Journal of Applied Ecology 1996, **33**, 339–347

An autologistic model for the spatial distribution of wildlife

N.H. AUGUSTIN*, M.A. MUGGLESTONE† and S.T. BUCKLAND*

**Research Unit for Wildlife Population Assessment, Mathematical Institute, University of St. Andrews, North Haugh, St. Andrews, Fife, KY16 9SS; and †Statistics Department, IACR-Rothamsted, Harpenden, Hertfordshire, AL5 2JQ, UK*

Summary

1. A new method for estimating the geographical distribution of plant and animal species from incomplete field survey data is developed.

2. Wildlife surveys are often conducted by dividing a study region into a regular grid and collecting data on abundance or on presence/absence from some or all of the squares in the grid. Generalized linear models (GLMs) can be used to model the spatial distribution of a species within such a grid by relating the response variable (abundance or presence/absence) to spatially referenced covariates.

3. Such models ignore or at best indirectly model dependence on unmeasured covariates, and the intrinsic spatial autocorrelation arising for example in gregarious populations.

4. We describe a procedure for use with presence/absence data in which spatial autocorrelation is modelled explicitly. We achieve this by extending a logistic model to include an extra covariate which is derived from the responses at neighbouring squares. The extended model is known as an autologistic model.

5. To allow fitting of the autologistic model when only a random sample of squares is surveyed, we use the Gibbs sampler to predict presence/absence at unsurveyed squares.

6. We compare the autologistic model with the ordinary logistic model using red deer census data. Both models are fitted to a subsample of 20% of the data and results are compared with the 'true' abundance and spatial distribution indicated by the full census. We conclude that the autologistic model is superior for estimating the spatial distribution of the deer, whereas the ordinary logistic model yields more precise estimates of the overall number of squares occupied by deer at the time of the survey.

Key-words: autologistic model, generalized linear model, Gibbs sampler, red deer, spatial autocorrelation, spatial distribution.

Journal of Applied Ecology (1996) **33**, 339–347

Introduction

Recent publications have proposed using generalized linear models (GLMs) for modelling wildlife distributions (Walker 1990; Osborne & Tigar 1992; Buckland & Elston 1993). Osborne & Tigar (1992) used a GLM with a logistic link function to predict probabilities of occurrence of bird species in Lesotho. Buckland & Elston (1993) used the same method to model deer census data, and investigated the case when only a sample of squares was surveyed. The logistic regression models of Osborne & Tigar (1992) and Buckland & Elston (1993) include habitat and other spatial covariates to allow for a heterogeneous environment, but spatial autocorrelation in the residuals is ignored.

A different approach to modelling wildlife distributions was described by Högmander & Møller (1995). In their model, strength of evidence of breeding was modelled as a function of effort by observers at each square to allow for variable effort in bird atlas surveys that rely on volunteer observers. They then adopted image analysis methods that use information from neighbouring squares to predict occurrence at any given square, and thus estimate the breeding range of a given species. Their model does not incorporate spatial covariates and assumes that habitat is homogeneous across squares. In this paper we develop the

GLM approach in a way that allows for spatial autocorrelation *and* a heterogeneous environment. We illustrate our methods using the red deer data set analysed by Buckland & Elston (1993). In common with those authors, we apply our spatial model to data on presence/absence of deer by 1 km square in the Grampian Region of Scotland.

Neighbouring squares tend to have similar conditions and if available covariates do not fully reflect the conditions (as perceived by the deer) then the residuals from a fitted model will exhibit spatial autocorrelation. Furthermore, quite apart from the effects of the environment, the probability of occurrence of deer in one square might not be independent of whether deer occur in a neighbouring square. This too will generate spatial autocorrelation that cannot be modelled satisfactorily by environmental covariates. By using models that allow for spatial autocorrelation, we might hope to require fewer covariates in an empirical model for distribution, and to obtain a better indication of which covariates influence the distribution.

Although in the case of the deer, all squares in the survey area were visited, wildlife surveys are frequently carried out on a sample of sites, selected according to some randomized scheme. We use the deer data, for which the true distribution is known, to develop and test a method suited to such surveys. We fit an autologistic model which allows for spatial autocorrelation in the presence/absence data. It would not ordinarily be possible to fit an autologistic model when only a sample of sites is surveyed, but we use the Gibbs sampler to estimate presence/absence at unsampled sites and so predict the full distribution of presence/absence.

Methods

THE DATA

The data are census counts of deer recorded by the Red Deer Commission in two deer management areas, the West and East Grampians. In common with Buckland & Elston (1993), we reduce these counts to presence/absence data, as modelling is simplified by modelling the count given presence as a separate exercise from modelling the presence/absence data. Covariate data are available as summaries by 1 km square for the Grampian Region of Scotland only, so the census counts were reduced to presence/absence by 1 km square (see Fig. 1), and squares outside the Region were excluded from the analysis.

