Computer-based spatial extension of forest fauna survey data: current issues, problems and directions

S. Ferrier[1]

ABSTRACT

Computers are increasingly being employed to extend data from surveys of forest fauna across unsurveyed areas, using models relating species occurrence or population density to mapped environmental attributes.

This chapter examines current issues and problems associated with this methodology:

- avoiding and overcoming bias in survey data used to derive computer-based models
- selecting an appropriate modelling technique
- incorporating spatial analysis into models
- testing and refining predictions

INTRODUCTION

Effective conservation and management of forest fauna requires a sound knowledge of the distribution of species across relatively large geographical areas. Depending on the problem at hand these areas may vary in size from an individual logging compartment, state forest or national park, through to an entire region, state or even the entire continent.

Acquiring knowledge about the distribution of forest fauna across large areas has always presented special problems. Forest fauna cannot normally be mapped remotely using aerial photography or satellite imagery, but must instead be directly surveyed on the ground. Ground surveys are expensive and time consuming due to the secretive or cryptic nature of many species, coupled with the general ruggedness and inaccessibility of the areas they inhabit. Typically, therefore, we can afford to directly survey only a very small proportion of any geographical area of interest (Margules and Stein 1989). The results of these surveys must then be extended or interpolated across the remaining unsurveyed parts of the area. This spatial extension of forest fauna survey data is usually based on correlations between known occurrences or population densities of a species and remotely mapped or interpolated environmental attributes such as topography, climate, geology, forest type and land-use history.

Over the last decade there has been an increasing interest in the use of computer-based techniques to optimize the spatial extension of environmental survey results. In Australia, relationships between plants or animals and mapped environmental attributes have been modelled using an array of statistical and related software packages. Spatial extension or predictive mapping has then been achieved using computer-based geographical information systems (GIS) to manipulate mapped environmental attributes in order to display all areas potentially suitable for the species of interest or to map probabilities of occurrence or potential population densities (see Fig. 1). For recent Australian examples of these techniques see Austin *et al.* (1984), Ferrier (1984), Nix (1986), Walker and Moore (1988), Mackey and Bayes (1989), Nicholls (1989), Stockwell (1989), Ferrier and Smith (1990), Neave and Norton (1991).

Computer-based techniques such as these offer an enormous, but as yet largely untapped, potential to efficiently extend the results of scattered ground surveys of forest fauna throughout areas not yet surveyed due to lack of funding, time or resources. These techniques do, however, have shortcomings and pitfalls which need to be recognized and addressed if we are to obtain maximum benefit from the approach. As with any new application of computer technology, there is a danger that the "gee whiz look at that" capabilities of the technology can obscure or divert attention from basic underlying problems. The more

---

[1] National Parks and Wildlife Service of New South Wales, P.O. Box 402, Armidale, New South Wales, Australia 2350.

Pages 221–27 *in* CONSERVATION OF AUSTRALIA'S FOREST FAUNA, ed Daniel Lunney. Published by Royal Zoological Society of NSW, Mosman. 1991.

