

Pale Rock Sparrow *Carpospiza brachydactyla* in the Mount Lebanon range: modelling breeding habitat

CHRISTOPHER G. KNIGHT^{1*} & COLIN M. BEALE^{2†}

¹Department of Biology, Imperial College, Silwood Park, Ascot, Berks. SL5 7PY UK

²A Rocha Lebanon, The Post Office, Qabb Elias, The Bekaa, Lebanon

The Pale Rock Sparrow *Carpospiza brachydactyla* is a poorly known species with very little documentation of its breeding habitat preferences. Using detailed surveys and habitat modelling for a population in southern Lebanon's Bekaa valley, we have identified aspects of the habitat associated both temporally and spatially with breeding. Static spatial habitat modelling using three fundamentally different statistical techniques (multiple linear regression, regression trees and artificial neural networks) agreed on positive associations of breeding density with 'pebbliness' of ground cover and the quantity of available habitat, and a negative association with trees. Temporal associations were found between breeding and a rise in temperature and peaks in grasshopper and beetle abundance, the two main prey that we observed the birds taking. These findings are discussed in the context of Pale Rock Sparrow conservation and implications for species-directed habitat assessment more generally.

The Pale Rock Sparrow *Carpospiza brachydactyla* is a species of conservation concern, enigmatic in its taxonomy and little known in its ecology (Evans 1994). Endemic to the Middle East, it is unclear whether phylogenetically they are sparrows (Passeridae), finches (Fringillidae), buntings (Emberizidae) or a sister group to one or more of those taxa (Cramp & Perrins 1994). The only published study on their ecology focuses on nestling development (Adamyant 1965). Cramp and Perrins (1994) could cite only one other general work (Dement'ev & Gladkov 1954) describing their breeding habitat. Inconspicuous plumage and erratic fluctuations in breeding populations in the few well-visited breeding areas (Martins 1989, Shirihai 1996, Griffin *et al.* 2001) have resulted in their breeding preferences remaining largely unknown. Nevertheless, intensive survey work is now underway in their main range in Turkey (Kirwan *et al.* 2003), which may clarify the distribution of breeding populations. This study aims to clarify the temporal and spatial factors that may influence breeding in a particular established population.

A previously undocumented Pale Rock Sparrow breeding population was identified in 1997 on the

southwestern side of southern Lebanon's Bekaa valley, close to the Aammiq wetland (Walley & Walley 1998). This population is apparently stable, with good numbers present in the area each breeding season since its discovery, up to and including 2004. This Pale Rock Sparrow breeding population is split between two, apparently very different, habitats. Near the base of the valley at around 1000 m asl, they occupy a strip of goat-degraded land, sandwiched between maquis on the slopes above and the villages, vineyards, orchards and road which occupy the valley floor and encroach up the lower slopes. Separated from this area by around 800 m of altitude are the tops of the Barouk ridge of the Mount Lebanon range where the Pale Rock Sparrows breed among sparse alpine vegetation. This study covers both these areas, using the differences between the two to elucidate the consistent features chosen by Pale Rock Sparrows (see Knight & Beale 2004 for further descriptions and a map of the study area).

In this study we set out to determine reliable correlates of both spatial and temporal variation in Pale Rock Sparrow breeding. We build static spatial models of Pale Rock Sparrow density in relation to habitat variables using three fundamentally different techniques (multiple regression, regression trees and artificial neural networks) to look for features of the habitat consistently associated with breeding, irrespective of the particular modelling technique. We also use a difference in the timing of

*Corresponding author. Current address: Department of Plant Sciences, South Parks Road, Oxford OX1 3RB, UK
Email: christopher.knight@plants.ox.ac.uk

†Current address: Division of Environmental & Evolutionary Biology, Graham Kerr Building, Glasgow University, Glasgow G12 8QQ, UK

breeding between the upper and lower sections of the population to reveal associations with temporally varying aspects of habitat, such as availability of prey species. This approach gives an overall view of the habitat features chosen for breeding within the population that is relevant, not only to our understanding of this little-known species, but to practical conservation and habitat management measures in an area undergoing rapid land-use changes.

STUDY AREA AND METHODS

The study area comprised two discrete sites at different altitudes on the east slope of the Barouk range of the Lebanon Mountains, overlooking Lebanon's Bekaa valley (33°42'N, 35°45'E). The mountains in this region rise from around 880 m in the valley to peaks of *c.* 1940 m. The vegetation is broadly divided into three altitudinal bands: (1) a lower band, bounded by cultivation, habitation or road at the valley floor, up to 950–1050 m, highly degraded by intensive goat grazing and characterized by small spiny bushes (principally *Sarcopoterium spinosum* and *Calicotome villosa*), grasses, rocky ground and bare earth; (2) a mid band of partially goat-grazed Mediterranean maquis becoming steppe woodland towards the top at around 1700 m; (3) low, alpine vegetation above ~1700 m characterized by small spiny bushes ('tragacanth'). Following observations in 2000 we considered the first and third of these to be potential Pale Rock Sparrow habitat. A stretch of these two bands of potential habitat 9 km long was selected as the study area, between the villages of Kefraya in the south and Aammiq in the north (Knight & Beale 2004).

We defined 40 straight-line transects running perpendicular to the ridge: 20 on the lower slopes and 20 on the mountain top. Each transect was at least 400 m from the next.