In total, 1277 1 km squares were surveyed in the southern and south-western parts of Grampian Region; 190 of these squares were found to contain deer. The area is mostly open moorland, as fences exclude the deer from most of their favoured woodland habitat. The covariates are physical and habitat-related attributes, such as area of native pinewood or

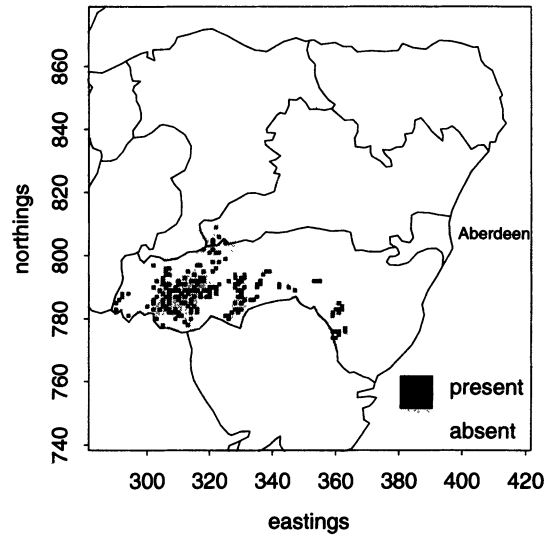

Fig. 1. Map of the observed spatial distribution of red deer in the intersection of the West Grampian and East Grampian Deer Management Group areas with Grampian Region (excluding Moray District).

mire within each square, or geometric attributes such as the cartesian coordinates (easting and northing).

MODELLING PROCEDURE

Buckland & Elston (1993) selected a simple random sample of 20% of the 1277 1 km squares in the study area, and assessed the effectiveness of logistic regression, coupled with Aitchison's (1955) method, to estimate abundance and spatial distribution of red deer from this sample. We assess the improvement that can be gained by using an autologistic model which incorporates both spatial autocorrelation and environmental covariates. Finally, the Gibbs sampler (Geman & Geman 1984) is incorporated into the autologistic model to allow the model to be fitted when only 20% of the squares are sampled.

Autologistic model

We define the response variable y_i to be 0 if deer are absent from square i and 1 if deer are present. The probability p_i that deer are present in square i is likely to depend on various habitat and climate covariates. In addition, it might be expected to depend on whether deer occur in neighbouring squares.

As a starting point for our autologistic model, we use Buckland & Elston's (1993) logistic regression model in which five covariates were selected:

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta_1\, altitude_i^2 + \beta_2\, northing_i + \beta_3\, mires_i + \beta_4\, easting_i + \beta_5\, pine_i.$$

This model was selected using a forward stepwise procedure applied to a single 20% sample of squares.

Our comparisons are based on the same 20% sample. Adding the term for autocorrelation, here called *autocov*, leads to the model

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta_1\, altitude_i^2 + \beta_2\, northing_i + \beta_3\, mires_i + \beta_4\, easting_i + \beta_5\, pine_i + \beta_6\, autocov_i$$

where

$$autocov_i = \frac{\sum_{j=1}^{k_i} w_{ij} y_j}{\sum_{j=1}^{k_i} w_{ij}} \qquad \text{eqn 1}$$

is a weighted average of the number of occupied squares amongst a set of k_i neighbours of square i. The weight given to square j is $w_{ij} = 1/h_{ij}$, where h_{ij} is the (Euclidean) distance between squares i and j.

The simplest scheme is to select $k_i = 4$ neighbours:

```
   *
*  yi  *
   *
```

A set of the eight nearest neighbours is given by

```
*  *  *
*  yi *
*  *  *
```

The size of the clique can be much bigger, depending on the range of distances over which the responses are thought to be correlated (see later).

Fitting an autologistic model is straightforward when presence/absence is recorded in every square. However, when data are available only from a sample of squares, the autocovariate cannot be evaluated, as the pattern of occupation in neighbouring squares is unknown. This difficulty does not arise in the ordinary logistic model, which does not utilize such information. One solution is to incorporate the Gibbs sampler into the autologistic model as described below.

Gibbs sampler

The Gibbs sampler allows us to estimate the presence/absence distribution in unsurveyed squares. It is a mechanism for generating an observation (presence or absence) for square i, given the pattern of occupation in the neighbouring squares. We can use an ordinary logistic model (ignoring spatial autocorrelation) to predict the distribution in unsurveyed squares, to provide a starting point. We then select one of the unsurveyed squares, delete its observation (predicted or actual), and generate a new value given the current estimated distribution in other squares. We now move to the next unsurveyed square, and repeat the process. This continues through all unsurveyed squares, and the whole process is iterated until convergence. The method is described in greater detail by Augustin *et al.* (in press). The full algorithm is summarized in Table 1.

