtained on a nation-wide digital format. We extrapolated the habitat model to a region in Denmark and developed a threshold-independent sett distribution algorithm to estimate sett densities. The habitat model was simpler than previously published models of badger sett habitat selection, but nevertheless had a predictive ability in excess of 80% judged against independent data. The sett distribution algorithm was able to simultaneously reproduce several observed patterns of sett density and distribution over the probability gradient. It thus represents a significant improvement over threshold-dependent methods used to discriminate between suitable and unsuitable habitat predicted by presence/absence regression models. Our approach demonstrates that a model of badger sett habitat suitability with high predictive power can be obtained using easily accessible map-variables and presence/absence data. This is a prerequisite for using habitat models as predictive tools over large areas. The use of a simple sett distribution algorithm circumvents the common problem of subjectively fixing a threshold to discriminate between suitable and unsuitable habitat. In conjunction the models presented here constitute an important contribution to the management of the badger in Denmark and, upon further validation, possibly to similar regions in Northern Europe.

Introduction

The importance of landscape structure and habitat characteristics for the spatial distribution of the European badger (*Meles meles* L.) has received considerable attention in recent years (e.g. Clements et al. 1988; Thornton 1988; Macdonald et al. 1996; Feore and Montgomery 1999; Virgos and Casanovas 1999; Wright et al. 2000; Hammond et al. 2001; Good et al. 2001; Revilla et al. 2001; Johnson et al. 2002; Revilla and Palomares 2002). From a management perspective the species attracts interest partly due to conservation concerns (e.g.

van der Zee et al. 1992) and partly due to its possible role in transmitting wildlife diseases to domestic animals (e.g. Nolan and Wilesmith 1994). From an ecological perspective the badger is intriguing due to its flexible social behaviour expressed through social group sizes in excess of 15 adult individuals in high-density populations in the UK (Johnson et al. 2001) and strict pair-living in low-density areas in Southern Europe (Rodríguez et al. 1996; Revilla et al. 1999).

The elusive behaviour of the badger renders direct censuses of population sizes difficult even in limited areas. Therefore the density of breeding setts is frequently used as a surrogate for badger population densities. As the badger's choice of sett habitat appears to be fairly consistent, predictive habitat models are attractive and useful tools in extrapolating information from a limited study area to larger regions. A number of statistical models are available describing the choice of sett habitat by badgers across the species' European range (UK: Thornton 1988; Macdonald et al. 1996; Wright et al. 2000; Ireland: Hammond et al. 2001; Switzerland: Good et al. 2001; Spain: Revilla et al. 2001). Some of these models (Thornton 1988; Macdonald et al. 1996; Hammond et al. 2001) are developed from sett density data, but most are – like the present – based on presence/absence data.

A statistical habitat model based on presence/absence data provides information on the habitat preferences of the species and allows extrapolation and prediction of the amount and location of suitable habitat in other areas. It provides directly an estimate of the probability that a given area is used. But the question how predicted probabilities of use relate to population densities (or in the case of the badger: to sett densities) remains unanswered. An answer requires that the modeller is able to judge, from the predicted continuous suitability gradient, whether habitat can be considered 'sufficiently suitable' for use. Only rarely can such a threshold be deduced from ecological knowledge and far too often the solution to the dilemma is an arbitrary cut-off level separating suitable from unsuitable habitat. We suggest a way to circumvent this dilemma by employing a simple algorithm that determines minimum acceptable distance between setts according to habitat suitability.

The majority of existing habitat models of badger sett use include a combination of map- and field-measured variables. For a habitat model to be useful in prediction in larger areas, however, the model parameters must be easy to obtain. While field-measured variables such as ground cover, vegetation height or areas of certain crops have in several cases been shown to be significant predictors for the location of badger setts (e.g. Macdonald et al. 1996; Good et al. 2001; Revilla et al. 2001), such variables can only rarely be determined for larger areas. They therefore place a severe limitation on the utility of the habitat model for predictive purposes. Moreover most such field characteristics change within time spans much shorter than the 'life span' of a badger sett. While they might be relevant for the activity around a given sett in a given year, they are unlikely to represent the conditions at the site at the time the sett was first established.

The aim of the current study has been to develop estimates of densities of European badger setts in a region of Denmark. It was done in a step-wise procedure. First, a statistical habitat model was developed based on presence–absence data predicting the probability that a given area is used as badger sett habitat. The habitat model was validated against independent field data. Second, the predictions of the habitat model were translated into realistic estimates of sett densities without using a pre-defined threshold for when a given area was suitable for use. The resulting density estimates were compared to observed patterns of sett distribution and densities in Denmark.

Methods

Study areas

The study areas were situated in the counties of Aarhus, Viborg and Vejle in East Jutland, Denmark (Figure 1, Table 1). The *Fussingø area* (15 km^2) was used to develop the habitat model. The area consists of a flat plateau (50–65 m above sea level) bordered by valleys (0–15 m above sea level) to the North and South. Mixed deciduous woodland is mainly situated on slopes, while the plateau is dominated by pasture and agricultural areas. The *Bjerringbro* study area served as a validation area. It consisted of 32 squares each of 0.5 × 0.5 km (in total 8 km^2), distributed randomly within a 110 km^2 area. The area is similar to the Fussingø area with respect to terrain, with flat plateaus (40–75 m above sea level) used for agriculture and a broad river valley (0–15 m above sea level) covered by pasture and extensively used areas. Wooded areas consist of mixed deciduous–coniferous stands and conifer plantations. Further validation data was obtained from an extensive survey of a large number (N = 33; in

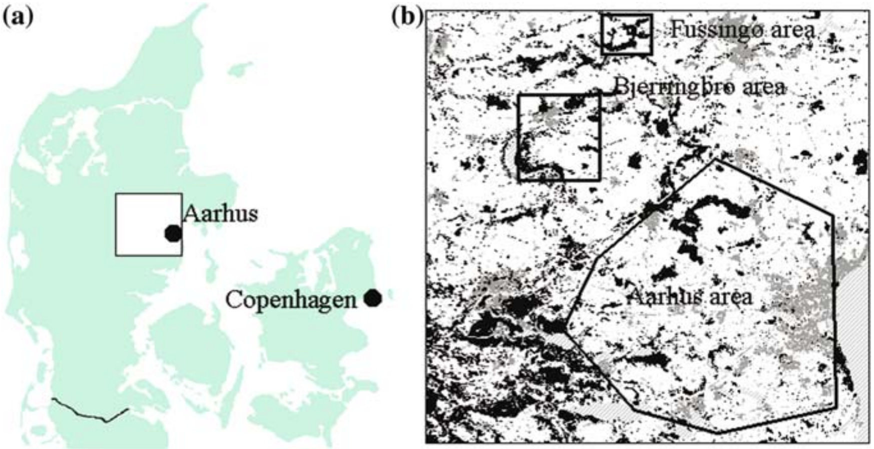

Figure 1. The location of the extrapolation area (55 × 55 km) in Denmark (a) and of the three study areas within the extrapolation area (b). Legends for (b): white, open land (mainly agriculture); grey, urban; black, forest; and hatched, water.