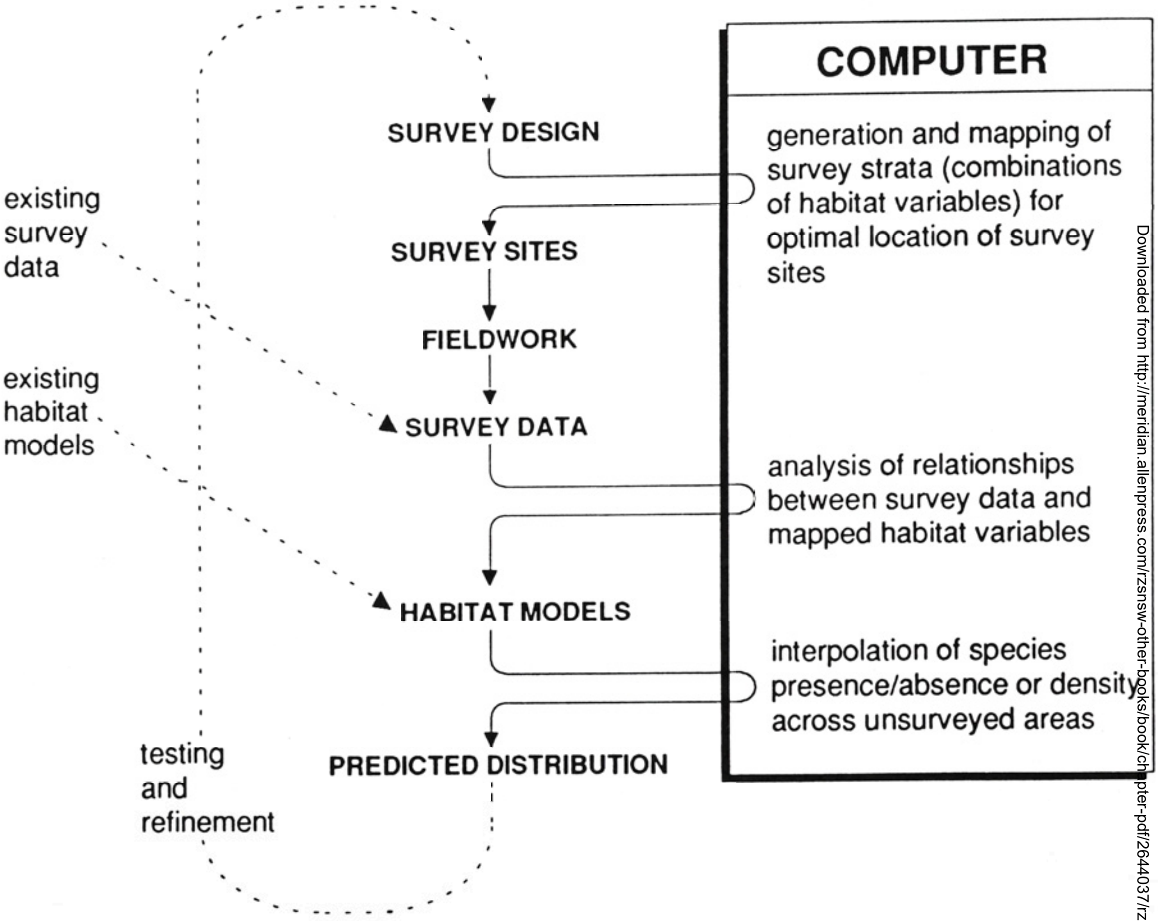

Fig. 1. An integrated procedural framework for applying computer-based techniques to the design, analysis and spatial extension of forest fauna surveys.

spectacular the complex mathematical models and glossy computer-generated maps become, the more cautious we need to be.

This chapter highlights and discusses some of the more important issues and problems currently associated with computer-based extension of forest fauna survey data.

CURRENT ISSUES AND PROBLEMS

Avoiding and Overcoming Survey Bias

Computer-based modelling and interpolation of forest fauna distributions are often based on the results of previous surveys never designed with this aim in mind. This situation is frequently unavoidable due to the cost of conducting major new field surveys. Often the only real alternative is to take advantage of existing survey data. However, the potential suitability of data from previous surveys to modelling and interpolation varies greatly. If adequate information is available about the survey design used, as well as the accurate locations of all survey sites or transects, then the survey data may be useful. At the other end of the spectrum are the extensive data sets derived from *ad hoc* or opportunistic surveying of forest fauna such as those held in museum collections, databases for wildlife atlases or in the notebooks of naturalists. Such data sets pose two major problems when subjected to computer-based modelling and interpolation. Firstly, the described geographical location of each record may not be sufficiently accurate to allow meaningful correlation with mapped environmental attributes. Secondly, these data sets normally contain information on recorded presences of a species but not recorded absences. We therefore know where surveys found the species but not where else they searched without finding the species. Without the latter information, analyses of relationships between forest fauna distribution data and mapped environmental attributes may suffer from bias. For example, a species may incorrectly appear to be confined to low elevation areas simply because nobody has searched for it at higher elevations.

Data sets generated from *ad hoc* or opportunistic survey work need not, however, be discarded simply because of a lack of data on the absence of species. Bias in such data sets can be at least partially removed using a number of techniques. For example it might be assumed that most survey work would have been conducted close to roads, thereby allowing the distribution of mapped environmental attributes along all roads to be used as an approximate control or domain with which to compare the environmental characteristics of sites where a species was actually recorded. Alternatively, the range of sites (or domain) within which searches were conducted for a particular frog species, for example, could be approximately reconstructed by pooling all sites at which frogs of any species were recorded.

Whenever time and resources permit new forest fauna surveys to be conducted, these should be designed to maximize information returned per unit of survey effort. This is most often achieved by stratifying survey sites on the basis of vegetation or land units derived from thematic, parametric or integrated land classifications (see reviews in Margules and Scott 1984; Myers *et al.* 1984; Gunn *et al.* 1988). Gillison and Brewer (1985) recently introduced the "gradsect" approach to environmental survey design as an efficient means of balancing the need for stratification against logistic restrictions such as accessibility and travel time. Ferrier and Smith (1990) present an example of the application of GIS technology in generating a cost-effective multivariate design for surveying rainforest vertebrates in north-eastern New South Wales.