The final presence/absence map is the predicted distribution of red deer, allowing for spatial autocorrelation. The repeated application of steps 5 and 6 of the algorithm allows us to assess the Monte Carlo variability introduced by the Gibbs sampler. In our example, we condition on the observed presence/absence from the 20% sample at each step, as we wish to predict presence/absence only for the 80% 'unsurveyed' squares. Thus, in steps 5 and 6 the stochastic method for generating presence/absence data is applied only to squares excluded from the 20% sample.

Note that steps 1–5 of the algorithm describe the process of fitting an autologistic model on its own, from which a stochastic prediction of the distribution is generated, and the Gibbs sampler is implemented at step 6. The selection of T, the number of iterations, depends on how much computing time is feasible and how fast the Gibbs sampler converges. For our example, each iteration involves the generation of 1021 random variates, one for each unsurveyed square (80% of 1277 = 1021). For the red deer data $T = 20$ iterations were found to be sufficient. The Gibbs sampling and the final prediction procedure (steps 5, 6 and 7 of the algorithm) were replicated $M = 120$ times, producing 120 independently and identically distributed samples of the map of presence/absence data. This number of replications was considered to be adequate in Buckland & Elston (1993).

The Gibbs sampler creates a stochastic realization of the whole map in each iteration and uses these stochastically generated observations to calculate the autocovariate. Consequently, the autologistic model fitted after each iteration depends very much on the random outcome of the stochastic map, which introduces considerable variability and reduces the rate of convergence of the Gibbs sampler. Instead, we can calculate the autocovariate using the predicted probability of occupation, instead of the generated presence/absence data, y_i:

$$autocov_i = \frac{\sum_{j=1}^{k_i} w_{ij} \hat{p}_j}{\sum_{j=1}^{k_i} w_{ij}} \qquad \text{eqn 2}$$

where k_i and w_{ij} are defined as in equation 1.

This approach is less computer intensive than using the Gibbs sampler since only the probability of presence is updated iteratively and there is no need to generate the responses for unsampled sites in each iteration. This modified method also convergences faster (Augustin *et al.* in press). In this example, $T = 10$ proved sufficient.

We now have four options for predicting the spatial distribution of presence/absence of red deer, given a random sample of counts.

Table 1. Algorithm for modelling presence/absence wildlife data, when just a random sample of grid squares is surveyed

1. Fit an ordinary logistic regression model to data from the random sample of squares. Calculate the fitted probability, $\hat{p}_i$, for all squares and create an initial map of presence/absence by generating for unsurveyed squares a value of 1 (presence) for site i with probability $\hat{p}_i$, and 0 (absence) otherwise.
2. Calculate the autocovariate for each square using equation 1 and the map obtained in the previous step.
3. Fit an autologistic model to the data from the random sample of squares using the autocovariates calculated from the most recent map of presence/absence.
4. Calculate the fitted probability, $\hat{p}_i$, for all squares using the parameter estimates from the autologistic model and the current map of presence/absence.
5. Update the map of presence/absence by generating observations from the $\hat{p}_i$ for unsurveyed squares.
6. Perform the Gibbs sampler.
   do T times:
   - repeat steps 2 and 3;
   - pick a random starting point in the map of squares.

   do for each unsurveyed square in turn:
   - calculate the autocovariate at square i;
   - calculate $\hat{p}_i$, the predicted conditional probability of presence at square i, and generate a new y_i from the $\hat{p}_i$.
7. Store the final map of presence/absence data, the fitted probabilities and the parameter estimates.

Method 1: *Logistic model* (Buckland & Elston 1993).
Method 2: *Autologistic model* (steps 1–5 of the algorithm in Table 1).
Method 3: *Autologistic model* combined with the *Gibbs sampler*.
Method 4: *Autologistic model* combined with the modified version of the *Gibbs sampler*.

For each method, the final map of presence/absence data can be generated stochastically from the estimated probabilities of occupation under that method.

COMPARISON OF MODELLING PROCEDURES

For each of the three new methods (i.e. the autologistic model on its own, with the Gibbs sampler, and with the modification to the Gibbs sampler), we repeatedly applied steps 5, 6 and 7 to generate $M = 120$ independently and identically distributed random samples of the map of presence/absence data and 120 sets of fitted probabilities. An overall estimate of the probability of occupation at each site can be obtained by averaging the 120 estimated probabilities for each site. Maps of the average predicted probability can then be compared with a map based on the (single) map of estimated probabilities of occupation from the logistic model. Similarly, a stochastic realization of presence/absence data can be created from the average probabilities of occupation from the Gibbs sampler. This map can be compared with a corresponding realization from the logistic model, or the autologistic model, or the 'true' spatial distribution of the deer.