Table 1. Characteristics of the three study areas and the extrapolation area.

	Fussingø	Bjerringbro		Aarhus		Extrapolation
	Total ($15\ km^2$)	Surveyed area ($8\ km^2$)	Total ($110\ km^2$)	Surveyed area ($15\ km^2$)	Total ($845\ km^2$)	Total ($3025\ km^2$)
Habitat composition						
Agricultural	39.5	40.7	62.7	8.3	64.4	61.7
Woodland	48.9	43.9	21.7	84.6	10.4	14.9
Pasture	3.8	1.6	1.8	2.4	2.2	2.7
Urban + roads	3.4	6.6	9.6	1.8	16.0	11.2
Other	4.4	7.2	4.2	2.9	7.0	9.5
Observed sett densities (setts/km^2)						
Total setts	5.0	4.8	–	5.4	–	–
Main setts[a]	2.0	0.88	–	–	–	–

[a] Main setts are defined as all active setts with ≥3 entrances. See 'Definition of used and unused sites' for details.

total 15 km^2) of smaller woodlots distributed across the southern part of Aarhus county. It is referred to here as the *Aarhus* study area. The area consists of primarily deciduous woodlots interspersed in agricultural land. The habitat model was extrapolated to an area including all three study areas and covering a total of 3025 km^2 (2885 km^2 land area) or approximately 10% of the area of Jutland. The proportional area use in the extrapolation area is representative for Denmark as a whole. We define the extent of the study as the size of the extrapolation area (3025 km^2). The grain of the study (20 × 20 m raster cells) was determined by the resolution of the available spatial data.

Sett surveys

The surveys for setts in the Fussingø study area was carried out in the period April–June 1997 as part of a larger monitoring and mapping of badger setts (A. Prang unpublished data). Four surveyors went systematically through the area on foot and recorded the presence of setts. The distance kept between the surveyors depended on the density of the vegetation cover. The location (UTM co-ordinates), size (number of entrances) and activity status (occupied/unoccupied) were recorded for each sett. We assume that all setts within the intensively searched area were found. The Bjerringbro area was surveyed for setts during the period June–July 2001. It was not logistically possible to search the whole study area. The intensively surveyed area was therefore chosen as 32 non-overlapping 0.25 km^2 squares placed randomly along a gradient from open land to total forest cover (8 squares with 0–25% forest cover, 8 squares with 25–50% forest cover and so forth). The intensive survey followed the same procedure as in the Fussingø surveys, but was done by only two surveyors. Again we assume that all setts within the intensively searched squares were found. The Aarhus area was surveyed extensively along transects. It can therefore not be assumed that all setts within the area were found.

Spatial information

All spatial information was handled in Arcview® GIS 3.1 (ESRI Inc.). Land cover data (1:25000) for the extrapolation area were available from a national mapping (the Area Information System (AIS); Nielsen et al. 2000). A digital elevation model (DEM) was calculated for the extrapolation area based on 5-m contour intervals. From the DEM we derived floating point raster maps of the slope (degrees) and aspect (compass direction, 0–360°) with a resolution of 20 m.

Definition of used and unused sites

All sett sites located in the study areas were included in the analysis, irrespectively of activity status in the survey year (= 'used' sites). This is based on the assumption that any given sett represents a choice of habitat at the time of establishment, even if the sett was not in use at the particular time of the survey. It was not known with certainty which setts were main setts and which were not, since no attempts were made to confirm reproduction or to map territories borders. In the Fussingø and the Bjerringbro areas we made a careful guess, by defining all active setts which had ≥3 entrances, as main setts (Kruuk 1978; Good et al. 2001). This was required to get an estimate of mean distance between main setts. Since secondary setts presumably are less important for successful reproduction and survival than main setts, it is likely that habitat requirements for secondary setts are somewhat relaxed, compared to main setts. There is no reason to think however that the habitat cues used by badgers to identify a suitable location for a secondary sett are different or conflicting with those used when establishing a main sett. Secondary setts are sometimes used for reproduction (Neal and Cheeseman 1996, p. 171) and former secondary setts are known to have been 'upgraded' to main setts in response to an increase in population density (Ostler and Roper 1998). For a habitat model intended to predict the probability that a given area is used for badger setts it is therefore reasonable to include both main and secondary setts in the analysis, given that it can be justified that no systematic differences exists between them.

In each study area a sample of randomly distributed points separated by a minimum of 50 m was generated to represent non-selected sites (= 'unused' sites). Based on the assumption that all setts were found within the intensively searched areas in Fussingø and Bjerringbro, unused random points were not ground-checked in the field. Since the Aarhus study area was surveyed extensively, the random non-selected points were ground-checked to confirm that no badger setts were located within a 50 m radius around the point. In the two validation areas random points were only placed within the surveyed sub-areas (e.g. the 32 squares in the Bjerringbro area and the 33 woodlots in the Aarhus area). There is thus no room for bias caused by differences in habitat composition in the surveyed sub-areas and the study area as a whole. The sample of unused sites was larger than the sample of used sites in all areas (Fussingø: used, 75; unused, 100; Bjerringbro: used, 39; unused, 100; Aarhus: used, 81; unused, 98).