Bird density estimates

All Pale Rock Sparrows seen during a transect were recorded and whether they were more or less than 40 m from the transect line. Density (D , km⁻²) was derived assuming complete detection on the transect line and a linear reduction in detectability with distance from the line (Bibby *et al.* 1992). This requires a detectability constant, k , that was calculated for each visit to each transect. The average value of k was then used for all estimates of D . The average value of D for each transect was used in subsequent analyses.

A preliminary dataset obtained from 37 transect walks (different transects from those in the main dataset) in the (2000) breeding season (C.M.B. & A. Sprenger unpubl. data) was used to determine the optimum layout of transects and the 40-m threshold mentioned above. This dataset used an arbitrary 10-m threshold, but analysis with the model described above indicated that the optimum threshold (where 50% of records fall within and 50% beyond) was 40 m. It also indicated that singing males could be detected up to approximately 150 m. Applying the same methodology to the main dataset from 2001 estimated the optimum threshold and maximum detectability at 40.2 m and 161 m, respectively. This helps justify the implicit assumption of the method that there is no systematic variation in detectability among observers, sites or occasion of data collection.

We used the maximum detectability estimate to define optimum transects for the site (well separated in terms of maximum detectability of the birds, whilst still giving thorough coverage). This comprised 20 straight-line transects perpendicular to the ridge at each site (upper and lower), all separated by 400 m. All transects were 500 m long, except for three in the lower site where potential habitat was only 400 m wide. All transects were walked twice, the lower site between 2 and 23 May 2001 and the upper site between 29 May and 19 June 2001. Each visit occurred between 07:00 and 14:00 h. The second visit to each transect was *c.* 10 days after the first. Between visits the direction of travel and order of transects in the day were reversed.

Habitat assessment

Fourteen habitat variables were recorded (Table 1). At two points approximately one-third of the way in from either end of each main transect, five radiating 50-m transects were walked while recording habitat variables. Point assessments of cover were made every 10 m along these short transects, using seven categories (corresponding to *Bar*, *Bus*, *Grs*, *Hrb*, *Peb*, *Rok*, *Tre*). Some points included more than one category (e.g. tree and grass). The per cent cover was arcsine transformed to give the variables used in the analyses. We also recorded *With*, the width of potential habitat at the transect, *Hgt*, the height of the highest object encountered by the 50-m transects, *Slp*, the gradient of the ground within the central third of the main transect continuing for the largest distance, and *Hmn*, the distance of the nearest human habitation to any part of the main transect. *Dst*

Table 1. Variables measured.

Variable name	Meaning
<i>Alt</i>	Altitude
<i>Bar</i>	proportion of bare earth
<i>Bus</i>	proportion of spiny bushes
<i>Dst</i>	distance along ridge
<i>Grs</i>	proportion of grass
<i>Hgt</i>	maximum height of object
<i>Hmn</i>	distance from human habitation
<i>Hrb</i>	proportion of herbs
<i>Peb</i>	proportion of pebbles (< 10 cm diameter)
<i>Rok</i>	proportion of rocks (> 10 cm diameter)
<i>Slp</i>	gradient of slope
<i>Tmp</i>	temperature at the time of the density transects
<i>Tre</i>	proportion of trees
<i>Wth</i>	width of potential habitat strip

and *Alt* were included to identify the locations of the transects, being, respectively, the number of the transect south to north along the mountain ridge and the approximate altitude in metres. Finally, *Tmp* was the average temperature (see 'Temperature data') for the time nearest that when the transect was walked. Explanatory variables were transformed to a scale of 0–1 for the artificial neural networks (ANN) analyses (Venables & Ripley 2002).

Prey assessment

Assessments of potential invertebrate prey available for Pale Rock Sparrows were made at two sites near the centre of the study area, one in the upper area and one in the lower area. Three techniques were used: (a) pitfall traps, (b) visual surveys and (c) sweep-netting. (a) Five pitfall traps (7 cm diameter, 8.5 cm depth with 100 mL ethanol) were set under stones in each of the upper and lower sites. Contents were removed and classified (to order for insects) and the ethanol replenished weekly throughout the breeding season (10 weeks from 23 April). (b) Weekly (10 weeks from 23 April) visual surveys in each of the upper and lower study sites used a 500-m transect loop walked for 1 h recording and classifying (as above) all terrestrial invertebrates observed. The quantity of observations was kept manageable by a priori excluding flying invertebrates and ants. (c) Each visual survey was broken three times to make ten net sweeps each of whatever vegetation was in the vicinity. All invertebrates captured were classified as above and released.

Temperature data

Temperature readings came from two weather stations in the Bekaa valley, usually at three hourly intervals, throughout the study period. The stations (at 34°03'N, 36°72'E and 33°82'N, 35°85'E) are at 920 m and 1333 m above sea-level, respectively. These figures were used linearly to extrapolate the likely temperatures at the study sites throughout the study period.

Statistical analysis and software

JMP 3.0 (SAS Institute 1995) and Excel 98 were used to explore the data and perform simple tests. R v1.3.2 (Ihaka & Gentleman 1996) was used to implement multiple regression models, regression trees and neural network analyses using the `lm()`, `tree()` and `nnet()` objects, respectively (Chambers & Hastie 1992, Venables & Ripley 2002). *G*-tests were used on 2 × 2 contingency tables unless the expected number in any category was < 5, in which case Fisher exact test was used. The individual *P* values resulting were adjusted for multiple testing by controlling for false discovery rate (Benjamini & Hochberg 1995). Programmes for multiple model fitting and simplification were written in S and implemented in R.