Selecting an Appropriate Modelling Technique

A wide variety of computer-based techniques can be used to analyse and model relationships between forest fauna and mapped environmental variables:

- Simple profile or homoclime matching. A range within each environmental variable is identified as being "suitable" for the species of interest. These ranges are then combined to form a climatic or habitat "envelope". Probably the best known example of this approach is the BIOCLIM package (Nix 1986; Busby 1991). A similar technique is provided in the E-RMS package (Ferrier 1988).
- Generalized linear statistical modelling. A number of multivariate statistical techniques have been used to model species-habitat relationships. For example, Austin *et al.* (1984) used generalized linear modelling to analyse the distributions of various eucalypt species. Ferrier (1984) used logistic regression to model the distribution of the Rufous Scrub bird *Atrichornis rufescens* in northeastern New South Wales (see Fig. 2). Other recent examples are given in Lindenmayer *et al.* (1990) and Nicholls (1991).
- Nonparametric statistical modelling. Mackey and Bayes (1989) used a set of algorithms for monotonic functions to model leaf size in tropical Queensland forest canopies. Skidmore and Turner (1988) have also developed a nonparametric classifier potentially applicable to species-habitat modelling.
- Decision tree induction. This is a form of machine learning which is gaining popularity as a technique for modelling wildlife habitat relationships (e.g., Walker and Moore 1988; Stockwell 1989; Ferrier and Smith 1990; Stockwell *et al.* 1990).

This diversity of approaches poses a potential problem for anybody wishing to model relationships between forest fauna and mapped environmental variables. Which technique is most appropriate? Are different techniques appropriate for different applications? Not enough comparative work has been undertaken to address these questions. We do, however, know that some of these techniques make different assumptions about how environmental variables combine to determine habitat suitability. For example, simple profile matching assumes that the environmental variables do not interact or compensate for one another, and therefore that potentially suitable habitat must have suitable values for all variables (Nix 1986). Generalized linear modelling assumes that variables can interact and compensate for one another. The