Environmental variables used in the habitat model

From knowledge of badger ecology we chose a number of potential explanatory variables related to land cover and terrain (Table 2). The spatial extent at which explanatory variables should be extracted is not clear. The choice of sett sites by badgers might be motivated both by on site characteristics and by neighbourhood attributes. We therefore chose to extract land cover and terrain variables for areas of different extent. The proportion of forest cover was calculated in circles of increasing radius (1–2000 m) around the site (F_1, F_{25}, F_{50}, F_{100}, F_{200}, F_{500}, F_{1000}, F_{2000}). F_1 is thus a binary variable indicating presence–absence of forest on site, while F_{25}–F_{2000} are floating point variables with values between 0 and 1. The basic terrain variables were ASPECT and SLOPE calculated from the DEM. There is good evidence that badgers select for terrain heterogeneity rather than slope and aspect in itself (e.g. Thornton 1988; Macdonald et al. 1996), and we therefore combined slope and aspect characteristics into a terrain heterogeneity index. The terrain index TRI (Nelleman and Fry 1995) is a commonly used measure based on the number of contour intercepts and the number of contour changes along either NS, EW or diagonal transects in grid cells. TRI has been applied successfully in a number of cases where terrain ruggedness is thought to be important (e.g. Nellemann and Fry 1995; Nellemann and Reynolds 1997; Eide et al. 2001). The disadvantage of this index is that the contour counts are not straightforward to automate in a GIS. Danks and Klein (2002) circumvented this by calculating Nellemanns TRI based on ruggedness measures derived from a DEM, rather than counts of contour lines. With slight modifications we followed the approach of Danks and Klein (2002). We calculated an automated terrain index (*Slope–Aspect Ruggedness Index*; SARI) for five areas of increasing extent ($s = 100$, 200, 500, 1000 and 2000 m) in the following way: Based on the floating point slope raster we calculated *slope heterogeneity* as the standard deviation (SD) of SLOPE (SLSD) in a quadratic moving window with side length s. A given 20 × 20 m cell in the *SLSD* raster thus contains the SD of the slope values found in all cells in the surrounding window. The integer values of *ASPECT* were grouped in bins of 10 (0–10°, 10–20° etc.) yielding a total of 36 aspect groups. *Aspect heterogeneity* was then calculated as the number of different aspect groups found within the quadratic moving window (ASPVAR). Based on the intermediate SLSD and ASPVAR raster layers, we calculated SARI for each spatial extent (s) as $(\text{SLSD}_s * \text{ASPVAR}_s)/(\text{SLSD}_s + \text{ASPVAR}_s)$ following Nellemann and Fry (1995). Though badgers (and badger setts) are sometimes found in suburban or built-up areas, human settlements and infrastructure such as large roads, probably represent sources of disturbance as well as additional mortality. Several studies have found significant negative relationships between various

Table 2. Summary of the explanatory variables considered in the habitat model.

Variable	Description	All used sites ($n = 75$)		Main setts only ($n = 31$)		Unused sites ($n = 100$)
		Mean ± SD	r	Mean ± SD	r	Mean ± SD
SLOPE	**Slope (0–90 °)**	**7.37 ± 6.66***	**0.30**	**7.16 ± 6.56***	**0.34**	**3.25 ± 4.81**
ASPECT	Aspect (0–360 °)	148.7 ± 135.5 ns	0.32	190.7 ± 139.9 ns	0.18	116.6 ± 129.3
DINFRA	**Distance to infrastr. (m)**	**337.6 ± 172.7***	**0.33**	**358.4 ± 192.2***	**0.28**	**232.8 ± 188.0**
Forest neighbourhood measures						
F_1	Forest on site (y/n)	0.920 ± 0.27*	0.49	1.0 ± 0.0 *	0.47	0.450 ± 0.50
F_{25}	% forest in 25 m radius	0.915 ± 0.26*	0.50	1.0 ± 0.0*	0.50	0.458 ± 0.48
$\boldsymbol{F_{50}}$	**% forest in 50 m radius**	**0.912 ± 0.26***	**0.51**	**1.0 ± 0.0***	**0.52**	**0.455 ± 0.47**
F_{100}	% forest in 100 m radius	0.868 ± 0.25*	0.45	0.966 ± 0.05*	0.48	0.448 ± 0.43
F_{200}	% forest in 200 m radius	0.779 ± 0.24*	0.45	0.866 ± 0.11*	0.48	0.427 ± 0.37
F_{500}	% forest in 500 m radius	0.600 ± 0.18*	0.40	0.640 ± 0.12*	0.43	0.409 ± 0.25
F_{1000}	% forest in 1000 m radius	0.56 ± 0.17*	0.36	0.594 ± 0.12*	0.39	0.41 ± 0.22
F_{2000}	% forest in 2000 m radius	0.38 ± 0.12 ns	0.08	0.395 ± 0.10 ns	0.12	0.36 ± 0.13
Terrain indices						
$\mathbf{SARI_{100}}$	**SARI index, 100 × 100 m**	**2.88 ± 1.37***	**0.52**	**3.14 ± 1.27***	**0.52**	**1.23 ± 1.22**
SARI_{200}	SARI index, 200 × 200 m	3.92 ± 1.78*	0.49	4.35 ± 1.58*	0.51	1.93 ± 1.64
SARI_{500}	SARI index, 500 × 500 m	4.83 ± 1.53*	0.50	4.94 ± 1.38*	0.46	2.90 ± 1.77
SARI_{1000}	SARI index, 1000 × 1000 m	4.52 ± 1.15*	0.40	4.57 ± 0.99*	0.36	3.45 ± 1.31
SARI_{2000}	SARI index, 2000 × 2000 m	4.21 ± 0.63*	0.18	4.18 ± 0.54 ns	0.14	3.93 ± 0.77

Mean values ± SD are given for all used sites, used sites classified as main setts, and unused sites. *indicate significance ($p < 0.001$) of a Mann–Whitney *U*-test for differences in medians between the relevant used and the unused sample. r values indicate the degree of correlation with the dependent variable (Spearman's r). Explanatory variables highlighted in bold were considered in the logistic models.

measures of infrastructure and the presence of badger setts (e.g. Wright et al. 2000; Hammond et al. 2001). We therefore included a variable indicating the distance to infrastructure (DINFRA), with infrastructure being defined as all types of buildings, roads wider than 3 m and railroads.