Multiple linear regression (MLR)

In order to keep the number of effects fitted in any one model to an appropriate maximum for the number of data points (13 given the 40 data points), a four-stage simplification strategy was used:

(1) Variables with negligible explanatory power were removed by fitting all 364 possible combinations of three variables with all interactions and first-order curvature effects, and then simplifying to a minimal adequate model using the Akaike information criterion (AIC) (Akaike 1974). The main effects dropped most frequently were sequentially dropped from consideration in subsequent stages of analysis unless over-represented in two- or three-way interactions or curvature effects.

(2) All possible models containing all the remaining main effects, a single three-way interaction (and all the necessary component two-way interactions) and one curvature effect were fitted. In each case a minimal adequate model was derived using AIC. The frequency with which effects were dropped formed a strongly bimodal distribution (at one end, effects dropped from over 55% of models, peaking at around 85% of models; at the other, effects dropped from fewer than half the models, peaking at only a 5% drop rate); only the peak of rarely dropped effects was used in the next stage.

(3) As each three-way interaction had been tested less frequently than other effects in the preceding stages, all pairs of three-way interactions (and the necessary component two-way interactions and main effects) were fitted. These models were simplified as before using AIC. The three-way interaction dropped the greatest proportion of the time was repeatedly removed from the analysis until all were retained in 50% or more of the models in which they were used. These were then compared directly.

(4) The final model was fitted using remaining effects and simplified to the minimal adequate model by sequentially removing the least significant effect not required by a higher order interaction and not itself significant at the 5% level.

Regression tree analysis (RT)

Constructing a regression tree is a non-parametric, maximum-likelihood-based analysis, and is highly appropriate for habitat-modelling questions (e.g. Rejwan *et al.* 1999, Andersen *et al.* 2000). It successively subdivides the data into high and low values of individual variables, starting with the one explaining the greatest proportion of the deviance. A regression tree was fitted to the data and then pruned to a size that minimized the deviance in a 'leave-one-out' cross validation (Venables & Ripley 2002).

Artificial neural networks (ANN)

ANN are highly flexible non-linear modelling tools whose value in ecology has been recognized only relatively recently (e.g. Lek *et al.* 1996, Lek & Guegan 1999). Although very effective as a 'black box' method for prediction of new values on the basis of an existing

data set, their drawback is a lack of agreement on the best methods of model selection and interpretation (Zhang 2000). A reproducible model selection algorithm was used as advocated by Anders and Korn (1999) starting with a 'bottom-up' choice of the number of hidden neurones in the model (i.e. starting with one and only increasing if deemed necessary), followed by a 'top-down' removal of unnecessary explanatory variables. The criterion chosen was mean square prediction error on 'leave-one-out' cross validation (CV-error, a cumbersome, but simple and widely applicable technique), i.e. changes were made in the model structure only if they decreased CV-error averaged across replicates (Anders & Korn 1999). In the case of more than one explanatory variable reducing CV-error when removed, only that which reduced it most was removed before re-evaluating CV-error. In each of 20 replicates, models were iterated using random starting weights in the range -0.3 to $+0.3$ and a decay constant of 5×10^{-4} (Venables & Ripley 2002). Iterations continued until convergence to six decimal places. A measure of the relative contributions of the different variables in the chosen model was derived by weight partitioning (Garson 1991, Goh 1995).

RESULTS

Pale Rock Sparrow density in the study area varied from 0 to about 100 birds/km² over the 40 transects with a distribution not differing significantly from normal (Shapiro–Wilk test for normality (Shapiro & Wilk 1965) $n = 40$, $P = 0.22$). Several of the measured habitat variables were correlated with each other and with Pale Rock Sparrow density (Table 2).

Table 2. Pearson product-moment correlation coefficients between all variables and Pale Rock Sparrow density.

	<i>Tre</i>	<i>Rok</i>	<i>Bar</i>	<i>Bus</i>	<i>Hrb</i>	<i>Grs</i>	<i>Peb</i>	<i>Slp</i>	<i>Hgt</i>	<i>Hmn</i>	<i>Wth</i>	<i>Tmp</i>	<i>Alt</i>	<i>Dst</i>
Density	-0.51	-0.33	0.08	0.02	0.08	-0.02	0.42	-0.22	-0.39	0.03	0.44	-0.28	0.09	-0.07
<i>Tre</i>		0.11	0.07	-0.12	0.28	-0.18	-0.14	0.37	0.57	0.29	-0.09	0.47	0.18	0.34
<i>Rok</i>			-0.04	-0.54	0.23	-0.26	-0.16	0.50	0.10	0.31	-0.18	0.49	0.38	-0.37
<i>Bar</i>				-0.19	0.11	0.24	-0.32	0.20	0.18	0.00	-0.06	0.28	-0.13	-0.07
<i>Bus</i>					-0.61	0.19	-0.25	-0.47	0.20	-0.59	-0.07	-0.58	-0.67	0.34
<i>Hrb</i>						-0.47	0.25	0.37	-0.17	0.78	0.41	0.48	0.77	0.07
<i>Grs</i>							-0.58	-0.35	0.13	-0.61	-0.34	-0.21	-0.66	-0.19
<i>Peb</i>								0.06	-0.44	0.45	0.45	-0.03	0.52	0.02
<i>Slp</i>									0.29	0.50	0.02	0.52	0.50	0.06
<i>Hgt</i>										-0.12	-0.16	0.19	-0.24	0.22
<i>Hmn</i>											0.58	0.52	0.93	0.05
<i>Wth</i>												-0.02	0.52	0.10
<i>Tmp</i>													0.45	-0.26
<i>Alt</i>														0.00

Table 3. Multiple regression (MLR) minimal adequate model.