An informal comparison of the estimated spatial distribution of deer can be obtained through visual examination of the maps described above. For these deer data, the true distribution is known, and so a more formal comparison can be obtained by examining the mis-classification rate of each method. For any stochastic realization of presence/absence data, this involves construction of the cross-tabulation shown in Table 2. Here $m1$ and $m2$ represent the frequencies of correctly classified squares, and $nm1$ and $nm2$ represent the frequencies of misclassified squares. The *matching coefficient* $(m1 + m2)/n$, represents the proportion of correctly classified squares. The $M = 120$ independent realizations of presence/absence maps produced under methods 2–4 can be used to calculate mean matching coefficients for each method. It is necessary to obtain 120 stochastic realizations of the presence/absence map for the logistic model to calculate a comparable mean matching coefficient for this model.

Another aspect of the modelling procedure that can be used to compare the four methods is the precision of abundance estimates. Calculation of an abundance estimate from a fitted model is based on Aitchison's (1955) approach. We use the above methods to estimate the number of occupied squares in the survey area, and to quantify the precision of that estimate. Separately, we estimate the mean number of deer per *occupied* square by the sample mean count from occupied squares in the 20% sample. Its standard error is estimated as the sample standard deviation of these counts divided by the square root of the number of occupied squares in the 20% sample. This two-stage process avoids the issue of how to model the counts when most are zero but non-zero counts exhibit substantially more variation than Poisson counts. The

Table 2. Matching counts: a method to assess the matching of predicted squares and observed squares

	Predicted classification		
True classification	Absent	Present	Total
Absent	$m1$	$nm1$	
Present	$nm2$	$m2$	
Total			n

second stage is common to all four models listed above, so the choice of model only affects the first stage. A bootstrap procedure (Efron 1979) is applied to quantify the precision of the number of occupied squares in the survey area as follows. First the estimated probabilities of occupation are obtained using the chosen method. Then $B = 120$ further maps of presence/absence are generated, in each of which observed data from the original 20% sample are retained. From each map, a new 20% sample of squares is selected and the chosen method is applied to each of these samples to give 120 estimates of the number of squares occupied, and hence 120 abundance estimates. For each method, a standard error of the mean of abundance is obtained using the standard deviation of the mean of the 120 bootstrap replicates of the abundance estimate. Conditioning on the observed counts in the 20% sample is analogous to applying a finite population correction in sampling theory.

Results

FITTING THE AUTOLOGISTIC MODEL

Autocovariates corresponding to various clique sizes, ranging from the simplest of size $k_i = 4$ to a square of side 9 km ($k_i = 80$), were computed. The most suitable autocovariate for the deer data was determined by looking at the change in deviance obtained by adding a specific autocovariate to Buckland & Elston's (1993) logistic model. In all cases, fitting the autocovariate gave a significant reduction in deviance. The autocovariate corresponding to a square clique of side 7 km produced the greatest reduction in deviance relative to the amount of extra computation involved and was used in all subsequent analyses. Once the autocovariate was fitted, the covariates northing, easting and mires did not have a significant effect on the conditional probability of presence. When we compared the precision of abundance estimates, we examined the effect of dropping these terms from the autologistic model (see later).

COMPARISON OF ABILITY TO PREDICT THE SPATIAL DISTRIBUTION

Visual comparison

A map of the observed spatial distribution of red deer is shown in Fig. 1. Figures 2–5 show maps of estimated probabilities of occupation and corresponding stochastic realizations of presence/absence data for the four modelling procedures. (In Figs 3–5, the estimated occupation probabilities represent the mean of the estimates obtained in $M = 120$ runs of the selected method.)