Statistical analysis

The explanatory variables could not be considered normally distributed. The degree of correlation between the response variable and all the explanatory variables was therefore calculated using Spearman's correlation coefficient (r; Table 2) to evaluate the explanatory power of the variables extracted for areas with different extent. Variables describing the same environmental characteristics – for instance forest cover – at different spatial extents (ex. F_{200} and F_{500}) were considered alternative rather than complementary. Consequently they were never entered into the same model. To avoid including explanatory variables with a too high degree of co-linearity, we calculated Spearman's r between all variables entered into the logistic models. The variables related to terrain (SLOPE and SARI_{100}) showed as expected a fairly strong co-linearity ($r = 0.649$), but in all other pairs of variables correlation was weak to modest (0.07–0.49).

Land use data will often show a certain degree of spatial autocorrelation, due to the nature of the processes that shape landscape pattern (e.g. Carroll and Pearson 2000; Lennon 2000). The presence of a strong spatial autocorrelation in both response and explanatory variable will cause the autocorrelated explanatory variable to be shown significant more often than it should (Lennon 2000; Legendre et al. 2002). Consequently, non-significant explanatory variables can be retained in the regression model, and in the worst case, be interpreted as of ecological significance. We evaluated the degree of spatial autocorrelation in the explanatory variables entered into the final models at increasing spatial extent (s = 0–1000 m). This was done by calculating the correlation coefficient (Pearson's r) between the value of the variable at the site, and the mean value of the variable in a ring of inner radius s and width 1 cell (= 20 m) around the site (see Schadt et al. 2002 for an application of the same approach). The spatial correlation coefficient in all cases decreased with increasing spatial lag and was ≤ 0.4, which indicates that only a weak spatial autocorrelation was present even over small distances (Figure 2). Spatial autocorrelation was thus not considered of concern in the present data set.

A set of logistic models (PROC LOGISTIC, SAS Institute Inc., Cary, NC, USA) was constructed by adding one variable at a time, beginning with the strongest variable as judged from the r values in Table 2. The significance levels for entering and removing a variable from a model was set to $p < 0.05$ and $p < 0.10$, respectively. The goodness of fit of a model was evaluated using the Akaike Information Criterion (AIC; Manly et al. 2002), choosing the model with the lowest AIC value.

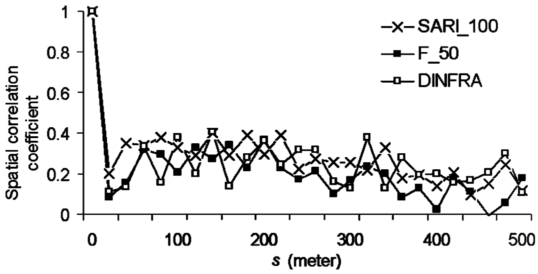

Figure 2. The spatial correlation coefficient (Pearson's r) of the three variables in the final model at increasing area extent s.

Habitat model validation procedures

Based on the final model the predicted probability of use (P) was calculated for each 20 × 20 m cell in the extrapolation area, excluding areas covered by open water (sea and lakes). The ability of the model to discriminate between used and unused sites both in the model data set (Fussingø) and in the two validation data sets (Bjerringbro and Aarhus) was evaluated using the area under the Receiver Operating Characteristics (ROC) curve (Fielding and Bell 1997; Pearce and Ferrier 2000). A ROC curve is obtained by plotting the *sensitivity* of the model (= the proportion of used sites correctly predicted to be used) against the *false positive fraction* (= the proportion of unused sites incorrectly predicted to be used) over a large number of threshold probabilities. The threshold probability is a predefined probability above which a cell is assumed to be suitable for use. We used 20 different threshold values (5% intervals). For a model with no discrimination capacity (random) the area under the ROC curve (AUC) will be 0.5, while for a model with perfect discrimination capacity AUC will be 1. AUC values between 0.7 and 0.8 indicate an acceptable discrimination capacity, while AUC values ≥0.8 indicate that the model has excellent discrimination capacity (Hosmer and Lemeshow 2000, p. 162). In addition we calculated the maximum model *accuracy* as the total proportion of correctly classified cells (Pearce and Ferrier 2000).

Sett distribution algorithm

The habitat model indicates the probability that a given cell contains a badger sett. To predict the distribution of setts in an area in which total sett density is unknown, an extrapolation algorithm is needed that is able to place setts in accordance with the probability map. We chose to do this by developing an algorithm that calculates a minimum acceptable distance (D) between a given sett and all of its neighbours, based on the probability score. This mimics the presence of an exclusive territory around each sett – a reasonable assumption for a territorial species such as the badger. We aimed at developing an algorithm that (i) Was able to reproduce observed patterns of sett densities in different probability classes in the Fussingø study area, (ii) Was functional over the whole range of probabilities, to avoid a predefined probability threshold for use and (iii) Required a minimum of additional information. The sett distribution algorithm that came closest to meeting all the criteria above was one assuming a logistic relationship between inter-sett distance (D) and probability of use (P) at a given site:

$$D = D_{\text{min}} + \frac{D_{\text{max}} - D_{\text{min}}}{1 + \exp(-a * P - b)},$$

where a and b are fitted constants ($a = -0.11$; $b = 4$) and D_{min} and D_{max} are the minimum and maximum allowed distance between main setts, defining the upper and lower limits of the distribution. In the Fussingø study area the distance between main setts and their closest main sett neighbour varies between 70 and 720 m with a mean of 212 m. We therefore fixed $D_{\text{min}} = 200$ m. While D_{min} is a parameter with a direct ecological relevance for a territorial species such as the badger, D_{max} is not. It was therefore varied over a large range (1000–6000 m). The sett distribution algorithm was initialised with a single sett randomly placed in the landscape. The rules for placing subsequent setts were then (i) Pick a random location in the landscape and get the value of P for this location, (ii) Calculate the minimum acceptable distance (D) to a potential neighbour following the equation above (iii) Evaluate the actual distance to every existing sett (iii) If a neighbour is closer than D pick a new random site and begin from (i). If no neighbour is closer than D, then establish a sett on location. For each value of D_{max} we ran 100 replicates.