Effect	Coefficient (± se)	Significance
<i>Slp</i>	24 (12)	*
<i>Peb</i>	870 (340)	*
(<i>Tre</i>) ²	-310 (94)	**
<i>With</i>	0.83 (0.36)	*
<i>Slp</i> * <i>Peb</i>	-47 (19)	*
<i>Slp</i> * <i>With</i>	-0.046 (0.020)	*
<i>Peb</i> * <i>With</i>	-1.4 (0.57)	*
<i>Slp</i> * <i>Peb</i> * <i>With</i>	0.085 (0.031)	**

P* < 0.05, *P* < 0.01.

Spatial analysis

MLR resulted in a minimal adequate model (*sensu* Crawley 2002) containing four variables (*Peb*, *Tre*, *Slp* and *With*) and eight effects (*Peb*, *Tre*², *With*, *Slp*, *With***Slp*, *With***Peb*, *Slp***Peb* and *With***Slp***Peb*). It explained 66% of the observed variance (Table 3, *R*² = 0.66, *P* < 0.001). In the creation of that model from a maximal model, variables were removed in the following order: *Bar*, *Hrb*, *Hmn*, *Bus*, *Rok*, *Dst*, *Alt*, *Hgt*, *Grs* and *Tmp*.

An RT model (Fig. 1) explained 81% of the deviance. The tree as initially derived contained five variables (*Bar*, *Hgt*, *Peb*, *Slp* and *With*) and required no pruning according to ‘leave-one-out’ cross validation.

To fit an ANN model we used a simple method based on cross-validation error. This was minimized

for a model using only the variables *Tre*, *Peb*, *Slp* and *With* using one ‘hidden layer neurone’ (Fig. 2a). The ANN model explains 76% (± 0.4% se, *n* = 100) of the variance in the data. Weight partitioning (Garson 1991, Goh 1995; Fig. 2b) suggests that the most important variable is *Peb*.

Temporal analysis

Temperature data from two weather stations at different altitudes in the Bekaa Valley mirrored each other closely over the study period (*R*² = 0.88, *n* = 394) with a consistent temperature difference equivalent to a fall of 6.2 °C per 1000 m increase in altitude. We used this to estimate (by linear interpolation and extrapolation) weekly temperature at the study sites (Fig. 3a). The temperature curves are S shaped, differing between sites by 5.2 °C. The temperature at the upper site when the birds arrive is very similar to the temperature at the lower site when they arrive (18 °C) although the events are separated by nearly 1 month.

Data on potential prey species were analysed for two periods (early = 4–25 May, late = 7–29 June) corresponding to the Pale Rock Sparrow breeding periods in the lower and upper areas, respectively, at both the lower and upper study sites. Data were collected using three different techniques. Although these techniques focused on different groups (more than 75% of records were in the visual survey for

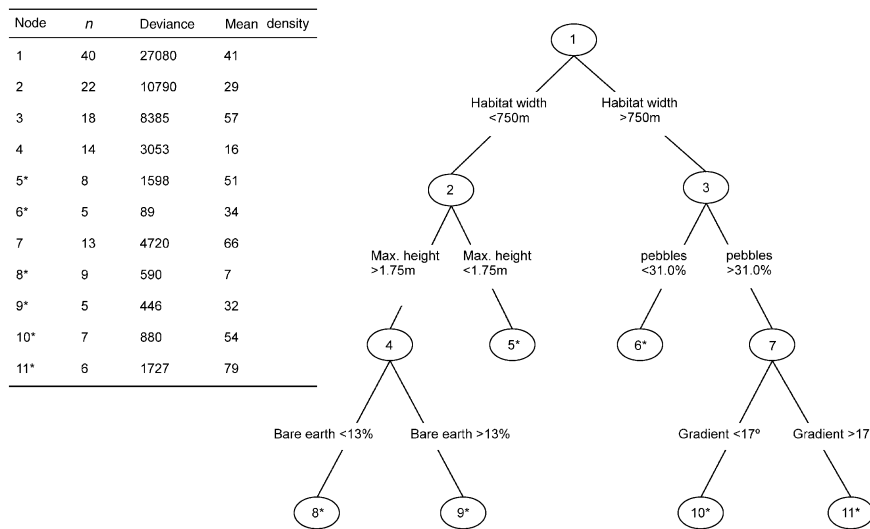


Figure 1. Regression tree model. The model subdivides the data according to the criteria marked on the branches starting with all the data at node 1. Each node has the number of data points (*n*), deviance and mean density of Pale Rock Sparrows shown in the adjacent list. Asterisks next to node numbers indicate terminal nodes.