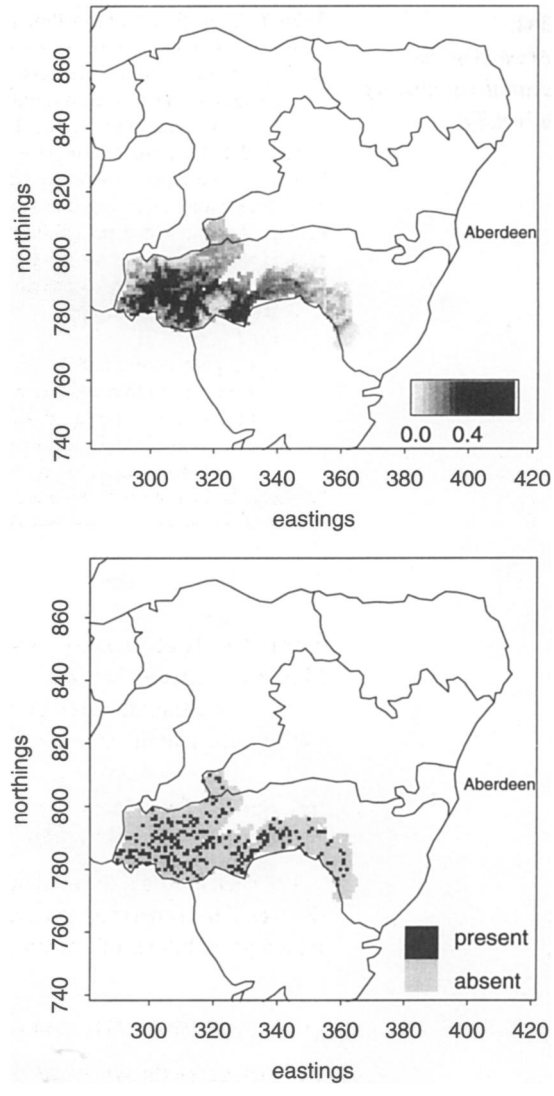

Fig. 2. (a) Estimated probabilities of occupation obtained using the logistic model (method 1) fitted to a 20% sample of the data. (b) Stochastic realization of presence/absence data obtained using the probabilities shown in (a).

Comparison of the maps in Figs 1 and 2–5 shows that both the Gibbs sampler (method 3) and its modification (method 4) produce less uniform estimated occupation probabilities than the logistic model; particularly low or high values of $\hat{p}_i$ are concentrated in appropriate areas to reflect the clustering in the true distribution of deer. The maps of presence/absence data from methods 2–4 also better reflect the true clustering in spatial distribution of deer than the map based on the ordinary logistic model. There are no obvious differences between the maps produced by methods 2–4.

Misclassification rates

The boxplots in Fig. 6 show the number of cases of incorrect classification of presence/absence obtained in 120 stochastic realizations of presence/absence data using the four different modelling methods. All the

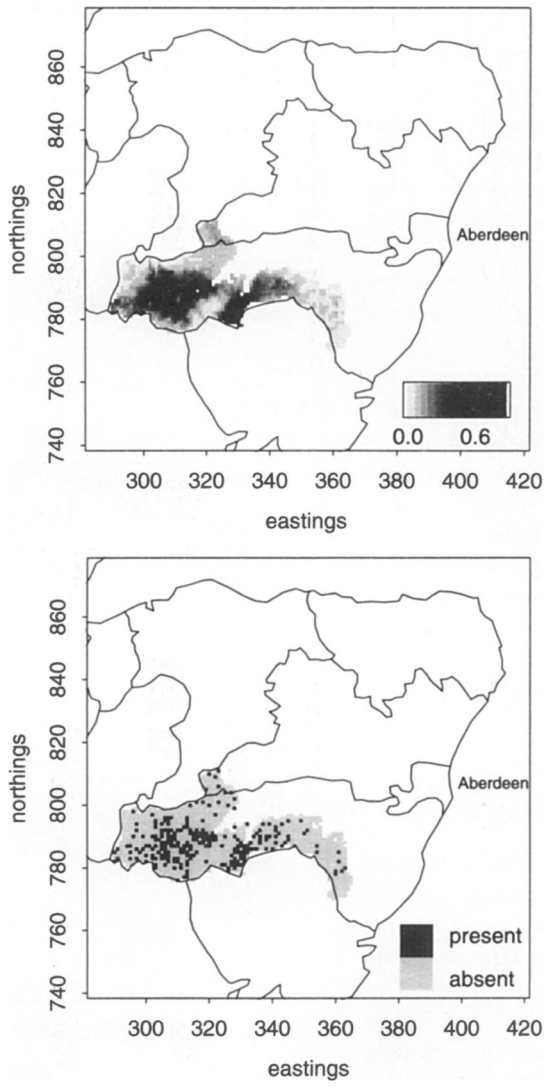

Fig. 3. (a) Estimated probabilities of occupation obtained using the autologistic model (method 2) fitted to a 20% sample of the data. (b) Stochastic realization of presence/absence data obtained using the probabilities shown in (a).