The predictions of the sett distribution algorithm were evaluated against three different observed patterns in the Fussingø study area by calculating the sum of squares of the difference between model prediction (M_i) and data (D_i) over all probability classes

$$\text{SSQRT} = \sqrt{\frac{1}{n}\sum (D_i - M_i)^2},$$

where n is the number of probability classes. The three patterns were: (i) The total density of setts, (ii) The density of setts in different probability classes (e.g. 10–20%, 20–30% etc), and (iii) The relative number of setts in different probability classes. The final value of D_{max} was chosen as the one that best reproduced all three patterns simultaneously.

Results

Environmental variables

Used and unused sites were significantly different with regard to the mean value of all environmental variables, with the exception of ASPECT and the forest and terrain variable calculated for a large area extent (F_{2000} and SARI_{2000}; Table 2). Used badger sett sites tended to be located on slightly steeper slopes, with larger distance to infrastructure, higher forest cover and more heterogeneous terrain than the random unused sites. The correlation between the forest cover variables and the response variable decreased with distance beyond 200 m from the site, but there were only marginal differences between the variables calculated for a small area extent. Highest r value was found for F_{50}. Highest r value for the SARI index was found at the smallest area extent (100 m). It is important to note that, since F_{50} indicates the forest cover within a radius of 50 m from the site, and SARI_{100} indicates the terrain heterogeneity in a 100 × 100 m square around the site, the two variables both identify the immediate neighbourhood (approx. 50 m from site) as the unit of concern. There were no significant differences between the used sample including all sites and the used sample including only sites defined as main setts. Both samples identified the same set of significant variables with very similar r values (Table 2).

Habitat model, predictions and validation procedures

Initially a set of logistic regression models was developed based on the strongest terrain index (SARI_{100}). Due to the high correlation between SLOPE and SARI_{100} a parallel set of models was developed based on SLOPE yielding a total of 6 regression models (Table 3). The model that presented the best fit to the data was based on the SARI_{100} index and contained in addition the forest cover variable F_{50} and DINFRA (Table 3, M3). All three variables were highly significant.

Table 3. Summary of the logistic models. The final model, highlighted in bold types, is the model with the lowest AIC value.

Model	AIC	df	Intercept	Variable (i)	β_i	SE_i	p-value
Null model	241.0	174					
M1	187.7	173	−2.068	SARI_{100}	0.876	0.1375	<0.001
M2	171.0	172	−3.291	SARI_{100}	0.672	0.1495	<0.001
				F_{50}	2.199	0.5644	<0.001
M3	**165.6**	**171**	**−3.978**	**SARI_{100}**	**0.761**	**0.1581**	**<0.001**
				F_{50}	**1.628**	**0.5883**	**0.005**
				DINFRA	**0.003**	**0.001**	**0.008**
M4	222.2	173	−0.909	SLOPE	0.123	0.0293	<0.001
M5	190.5	172	−2.546	SLOPE	0.062	0.0314	0.049
				F_{50}	2.659	0.5371	<0.001
M6	189.0	171	−2.904	SLOPE	0.064	0.0318	0.041
				F_{50}	2.364	0.5492	<0.001
				DINFRA	0.002	0.001	0.066

Table 4. Validation of the habitat model predictions against the model data and the two independent validation data sets.

Study area	*n*		AUC	Accuracy	Threshold
	Used	Unused			
Fussingø	75	100	0.892	94.0	0.40
Bjerringbro	39	100	0.833	86.3	0.40
Aarhus	81	98	0.744	77.0	0.35
All validation data	**120**	**198**	**0.826**	**79.6**	**0.40**

AUC indicates the area under the ROC curve. The measure for model accuracy is threshold dependent. The value stated here is the maximum accuracy obtained. The 'Threshold' column indicates the corresponding threshold level.

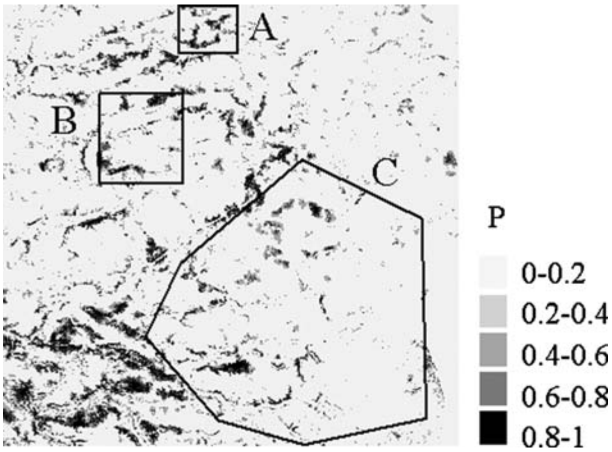

Figure 3. The predicted probability (*P*) of use in 0.2 intervals. The outlines of the three study areas are indicated (a) Fussingø, (b) Bjerringbro and (c) Aarhus.

The validation procedures indicated that the final model (Table 3, M3) had a satisfactory ability to discriminate between used and unused sites (Table 4). The model was extrapolated to the entire extrapolation area and the external validation against independent data was performed separately on data from each of the two validation areas and on the pooled data for both areas. Judged from the AUC values the model had a very high discrimination capacity: 0.892 for the Fussingø data set, 0.833 for the Bjerringbro data set, 0.744 for the Aarhus data set and 0.826 for the pooled validation data set.

A map of the predicted probability of use (*P*) for the extrapolation area is shown in Figure 3. The amount of suitable habitat predicted to be present in the extrapolation area varies between 0.2 and 23.7% of the total area depending on the value chosen for the threshold for use (e.g. the value of *p*