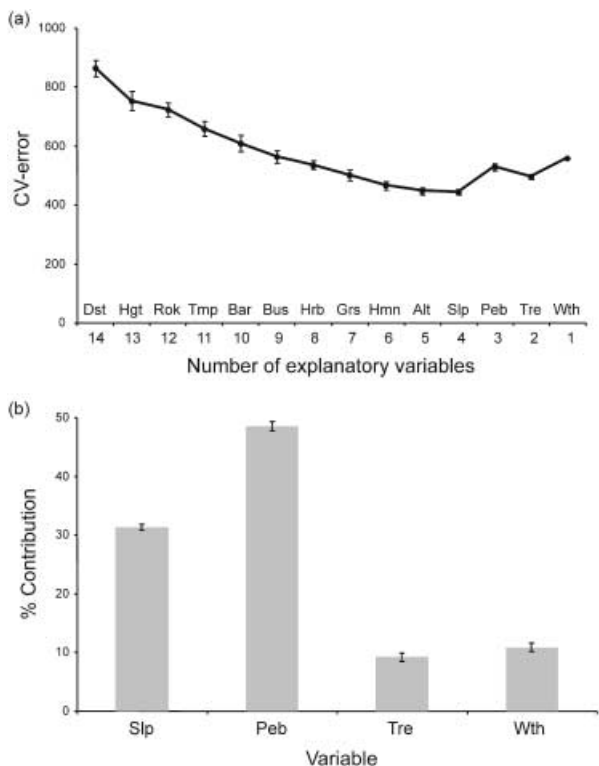
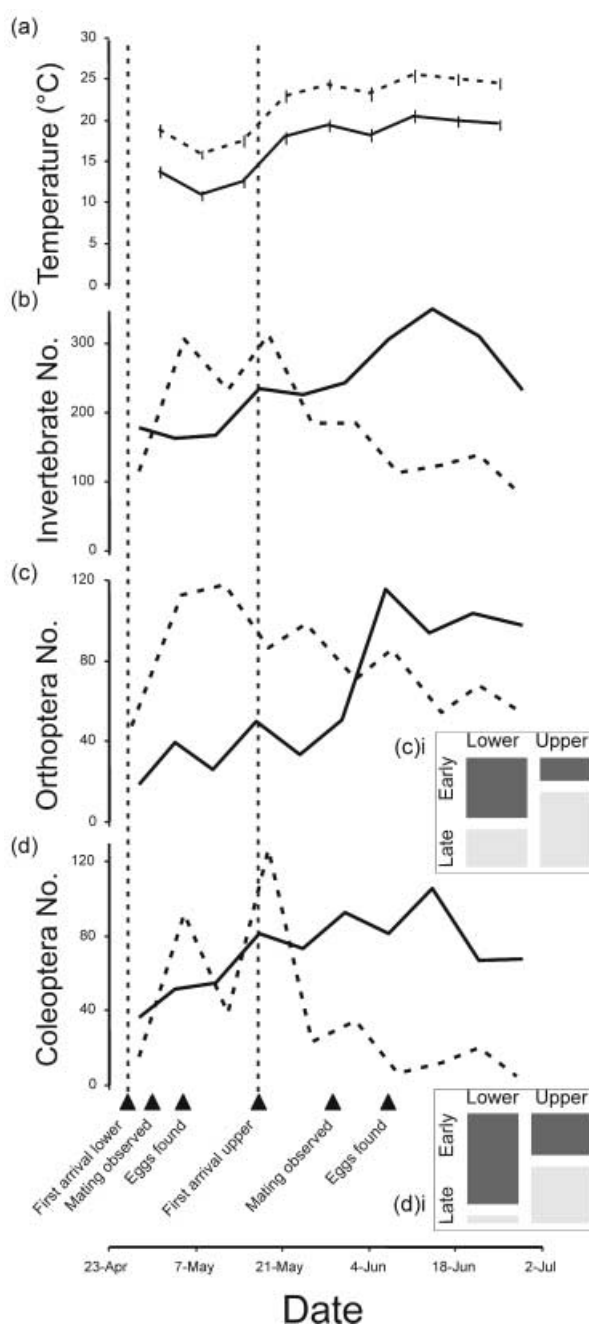


Figure 2. Artificial neural network model. (a) Model selection by ‘leave-one-out’ cross validation. Each point is the mean (\pm se, $n = 20$) cross-validation error for a given size of model. The variable whose removal results in the lowest CV-error (and is therefore removed at the next point) is indicated above the explanatory axis. (b) Per cent contribution of variables in the ANN model which minimizes CV-error. Points are mean (\pm se, $n = 200$) contribution as defined by partitioning of network weights (Garson 1991, Goh 1995).

Orthoptera, in the sweep net survey for Hemiptera and in the pitfall survey for Hymenoptera), they can be considered neither as independent nor as absolute. A contingency table test (corrected for multiple testing) was used for each technique for each identified taxon of potential prey species to determine whether records were independently distributed between the periods in the upper and lower areas.

Of the 16 invertebrate taxa identified in the samples (nine insect orders, plus Acari, Araneae, Chilopoda, Diplopoda, Isopoda, Opiliones and Scorpiones), five (Coleoptera, Diptera, Hemiptera, Hymenoptera

Figure 3. Key correlates of timing of breeding between the two altitudes. Dashed lines correspond to the lower site, solid lines to the upper site. The dotted vertical lines indicate the date of arrival of Pale Rock Sparrows at each site; other breeding observations are shown by the black triangles at the bottom.



(a) Average temperature at the study sites throughout the breeding season as extrapolated from weather stations in the Bekaa valley. Weekly means with se bars. (b) Total numbers of invertebrates recorded by all methods. (c) Number of Orthoptera recorded summed over all methods. Inset: mosaic plot of Orthoptera counts in the visual survey at the upper and lower sites in the early and late Pale Rock Sparrow breeding periods as used in the analysis. (d) Number of Coleoptera recorded summed over all methods. The dip on 12 May corresponds to greatly reduced numbers of small rove beetles (Staphylinidae) present on the preceding and succeeding weeks. Inset: mosaic plot of Coleoptera counts in the pitfall traps.