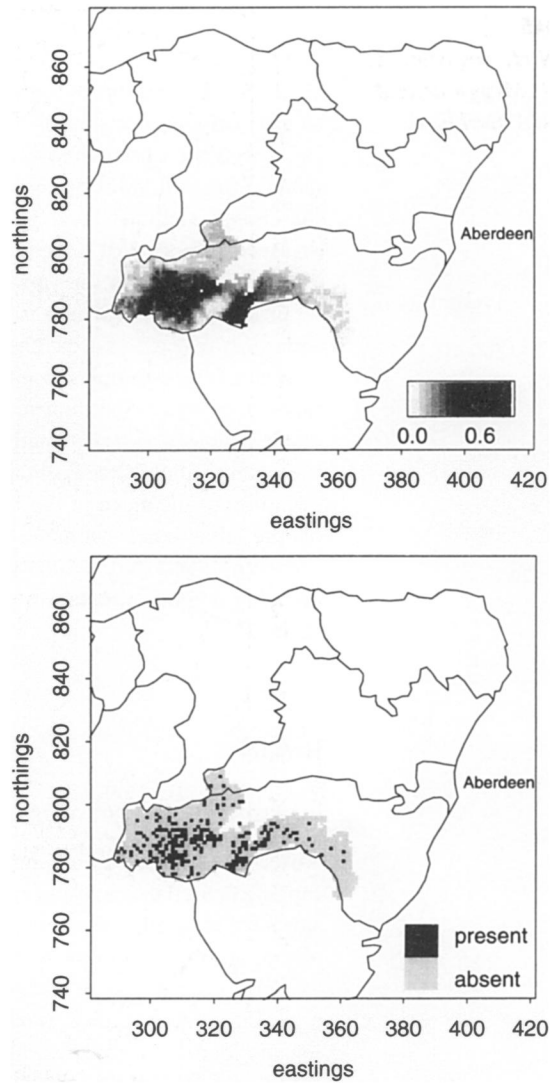

Fig. 4. (a) Estimated probabilities of occupation obtained using the autologistic model combined with the Gibbs sampler (method 3) fitted to a 20% sample of the data. (b) Stochastic realization of presence/absence data obtained using the probabilities shown in (a).

new methods have higher matching coefficients than the ordinary logistic model, with method 4 performing best. This method also has the lowest mean number of misclassified squares of each type and is the least variable method overall. The autologistic model on its own (method 2) performs better than in conjunction with the Gibbs sampler (method 3), especially in terms of correctly predicting absence. This suggests that the variability introduced by the Gibbs sampler outweighs any benefit of the theoretically superior method; we eliminate that undesirable variability under method 4.

COMPARISON OF REPEATABILITY FOR A GIVEN SAMPLE

Given a particular sample of squares, the estimate of the number of occupied squares $\Sigma\hat{p}_i$, produced by the logistic model is fixed because the model-fitting procedure does not involve any stochastic simulation of presence/absence. The autologistic model on its own, or combined with the Gibbs sampler or its modification, introduces extra variability in the estimate of the number of occupied squares since these models require one or more randomly generated presence/absence distributions. Table 3 shows that method 3 leads to far more variation in the estimated number of occupied sites compared with method 2. Method 4 is the least variable of these three methods.

COMPARISON OF THE PRECISION OF ABUNDANCE ESTIMATES

Having identified method 4 as the least variable of the new methods, we compare the precision of abundance estimates obtained using this method with those obtained from the ordinary logistic model (Table 4). The comparison is based on $B = 120$ bootstrap sam-

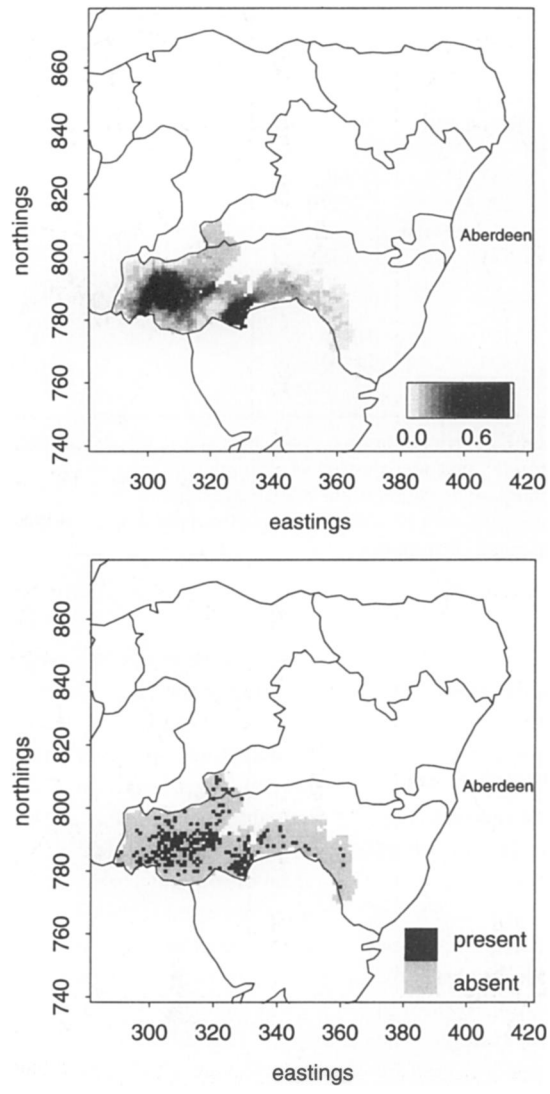

Fig. 5. (a) Estimated probabilities of occupation obtained using the autologistic model combined with the modification of the Gibbs sampler (method 4) fitted to a 20% sample of the data. (b) Stochastic realization of presence/absence data obtained using the probabilities shown in (a).