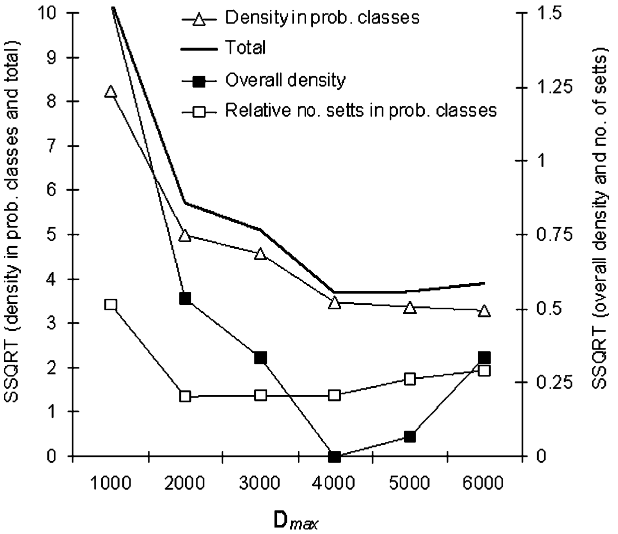

Figure 4. The comparison of the three predicted (for the whole extrapolation area) and observed (in the Fussingø area) patterns of sett density and distribution using different values of D_{max}. The SSQRT value indicate the degree of error between predicted and observed patterns (see text for details). The best fit to all three patterns is thus obtained where the total error (bold line) is at its minimum.

above which habitat is considered suitable). At the probability threshold where the model best describes the data ($p = 0.4$) the model predicts that 6.4% of the extrapolation area is suitable habitat.

Sett distribution and density

The comparison of the predictions of the sett distribution algorithm for different values of D_{max} with the three observed patterns of sett distribution in the Fussingø area is shown in Figure 4. The SSQRT values indicate the degree of error between the pattern predicted by the model and the observed pattern. An SSQRT value of zero indicates perfect fit. All three patterns confirmed that the sett algorithm was able to predict the distribution and density of setts reasonably well at D_{max} values above 2000 m. Both the relative number of setts and the density in different probability classes had limited sensitivity to the value of D_{max}. Overall density however was clearly best reproduced with a D_{max} value of 4000 m. All further analyses were therefore performed with a D_{max} value of 4000 m. The resulting map of predicted densities of main setts in the extrapolation area is given in Figure 5. An independent validation against the estimated main sett density in the Bjerringbro area (Table 5) indicated a good correspondence between model prediction (0.75 ± 0.18 setts/km^2) and observed density (0.88 setts/km^2).

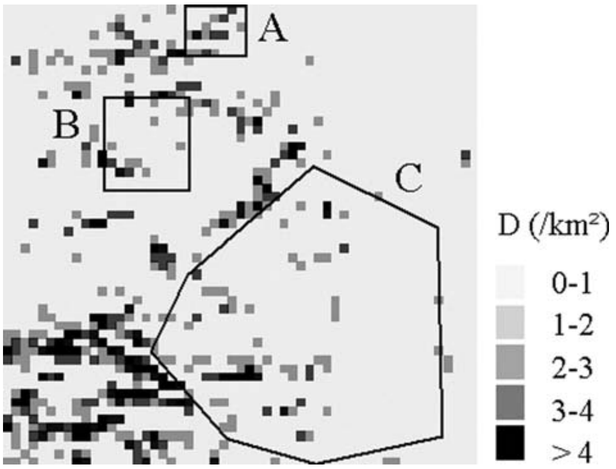

Figure 5. The predicted mean density (D) of main setts for the extrapolation area. The outlines of the three study areas are indicated (a) Fussingø, (b) Bjerringbro and (c) Aarhus.

Table 5. Final estimates of predicted main sett densities (mean ± SD of 100 replicates) in the study areas and the extrapolation area.

Area	Predicted sett density (setts/km^2)	Observed sett density(setts/km^2)
Extrapolation area	0.62 ± 0.01	–
Fussingø	1.95 ± 0.89	2.0
Bjerringbro – surveyed area	0.75 ± 0.18	0.88
Bjerringbro – total area	0.47 ± 0.86	–
Aarhus – surveyed area	0.43 ± 0.22	–
Aarhus – total area	0.30 ± 0.71	–

The observed main sett densities (see Table 1) added for comparison.

Discussion

Predictive models in management of the european badger

The European badger follows a fairly consistent pattern when selecting sett habitat. This renders predictive habitat modelling a very useful tool in the management of this species. We have presented a simple statistical model with high predictive ability and applied it in predicting sett distribution and densities in a region in central Denmark. We did this using a simple threshold-independent algorithm linking the predicted probability of use to inter-sett distances. Our aim was to optimise not only the predictive ability of the habitat model, but also the utility of the model in management. To achieve the latter we limited the choice of habitat model parameters to such that could be obtained on a digitised format on a national scale. This was done well knowing that certain variables that could potentially increase the fit of the model, such as on-site soil conditions, would have to be ignored. Our approach demonstrates however that it is possible to obtain a predictive ability that is above or comparable to previously published models of badger sett distributions (Wright et al. 2000), using just three explanatory variables. Each of the variables has documented relevance to badger ecology and is readily calculated in any GIS system. This renders the model economic to parameterise and allows resources to be channelled towards testing and validation rather than developing predictions. As such the habitat model fulfils the ideal requirements to a predictive management model (e.g. Mosher et al.1986).

Using presence/absence data to predict densities

Several previous predictive habitat models for the badger have been based on observed sett densities rather than presence/absence data (Thornton 1988; Macdonald et al. 1996). In a recent review of a large number of predictive habitat models developed for both plant and animal species, Pearce and Ferrier (2001) concluded that models based on abundance data generally had a poorer performance when applied to independent data than models based on presence/absence data. For species where reasonably accurate abundance models were developed, the corresponding (cheaper) presence/absence model performed equally well in indicating relative abundance, as did the abundance models. The present study suggests a method by which sett densities can be predicted from presence/absence data without including a predefined threshold for when habitat is sufficiently suitable for use. This approach rests on the assumption that sett densities are higher in areas predicted by the habitat model to have a high probability of use. This assumption is in agreement with data. The sett distribution algorithm predicts an overall sett density of 0.62 setts/km^2 in the extrapolation area. This is comparable to reported densities in Scotland, Ireland and low-density population in England at similar latitudes (Johnson et al. 2002 and references therein). Although limited data is currently available to validate this overall estimate of sett density, the fact that the sett distribution algorithm accurately predicts both the local density and the distribution of setts along the suitability gradient in the Fussingø area, as well as the overall density in the Bjerringbro area, lends credibility to the results. However, additional information about local sett densities is required to evaluate the general utility of the approach for developing threshold-independent predictions of sett densities.