and Orthoptera) showed counts which were non-independently distributed as recorded by at least one technique (G -tests, adjusted $P < 0.01$ in each case: Coleoptera: sweep, $n = 353$; pitfall, $n = 466$; Diptera: sweep, $n = 208$; Hemiptera: pitfall, $n = 39$; Hymenoptera: pitfall, $n = 243$, Orthoptera: visual, $n = 1120$). In each case the number of invertebrates recorded was greater than expected when Pale Rock Sparrows were breeding at each site (see Fig. 3 insets for examples). In these cases, the timing of peak abundances was derived from the data of the method yielding the most significant result. In three cases (Diptera, Hemiptera and Hymenoptera) peak abundance was a long time from Pale Rock Sparrow breeding periods (e.g. in the lower area, peak abundances for these taxa were found on 4 May and 16 June, well before and after, respectively, Pale Rock Sparrow breeding). This left Coleoptera and Orthoptera for which abundance peaks were closely associated with Pale Rock Sparrow breeding events: Coleoptera abundance peaked about 1 week after Orthoptera in the period Pale Rock Sparrows were feeding young at both upper and lower sites. Abundance peaks for these taxa at the two sites were separated by 26–27 days, which compares with Pale Rock Sparrow breeding events separated by 21–33 days (Fig. 3). On nine occasions (in the 2000 and 2001 breeding seasons), Pale Rock Sparrows were seen pursuing and/or carrying identifiable food items. Eight of these were Orthoptera and one a large Coleopteran.

DISCUSSION

Habitat management decisions are frequently taken at a small scale, affecting particular populations of particular species. In this study we have investigated the associations of Pale Rock Sparrow breeding with habitat variables at such a within-population scale. To avoid the problems inherent at this small scale we utilized the subdivision of this population into sections living at different altitudes and employed multiple, fundamentally different, modelling techniques to clarify the consistently reliable features of the data.

There has been little previous study of Pale Rock Sparrow breeding habitat and no quantitative studies. However, past work suggested that for both adult and young Pale Rock Sparrows, insects are the main food in the breeding season (Adamyán 1965, Cramp & Perrins 1994). We found striking associations between timing of Pale Rock Sparrow breeding at the top and bottom of the mountain ridge with Coleopteran and Orthopteran abundance and average

temperatures (Fig. 3). The fact that the invertebrate groups showing this association are exclusively those observed being collected by Pale Rock Sparrows suggests that their link with Pale Rock Sparrow breeding is probably causal. The similarity in absolute numbers of Coleoptera and Orthoptera between the upper and lower sites (Fig. 3) associated with similar numbers of Pale Rock Sparrows is consistent with this hypothesis.

The close proximity, but differing altitudes, of the two sections of the population mean that the most significant climatic variable likely to differ between the sites is temperature. Pale Rock Sparrow arrival and breeding at the upper site is associated with a rise in temperature to a level similar to that when they arrived and bred at the lower site (Fig. 3a). Whilst temperature could affect Pale Rock Sparrows directly, the association with Coleoptera and Orthoptera may be relevant. Temperature is invariably crucial for development and hatching in these taxa (e.g. Fisher 1994, McAvoy & Kok 1999). Typical threshold development temperatures are around 9–13 °C, which compares to the rise from 12 to 18 °C that occurs at the upper site in the period between Pale Rock Sparrows first arriving in the area and appearing at the upper study site (Fig. 3a). This change undoubtedly also causes a substantial increase in insect development rate. It is thus plausible that the temperature difference between the upper and lower sites determines the variation in the timing of Coleopteran and Orthopteran hatching, which in turn determines the breeding season of Pale Rock Sparrows.

To assess the association of Pale Rock Sparrow breeding with small-scale habitat variation, we used a modelling approach. Ideally such models should be validated on independently collected datasets; however, this may not be possible if, as in this case, one is concerned with a particular population. The iterative and interactive nature of the model fitting process also limits the possibilities for assessing the model's robustness with randomization or simulated data. We took a different approach, using three fundamentally different modelling techniques (MLR, RT and ANN). We suggest that greater confidence can be placed in the reality of associations consistently identified when assessing the data in very different ways. Quite how different the three techniques are is illustrated by the way in which *Bar* (proportion of bare earth) was the first variable to be discarded from the MLR model, the fifth from the ANN model (Fig. 2) and yet was retained in the final RT model

(Fig. 1). Indeed the three-way interaction that the RT model suggests (between *Wth*, *Hgt* and *Bar*) was explicitly dropped when simplifying an MLR model containing it to a minimal adequate model. Despite these differences, however, several associations were consistently identified by each method, and we can have increased confidence that these are not artefacts but represent real, potentially causal, associations.

MLR and ANN identified exactly the same four variables as being important and all three techniques agreed that *Peb* (proportion of pebbles), *Wth* (quantity of habitat) and *Slp* (gradient of the slope) were key variables associated with Pale Rock Sparrow density. Density was clearly associated negatively with trees because both MLR and ANN identified the proportion of trees (*Tre*) and RT identified the maximum height of an object (*Hgt*, usually of a tree) as key variables. The existence of an interaction between *Peb*, *Slp* and *Wth* was agreed by MLR and RT, the only methods for which the interactions can be unravelled. MLR and ANN agree that *Peb* is the variable most influencing the predicted density of Pale Rock Sparrows, but they disagree on the relative contributions of the other variables, whilst RT offers no equivalent measures.