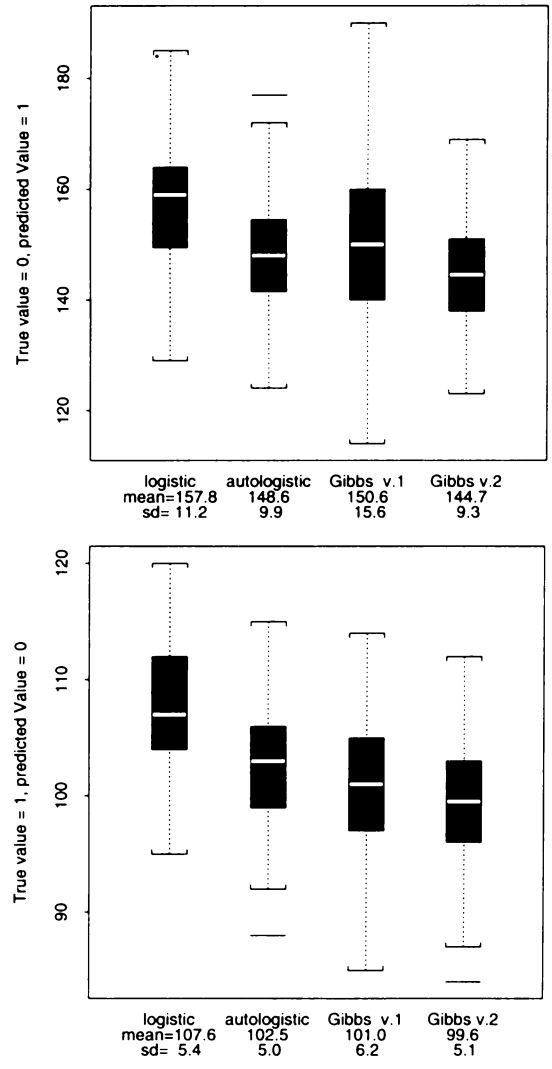

Fig. 6. Boxplots of numbers of sites for which deer were (a) absent but predicted to be present, and (b) present but predicted to be absent under methods 1–4. The models gave matching coefficients 0·792, 0·803, 0·808 and 0·809, respectively. The box represents the interquartile range, with the line inside representing the median. Whiskers are drawn to the nearest values not beyond a standard span of 1·5* interquartile range from the quartiles, and lines beyond the whiskers represent outlying data points.

ples for each method. The standard error of the number of occupied sites is substantially higher under method 4, although the estimate of abundance is slightly lower (i.e. closer to the true value). The reason for the higher standard error is as follows. The map of fitted probabilities from the logistic model (Fig. 2a) appears rather uniform. It has many squares with similar fitted probabilities. The map of the average fitted probabilities from method 4 (Fig. 5a) is less uniform and gives a closer reflection of the features of the observed data. Consequently, a 20% sample from this map will tend to be more variable than a 20% sample from the map corresponding to the logistic model. Bootstrapping the logistic model does not pick up the true variability of individual squares because it ignores autocorrelation in the residuals. Under this model, fitted probability of presence for any given square simply represents the average probability of all squares with the same (or similar) values of the chosen covariates. Thus, although method 4 performs better

Table 3. Mean and standard deviation of estimated number of sites occupied under methods 2–4, based on $M = 120$ repeated applications of each method. The standard deviation of each estimator indicates Monte Carlo variability only; it does not include a component for sampling error. Method 1 has no Monte Carlo variability associated with it

	Method 2*	Method 3	Method 4
Mean of estimated no. of sites occupied	238	237	237
Standard deviation	3·4	13·9	0·3

*Equivalent to one iteration of the Gibbs sampler.

Table 4. Comparison of precision of the logistic model (method 1) with the preferred method for incorporating spatial correlation (method 4) for estimating overall abundance

Parameter	Estimate Method 4*	Method 4 with reduced model†	Method 1	True values
No. of sites occupied	237	240	241	190
Standard error‡	59·4	59·5	24·8	NA
Mean no. of deer per occupied square	131	131	131	104
Standard error§	30·5	30·5	30·5	NA
Abundance	31 100	31 400	31 600	19 700
Standard error¶	10 600	10 700	8100	NA
Coeff. of Var. for abundance estimator	0·342	0·340	0·255	NA

*Using the autologistic model with altitude^2, northing, easting, mires, pine and autocov as covariates.
†The reduced autologistic model has altitude^2, pine and autocov as covariates.
‡The standard error is obtained as the standard deviation of the bootstrap estimates.
§The standard error is obtained as the standard error of non-zero counts from the simple random sample of squares.
¶The standard error includes both components of variation.

in predicting the local characteriztics of the spatial distribution, the ordinary logistic model gives more precise estimates of the overall number of deer.