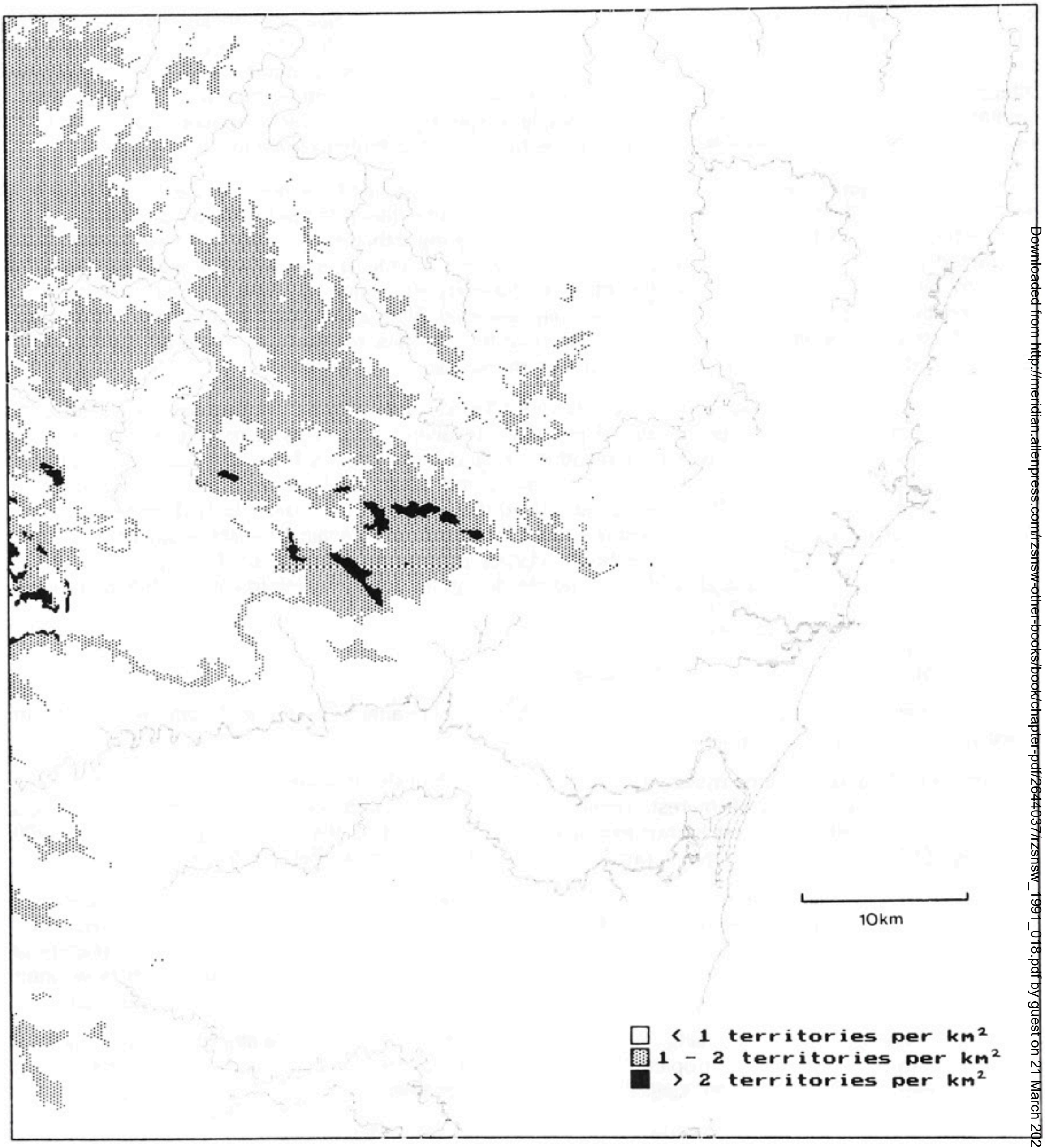

Fig. 2. Computer-generated map of the potential distribution of the Rufous Scrub-bird in the Dorrigo/Coffs Harbour region, New South Wales. The map was produced by linking a generalized linear model derived from surveys of scrub-bird density throughout northeastern New South Wales with a geographical information system containing information on terrain, climate and vegetation. For details of the model, see Ferrier (1984).

net suitability of an area is determined by summing the effects of individual variables (Nicholls 1991). Unfortunately not enough is yet known about the nature of relationships between forest fauna and environmental variables to determine which of these assumptions is the most appropriate for any given application. The nonparametric statistical and decision tree induction techniques make less stringent assumptions and therefore perhaps represent good interim alternatives as robust, general purpose modelling tools. Decision tree induction is particularly attractive as it can generate models in the form of "if-then-else" rules for predicting the occurrence or population densities of species. These are more easily understood by forest managers (such as rangers and foresters) than complex multivariate statistical models.

Incorporating Spatial Analysis

Many of the currently popular approaches to habitat analysis and interpolation do not adequately handle the spatial dimension of interactions between forest fauna and habitat. These techniques usually treat survey sites as a sample of statistically independent point observations, and attempt to correlate the occurrence or density of species at each site with the vegetation, geology, climate etc at that site. In reality, however, the occurrence of a species is likely to be determined not only by the environment at a particular point but also by the environment around the point. For example, in northeastern New South Wales the Eastern Bristlebird *Dasyornis brachypterus* and Rufous Scrub-bird can both occur in eucalypt open forest but only where this forest is near to rainforest. Various species of owls, gliders and macropods have large home ranges and therefore require a large area of suitable habitat often containing a number of diverse components. Habitat suitabiity for these species must be assessed not only on the basis of environmental conditions at a point or grid square but also on the basis of environmental conditions in the surrounding area.

Another potential problem is that distributions of most species of forest fauna are likely to exhibit at least some degree of spatial autocorrelation (Sokal and Oden 1978). For historical and other reasons including stochastic processes of colonization and extinction, distributions of plants and animals are often patchy regardless of habitat suitability. The occurrence or population density of a species at a particular site will therefore be correlated not only with the habitat at that site but also with the occurrence of other individuals of that species in the surrounding area.

Statistical and related packages rarely provide adequate techniques to handle spatial problems such as these. Even if they do, these packages cannot easily access required information about the environment surrounding a point because this is normally stored separately on the computer in a GIS. Better tools for spatial analysis and model building need to be incorporated directly into the GIS environment, thereby allowing a truly spatial approach to species-habitat analysis. Unfortunately, despite frequent references to "analysis" and "modelling" in the GIS literature, currently available GIS packages rarely provide adequate tools for analysis and model building but instead specialize in tools with which to apply or interpolate existing models developed outside the GIS.