In the limited number of previous studies where several modelling techniques have been used on the same data it has invariably been for the purposes of technique evaluation. Congruence of the results between the techniques used here has been mixed (e.g. Brosse *et al.* 1999, Manel *et al.* 1999, Brosse & Lek 2000, Guisan & Zimmermann 2000, Clermont *et al.* 2001). However, a similar degree of care is often not used for model selection among the different methods. Attempting to use similar care in different methods here meant considering all possible two- and three-way interaction terms in MLR to compare with RT and ANN in which interactions are inherent. Of the 364 possible three-way interactions, both MLR and RT identify exactly the same interaction, between width of habitat (*Wth*), slope (*Slp*) and proportion of pebbles (*Peb*). ANN has no means for identifying the existence of that particular interaction in the model, but did identify all three constituent variables as significant. This is reassuring both that the information inherent in the data is being read correctly and that all methods are capable of uncovering it.

However, all methods were not equally easy to implement or interpret: MLR was a long multistep process (see Methods), and ANN required fitting, by

several hundred iterations each, over 11 000 separate networks. The interpretation and, in the case of ANN, the identification of interactions are also difficult in these methods. In contrast, RT could be implemented rapidly and clearly showed the nature of interactions. The principal factor weighing against use of RT on its own is that it highlighted two variables (*Bar* and *Hgt*) not corroborated by the other methods. Interpretation problems for ANN may be eased in future with the introduction of new techniques (Gevrey *et al.* 2003), but in the meantime, RT provided the most efficient approach to this dataset.

Although there is a case for causality between the populations of Coleoptera and Orthoptera and timing of Pale Rock Sparrow breeding, the causality for the variables highlighted by habitat modelling is unknown. The invertebrate results might suggest that *Peb*, the only significant ground cover variable, affects Pale Rock Sparrows via an effect on insects, but Dement'ev and Gladkov (1954) noted a more direct association: Pale Rock Sparrows 'swallow considerable quantities of small pebbles'. Similarly, although there is a good case for the existence of a three-way interaction between *Wth*, *Peb* and *Slp* (identified by MLR and RT and implicit in ANN), the biological causality and meaning of this interaction is unknown and cannot be hypothesized from these data because MLR and RT models disagree on the details of the interaction.

The only significant effects greatly influenced by humans are *Tre* (tree cover) and *Wth* (habitat width). Thus, despite no detectable effect of the proximity of human habitation, habitat narrowing (through encroachment of orchards and villages in the lower study site) is associated with reduced densities of Pale Rock Sparrows. Decreasing habitat dimensions reducing the value of the habitat is commonly highlighted by habitat modelling studies both practical and simulated, particularly in birds (e.g. Estades 2001, van den Berg *et al.* 2001, Herrando & Brotons 2002, Muchai *et al.* 2002, Stephens *et al.* 2004).

In conclusion, we found that Pale Rock Sparrows breeding in the southwestern Bekaa valley were clearly associated with specific aspects of habitat. These associations were both temporal (presence was associated with temperature and grasshopper and beetle abundance) and spatial (density was associated positively with pebbliness of ground cover and quantity of available habitat, and negatively with trees). We found that Pale Rock Sparrow densities were associated with the quantity of available

habitat, which suggests that encroachment from rapidly growing villages and cultivation is a threat to the population. Similarly, in the areas that do have some level of protection, such as the Arz el Shouf National Park and some of the land above the Aammiq marsh, there is an emphasis on reforestation. In the long term, this seems unlikely to benefit a species associated with a lack of trees.

This work was supported by London University Convocation Trust and the Ornithological Society of the Middle East Conservation Research Fund. It was carried out jointly with, and made possible by, the Lebanon project of A Rocha International. Specific thanks for enabling the fieldwork go to Andy and Laurel Sprenger, Chris Naylor, the Skaff family, the Arz el Shouf National Park and Feisal Halabi. For assistance in the UK we thank Donald Quicke, Mick Crawley, Ian Owens and Armand Leroi. Antje Dethof at the European Centre for Medium Range Weather Forecasts kindly supplied temperature data. We thank Dan Chamberlain and an anonymous reviewer for constructive comments on the manuscript.