If the significance of covariate terms is reassessed after fitting an autocovariate term, we find that covariates northing, easting and mires can be omitted, suggesting that these entered the ordinary logistic model only as proxies for spatial autocorrelation in the presence/absence data. If we exclude these covariates when fitting model 4, there is little difference in the standard error of the estimated number of occupied sites (Table 4); inclusion of non-significant covariates (overfitting) has had little effect on precision in this example. The apparent over-estimation of abundance is caused by over-representation of occupied sites in the particular 20% sample used in this analysis. Looking at the standard error obtained from method 4 confirms that the sample taken is not extreme, because a 95% confidence interval for the abundance estimate (13 109, 48 985) still includes the true abundance (19 700).

Discussion

The drawback of Buckland & Elston's (1993) logistic modelling approach is that it ignores the possibility of spatial autocorrelation in the residuals. The red deer data exhibit positive autocorrelation (i.e. if one square is occupied, then neighbouring squares are more likely to be), which is only partly explained by the habitat-related covariates in Buckland & Elston's (1993) model. An obvious consequence of this is that the fitted probabilities and stochastic realizations of presence/absence data obtained from the ordinary logistic model may not reflect the true level of clustering in the distribution of deer. Our results show that the autologistic model outperforms the logistic model in this respect, especially when it is combined with our modification of the Gibbs sampler (method 4). All three of the new methods are better than the ordinary logistic model in the way they mimic the true level of clustering in the distribution of deer. Method 4 is the most precise of the new methods, and yields the highest matching coefficient (i.e. fewest misclassified squares), suggesting that it should be used when the main objective of an investigation is to map the spatial distribution of a species.

For estimating global characteristics, such as the total number of occupied squares or the overall abundance of deer in our example, the performance of the two types of model is reversed, and method 4 gives higher standard errors than the ordinary logistic model. This is because the estimates from the logistic model reflect the average response to the habitat-related covariates across the entire region, whereas the autologistic approach incorporates local variations by adjusting for the response at neighbouring squares. The ordinary logistic model is therefore to be preferred when the main priority of the modelling procedure is to estimate global characteristics of wildlife distributions.

An important issue in the choice of an appropriate modelling strategy is the practicality of a particular approach. Methods 3 and 4 are computationally intensive relative to logistic regression. Just fitting a basic autologistic model (method 2) gives better results than the logistic model in terms of predicting the spatial distribution, and involves far less computation than methods 3 and 4.

The red deer data set is useful because the true spatial distribution of the deer at the time of the counts is known. We have attempted with reasonable success to recreate the known distribution, given presence/absence data on just a 20% random sample of 1 km squares. Our methods are likely to be of considerable value in situations where the resources to conduct a survey are limited, ruling out a complete count of the population. We have developed a method for modelling presence/absence data, whereas the

original red deer data were counts. Although we have noted that counts conditional on presence can be modelled as a separate exercise, extension of our methods to analyse counts directly would be useful. In our case, correlation between count and estimated probability of occurrence was not strong with an estimated correlation coefficient of $r = -0{\cdot}086$, probably because herd size was governed more by social factors than by features of the habitat. In many applications, such correlation can be expected, and spatial modelling of the counts might seem preferable. However, the natural distribution to assume for counts is the Poisson (possibly with overdispersion), and it can be shown mathematically that counts cannot be positively correlated under this model (Besag 1974). One option might be to model presence/absence data as described, then to fit a Poisson regression to counts conditional on presence, assuming independence between these non-zero counts. This is a weaker assumption than to assume independence between all counts, including zeros, and allows us to map abundance, rather than just species distribution. Further work on modelling counts in the presence of spatial autocorrelation is needed.

Acknowledgements

We are grateful to the Red Deer Commission for allowing us to use their red deer counts. N.H. Augustin thanks the SERC for financial support whilst at the University of Reading IACR receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom. We would also like to thank the referees for their helpful comments.

References

Augustin, N.H., Mugglestone, M.A. & Buckland, S.T. (in press) The role of simulation in modelling spatially correlated data. *Envirometrics*.

Aitchison, J. (1955) On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of the American Statistical Association*, **50**, 901–908.

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society B*, **36**, 192–236.

Buckland, S.T. & Elston, D.A. (1993) Empirical models for the spatial distribution of wildlife. *Journal of Applied Ecology*, **30**, 478–495.

Efron, B. (1979) Bootstrap methods: another look at the jackknife. *Annals of Statistics*, **7**, 1–26.

Geman, S. & Geman, D. (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.

Högmander, H. & Møller, J. (1995) *Estimating distribution maps from atlas data using statistical methods of image analysis. Biometrics*, **51**, 393–404.

Osborne P.E & Tigar, B.J. (1992) Interpreting bird atlas data using logistic models: an example from Lesotho, Southern Africa. *Journal of Applied Ecology*, **29**, 55–62.

Walker, P.A. (1990) Modelling wildlife distributions using a geographic informations system: kangaroos in relation to climate. *Journal of Biogeography*, **17**, 279–289.

Received 26 May 1994; revision received 3 February 1995