Model performance and variable relevance

Local forest cover and terrain heterogeneity were the primary explanatory variables in the present model. The habitat model had a better performance in the Bjerringbro than in the Aarhus validation area. This is primarily due to differences in the forest cover in the surveyed area (Bjerringbro: 43.9% forest, Aarhus: 84.6% forest). An environmental constraint (here availability of forest cover) is only useful for prediction as long as it is limited. However, even in close to continuous forest cover (Aarhus area) the habitat model obtained a satisfying discriminating capacity.

That terrain attributes are important for the choice of sett sites for the badger is well founded in the literature (e.g. Macdonald et al. 1996; Good et al. 2001). It is less clear which attributes related to local terrain badgers are selecting for. Sloping areas offer better draining conditions, and facilitate digging. A favourable aspect increases microclimatic conditions. Local terrain heterogeneity shields the sett from view and might thus decrease disturbance (Good et al. 2001) and topographically heterogeneous areas are likely to consist of more varied substrate, thus increasing the chance that a suitable geological microsite is available. The terrain heterogeneity index (SARI) tested in the present study combines all these features into a single aggregated variable. It is worth nothing that all three variables in the final model describe local neighbourhood rather than on-site characteristics. Both the forest cover and the terrain variable identify the same extent: the area within approx. 50 m from the sett site appears to be the unit of concern. This area exceeds the extent of most badger setts, but is much smaller than a home range. This indicates firstly that very small and perhaps isolated forest patches are less preferred than larger or more connected ones. Secondly it shows that the cues (e.g. forest cover, terrain heterogeneity) used for sett site selection are perceived locally and their presence in the surrounding home range is of little or no importance. This highlights the importance of investigating scale in habitat modelling as also recently demonstrated by Schadt et al. (2002).

Sett density as a surrogate for badger densities?

The density of breeding setts has often been used as a surrogate for population densities (e.g. Thornton 1988; Cresswell et al. 1989; Roper 1993; Virgos 2001). This in spite of the fact that the actual relationship between sett densities and badger densities across the observed density gradient in Europe is poorly documented in the literature (Macdonald et al. 1996). A number of reasons exist why this relationship might be neither consistent nor straightforward: First, the badger is tremendously flexible with respect to the size of the social group. It has been suggested that suitable sett sites may be a limited resource in some areas (Roper 1993). Once this is true any increase in population size would have to be in the form of an increase in social group size rather than in the number of social groups (as observed by Rogers et al. 1997, 2000). Second, sett occupancy rates may be below one. At low densities, where nearest neighbour distances between setts are large, successful dispersal of individuals between social groups is likely to become constrained by the spatial configuration of setts. This is a common phenomenon in spatially structured populations. Fewer social groups – and thus breeding opportunities – in the local neighbourhood may discourage dispersal (as suggested by Revilla and Palomares 2002) and thus increase social group size in occupied setts. Uncertainty regarding occupancy rates is less of a problem when dealing with empirical data on sett densities, as it is fairly easy to determine whether a given sett is in use or not. It becomes an issue however, if we attempt to extrapolate from predicted sett densities to population densities. For such an exercise to be successful more attention to the spatial aspect of sett distribution and the possible consequences for the occupancy of breeding setts is needed.

Synthesis and recommendations

Decisions in management and conservation very often have to be made based on a less than optimal information about the system in question. Limited available resources or urgency in the need for answers, means that improving our abilities to produce reasonable answers based on current knowledge of the system is high priority. Predictive models are important tools in achieving this. The approach used here is an attempt to produce, in a step-wise manner, useful predictions about habitat suitability and densities of badger setts in a region in Denmark. In doing so we have aimed at optimising the utility of the model in management, by carefully selecting parameters and improving the incorporation of terrain characteristics in the habitat model. We have attempted to solve some of the common problems related to the inference of abundance estimates from presence/absence data. We are confident that this approach will be useful in management of the badger in other regions of Northern Europe. The concrete models presented here have, however, been developed and validated for local conditions in Denmark. Further validation is needed to determine the utility of the current model parameterisation in other regions. The relevance of each predictive variable for the local environment should be considered (e.g. a variable related to forest coverage is meaningless in continuous forest, geological conditions will increase in importance if large regions are entirely unsuitable habitat e.g. exposed bedrock). Denmark is dominated by a mosaic of agriculture, build-up areas and woodlots. Similar areas are found for instance in parts of the UK, the Benelux region and Northern Germany. In these regions the current model parameterisation may very well have a predictive ability similar to that found for the study areas in Denmark.

Acknowledgements

This work was supported by a Marie Curie Individual Fellowship provided by the European Commission (EVK2-CT-2000-57004) to the first author.

Dr T. Wiegand, Dr K. Frank and one anonymous reviewer provided helpful advice and valuable comments on an earlier version of the manuscript. We thank P.N. Andersen for GIS support. H. Baekgaard, S.A. Madsen, J.L. Jeppesen and M. Hammershøj provided important assistance collecting the field data.

References

Carroll S.S. and Pearson D.L. 2000. Detecting and modeling spatial and temporal dependence in conservation biology. Conserv. Biol. 14: 1893–1897.

Clements E.D., Neal E.G. and Yalden D.W. 1988. The national badger sett survey. Mammal Rev. 18: 1–9.

Cresswell P., Harris S., Bunce R.G.H. and Jefferies D. 1989. The badger, *Meles meles*, in Britain: present status and future population changes. Biol. J. Linn. Soc. 38: 91–101.

Danks F.S. and Klein D.R. 2002. Using GIS to predict potential wildlife habitat: a case study of muskoxen in northern Canada. Int. J. Remote Sens. 23: 4611–4632.

Eide N.E., Nellemann C. and Prestrud P. 2001. Terrain structure and selection of denning areas by arctic foxes on Svalbard. Polar Biol. 24: 132–138.

Feore S. and Montgomery W.I. 1999. Habitat effects on the spatial ecology of the European badger (*Meles meles*). J. Zool. Lond. 247: 537–549.

Fielding A.H. and Bell J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24: 38–49.

Good T.C., Hindenlang K., Imfeld S. and Nievergelt B. 2001. A habitat analysis of badger (*Meles meles* L.) setts in a semi-natural forest. Mammal. Biol. 66: 204–214.