REFERENCES

- Adamyan, M.S.** 1965. On the ecology of the Pale Rock Sparrow (*Carpospiza brachydactyla* BP) in the Armenian SSR. *Zool. Zhurnal* **44**: 569–577.
- Akaike, H.** 1974. A new look at statistical model identification. *IEEE Trans. Automatic Control* **19**: 716–722.
- Anders, U. & Korn, O.** 1999. Model selection in neural networks. *Neural Netw.* **12**: 309–323.
- Andersen, M.C., Watts, J.M., Freilich, J.E., Yool, S.R., Wakefield, G.I., McCauley, J.F. & Fahnestock, P.B.** 2000. Regression-tree modeling of desert tortoise habitat in the central Mojave Desert. *Ecol. Appl.* **10**: 890–900.
- Benjamini, Y. & Hochberg, Y.** 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Statist. Soc. Series B* **57**: 289.
- van den Berg, L.J., Bullock, J.M., Clarke, R.T., Langston, R.H. & Rose, R.J.** 2001. Territory selection by the Dartford warbler (*Sylvia undata*) in Dorset, England: the role of vegetation type, habitat fragmentation and population size. *Biol. Conserv.* **101**: 217–228.
- Bibby, C.J., Burgess, N.D. & Hill, D.A.** 1992. *Bird Census Techniques*, 1st edn. London: Academic Press.
- Brosse, S. & Lek, S.** 2000. Modelling Roach (*Rutilus rutilus*) microhabitat using linear and nonlinear techniques. *Freshw. Biol.* **44**: 441–452.
- Brosse, S., Lek, S. & Dauba, F.** 1999. Predicting fish distribution in a mesotrophic lake by hydroacoustic survey and artificial neural networks. *Limnol. Oceanogr.* **44**: 1293–1303.
- Chambers, J.M. & Hastie, T.J. (eds)** 1992. *Statistical Models in S*. New York: Chapman & Hall.
- Clermont, G., Angus, D.C., DiRusso, S.M., Griffin, M. & Linde-Zwirble, W.T.** 2001. Predicting hospital mortality for patients in the intensive care unit: a comparison of artificial neural networks with logistic regression models. *Crit. Care Med.* **29**: 291–296.
- Cramp, S. & Perrins, C.M. (eds)** 1994. *Birds of the Western Palearctic*, Vol. 8. Oxford: Oxford University Press.
- Crawley, M.J.** 2002. *Statistical Computing*. Chichester: John Wiley & Sons.
- Dement'ev, G.P. & Gladkov, N.A. (eds)** 1954. *Birds of the Soviet Union*, Vol. 5. Moscow: Gosudarstvenna Izdatel'stvo 'Sovetskaya Nauka'. Translated Gordon, E.D. 1970. Jerusalem: Keter Press.
- Estades, C.F.** 2001. The effect of breeding-habitat patch size on bird population density. *Landscape Ecol.* **16**: 161–173.
- Evans, M.I. (ed.)** 1994. *Important Bird Areas in the Middle East*. Cambridge: Birdlife International.
- Fisher, J.R.** 1994. Temperature effect on postdiapause development and survival of embryos of 3 species of *Melanoplus* (Orthoptera, Acrididae). *Ann. Entomol. Soc. Am.* **87**: 604–608.
- Garson, G.D.** 1991. Interpreting neural-network connection weights. *Artif. Intell. Expert* **6**: 47–51.
- Gevrey, M., Dimopoulos, I. & Lek, S.** 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol. Model.* **160**: 249–264.
- Goh, A.T.C.** 1995. Backpropagation neural networks for modeling complex-systems. *Artif. Intell. Eng.* **9**: 143–151.
- Griffin, P., Mayrose, A. & Tsurim, I.** 2001. Breeding Pale Rock Sparrow *Carpospiza brachydactyla* in the Samaritan Hills and Judean Desert, Israel. *Sandgrouse* **23**: 66–67.
- Guisan, A. & Zimmermann, N.E.** 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* **135**: 147–186.
- Herrando, S. & Brotons, L.** 2002. Forest bird diversity in Mediterranean areas affected by wildfires: a multiscale approach. *Ecography* **25**: 161–172.
- Ihaka, R. & Gentleman, R.** 1996. R.: a language for data analysis and graphics. *J. Computational Graph. Statistics* **5**: 299–314.
- Kirwan, G.M., Ozen, M., Kurt, B. & Martins, R.P.** 2003. Turkey Bird Report 1997–2001. *Sandgrouse* **25**: 8–31.
- Knight, C.G. & Beale, C.M.** 2004. Breeding habitat of the Pale Rock Sparrow *Carpospiza brachydactyla* in the Mount Lebanon range. *Sandgrouse* **26**: 127–131.
- Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. & Aulagnier, S.** 1996. Application of neural networks to modelling nonlinear relationships in ecology. *Ecol. Model.* **90**: 39–52.
- Lek, S. & Guegan, J.F.** 1999. Artificial neural networks as a tool in ecological modelling, an introduction. *Ecol. Model.* **120**: 65–73.
- Manel, S., Dias, J.M. & Ormerod, S.J.** 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecol. Model.* **120**: 337–347.
- Martins, R.P.** 1989. Turkey Bird Report 1982–86. *Sandgrouse* **11**: 1–41.
- McAvoy, T.J. & Kok, L.T.** 1999. Effects of temperature on eggs, fecundity, and adult longevity of *Hylobius transversovittatus* Goeze (Coleoptera: Curculionidae), a biological control agent of purple loosestrife. *Biol. Control.* **15**: 162–167.
- Muchai, M., Lens, L. & Bennun, L.** 2002. Habitat selection and conservation of Sharpe's Longclaw (*Macronyx sharpei*), a threatened Kenyan grassland endemic. *Biol. Conserv.* **105**: 271–277.
- Rejwan, C., Collins, N.C., Brunner, L.J., Shuter, B.J. & Ridgway, M.S.** 1999. Tree regression analysis on the nesting habitat of smallmouth bass. *Ecology* **80**: 341–348.

- SAS Institute.** 1995. *JMP Statistics and Graphics Guide*, Version 3. Cary, NC: SAS Institute Inc.
- Shapiro, S.S. & Wilk, M.B.** 1965. An analysis of variance test for normality (complete samples). *Biometrika* **52**: 591–611.
- Shirihai, H.** 1996. *The Birds of Israel*, 1st edn. San Diego: Academic Press.
- Stephens, S.E., Koons, D.N., Rotella, J.J. & Willey, D.W.** 2004. Effects of habitat fragmentation on avian nesting success: a review of the evidence at multiple spatial scales. *Biol. Conserv.* **115**: 101–110.
- Venables, W.N. & Ripley, B.D.** 2002. *Modern Applied Statistics with S*, 4th edn. New York: Springer-Verlag.
- Walley, C. & Walley, A.** 1998. Birding in the Bekaa valley, Lebanon. *Sandgrouse* **20**: 13–16.
- Zhang, G.Q.P.** 2000. Neural networks for classification: a survey. *IEEE Trans. Syst. Man Cybern. Part C – Appl. Rev.* **30**: 451–462.

Received 12 January 2004; revision accepted 18 November 2004.