Hammond R.F., McGrath G. and Martin S.W. 2001. Irish soil and land-use classifications as predictors of numbers of badgers and badger setts. Prev. Vet. Med. 51: 137–148.

Hosmer D.W. and Lemeshow S. 2000. Applied Logistic Regression. Wiley, New York.

Johnson D.D.P., Macdonald D.W., Newman C. and Morecroft M.D. 2001. Group size versus territory size in group living badgers: a large-sample field test of the Resource Dispersion Hypothesis. Oikos 95: 265–274.

Johnson D.D.P., Jetz W. and Macdonald D.W. 2002. Environmental correlates of badger social spacing across Europe. J. Biogeogr. 29: 411–425.

Kruuk H. 1978. Spatial organisation and territorial behaviour of the European badger *Meles meles*. J. Zool. (Lond.) 184: 1–19.

Legendre P., Dale M.R.T., Fortin M., Gurevitch J., Hohn M. and Myers D. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. Ecography 25: 601–615.

Lennon J.J. 2000. Red-shifts and red herrings in geographical ecology. Ecography 23: 101–113.

Macdonald D.W., Mitchelmore F. and Bacon P.J. 1996. Predicting badger sett numbers: evaluating methods in East Sussex. J. Biogeogr. 23: 649–655.

Manly B.F.J., McDonald L.L., Thomas D.L., McDonald T.L. and Erickson W.P. 2002. Resource Selection by Animals: Statistical Design and Analysis for Field Studies, 2nd ed. Kluwer Academic Publishers, Dordrecht.

Mosher J.A., Titus K. and Fuller M.R. 1986. Developing a practical model to predict nesting habitat of woodland hawks. In: Verner J., Morrison M.L. and Ralph C.J. (eds), Wildlife 2000: Modelling Habitat Relationships of Terrestrial Vertebrates. The University of Wisconsin Press, Madison, Wisconsin, pp. 31–35.

Neal E. and Cheeseman C. 1996. Badgers. T & A.D. Poyser Ltd., London.

Nellemann C. and Fry G. 1995. Quantitative analysis of terrain ruggedness in reindeer winter grounds. Arctic 48: 172–176.

Nellemann C. and Reynolds P. 1997. Predicting late winter distribution of muskoxen using an index of terrain ruggedness. Arctic Alpine Res. 29: 334–338.

Nielsen K., Stjernholm M., Olsen B.Ø., Müller-Wohlfeil D., Madsen I., Kjeldgaard A., Groom G., Hansen H.S., Rolev A.M., Hermansen B., Skov-Petersen H., Johannsen V.K., Hvidberg M., Jensen J.E., Bacher V. and Larsen H. 2000. Areal Informations Systemet – AIS. Technical report from the Danish Ministry of Environment. http://www.ais.dk. (in Danish)..

Nolan A. and Wilesmith J.W. 1994. Tuberculosis in badgers (*Meles meles*). Vet. Microbiol. 40: 179–191.

Ostler J.R. and Roper T.J. 1998. Changes in size, status and distribution of badger *Meles meles* L. setts during a 20-year period. Zeitschrift für Saugetierkunde/ Int. J. Mammal. Biol. 63: 200–209.

Pearce J. and Ferrier S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecol. Model. 133: 225–245.

Pearce J. and Ferrier S. 2001. The practical value of modelling relative abundance of species for regional conservation planning: a case study. Biol. Conserv. 98: 33–43.

Revilla E., Delibes D., Traviani A. and Palomares F. 1999. Physical and population parameters of Eurasian badgers, *Meles meles*, from Mediteranean Spain. Zeitschrift für Säugetierkunde/Int. J. Mammal. Biol. 64: 269–276.

Revilla E., Palomares F. and Fernández N. 2001. Characteristics, location and selection of diurnal resting dens by Eurasian badgers (*Meles meles*) in a low-density area. J. Zool. (London) 255: 291–299.

Revilla E. and Palomares F. 2002. Spatial organisation, group living and ecological correlates in low-density populations of Eurasian badgers, *Meles meles*. J. Anim. Ecol. 71: 497–512.

Rodríguez A., Martin R. and Delibes M. 1996. Space use and activity in a Mediterranean population of badgers *Meles meles*. Acta Theriol. 41: 59–72.

Rogers L.M., Cheeseman C., Mallinson P.J. and Clifton-Hadley R.S. 1997. The demography of a high-density badger (*Meles meles*) population in the west of England. J. Zool. 242: 705–728.

Rogers L.M., Delaney R.J., Hounsome T.D. and Cheeseman C. 2000. Changes in badger, *Meles meles*, social organisation in response to increasing population density at Woodchester Park, south-west England. In: Griffiths H.I. (ed.), Mustelids in a Modern World. Backhuys Publishers, Leiden, The Netherlands, pp. 267–279.

Roper T.J. 1993. Badger setts as a limiting resource. In: Hayden T.J. (ed.), The Badger. Royal Irish Academy, Dublin, pp. 26–34.

Schadt S., Revilla E., Wiegand T., Knauer F., Kaczensky P., Breitenmoser U., Bufka L., Červeny J., Koubek P., Huber T., Staniša C. and Trepl L. 2002. Assessing the suitability of central European landscapes for the reintroduction of Eurasian lynx. J. Appl. Ecol. 39: 189–203.

Thornton P.S. 1988. Density and distribution of Badgers in south-west England – a predictive model. Mammal Rev. 18: 11–23.

Van der Zee F.F., Wiertz J., Ter Braak C.J.F. and Apeldoorn van R.C. 1992. Landscape change as a possible cause of the badger *Meles meles* L. decline in the Netherlands. Biol. Conserv. 61: 17–22.

Virgos E. and Casanovas J.G. 1999. Environmental constraints at the edge of a species distribution, the Eurasian badger (*Meles meles* L.): a biogeographic approach. J. Biogeogr. 26: 559–564.

Virgos E. 2001. Role of isolation and habitat quality in shaping species abundance: a test with badgers (*Meles meles*) in a gradient of forest fragmentation. J. Biogeogr. 28: 381–389.

Wright A., Fielding A.H. and Wheater C.P. 2000. Predicting the distribution of European badger (*Meles meles*) setts over an urbanized landscape: a GIS approach. Photogramm. Eng. Remote Sens. 66: 423–428.