Field Validation and Model Refinement

Spatial extensions of forest fauna survey data based on correlations with mapped environmental attributes are merely predictions and therefore should, wherever possible, be subjected to field validation. Nix and Gillison (1985) have advocated a cyclical or iterative approach to the modelling and interpolation of faunal survey data. Ferrier and Smith (1990) have recently applied this approach to a survey of rainforest amphibians, reptiles and mammals in northeastern New South Wales. Models developed from initial surveys were tested during subsequent surveys. Repeated surveys not only allow model testing but also cost-effective refinement. Initial surveys of presence/absence can generate maps depicting broad limits of distribution. Subsequent surveys can focus within these predicted areas and if necessary employ refined measures of density as opposed to presence/absence.

Later surveys can also incorporate direct measurement of microhabitat variables that cannot be detected or mapped remotely. Effective conservation and management of forest fauna requires information not only on macrohabitat relationships derived from correlations with remotely mapped attributes, but also on microhabitat requirements derived from direct field measurement. The latter enables better assessment of the potential impact of proposed land-use activities such as logging and prescribed burning. An example of the integrated development and application of macrohabitat and microhabitat models is that for the Rufous Scrub-bird (Ferrier 1984, 1985). A logistic regression model relating scrub-bird occurrence to average annual rainfall, elevation and forest type has facilitated broad area mapping of potential distribution (see Fig. 2). A discriminant function model classifying scrub-bird habitat suitability on the basis of direct measurement of microhabitat variables has enabled rapid field assessment of habitat within broad areas mapped using the macrohabitat model.

FUTURE DIRECTIONS

Computers offer an enormous potential to use existing survey data to predict the occurrence of forest fauna in unsurveyed areas. By taking advantage of this potential we can maximize the benefit and cost-effectiveness of fauna surveys. However, despite many recent technological advances, we still have a long way to go in this field. This paper has highlighted just some of the problems involved. It will be interesting to see how these and other related problems are handled over the next decade. A few of the developments that I think will, or at least should, occur are:

- Forest fauna surveys will be better designed to ensure cost-effective spatial extension of results. GIS will be increasingly employed to help optimize survey designs rather than just for subsequent data storage and analysis.
- Modelling techniques will be intensively tested and compared to determine the suitability of these methodologies for different forest fauna applications.
- Sophisticated tools for spatial analysis and model building will be incorporated into the GIS software environment, facilitating more efficient and realistic modelling of relationships between forest fauna and mapped environmental attributes.
- Macrohabitat models developed for broad area spatial extension of forest fauna data will be increasingly integrated with microhabitat models and refined field surveys of species density. This will allow better assessment of the potential impact of alternative conservation and management options across large geographical areas.

REFERENCES

Austin, M. P., Cunningham, R. B. and Fleming, P. M., 1984. New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio* **55:** 11–27.

Busby, J. R., 1991. BIOCLIM — a bioclimate analysis and prediction system. Pp. 64–68 in *Nature conservation: cost effective biological surveys and data analysis* ed by C. R. Margules and M. P. Austin. CSIRO: Canberra.

Ferrier, S., 1984. The status of the Rufous Scrub-bird *(Atrichornis rufescens)*: habitat, geographical variation and abundance. PhD Thesis, University of New England, Armidale.

Ferrier, S., 1985. Habitat requirements of a rare species, the Rufous Scrub-bird. Pp. 241–48 in *Birds of the eucalypt forests and woodlands: ecology, conservation, management* ed by A. Keast, H. F. Recher, H. Ford and D. Saunders. Surrey Beatty & Sons: Sydney.

Ferrier, S., 1988. Environmental resource mapping system (E-RMS), users manual for version 1.2. National Parks and Wildlife Service of New South Wales.

Ferrier, S. and Smith, A. P., 1990. Using geographical information systems for biological survey design, analysis and extrapolation. *Australian Biologist* **3:** 105–16.

Gillison, A. N. and Brewer, K. R. W., 1985. The use of gradient directed transects or gradsects in natural resource surveys. *J. of Env. Manage.* **20:** 103–27.

Gunn, R. H., Beattie, J. A., Reid, R. E. and van de Graaff, R. H. M. (eds), 1988. Australian soil and land survey handbook: guidelines for conducting surveys. Inkata Press: Melbourne.

Lindenmayer, D. B., Nix, H. A., McMahon, J. P. and Hutchinson, M. P., 1990. Bioclimatic modelling and wildlife conservation and management — a case study on Leadbeater's Possum. Pp. 253–74 in *Proceedings of a conference on small mammals.* Melbourne Zoo: Melbourne.

Mackey, B. G. and Bayes, T., 1989. A modelling framework for the spatial extension of ecological relations in vegetation studies. Pp. 93–97 in *Proceedings, SSA-IAMCS 1989 Biennial Conference.*

Margules, C. R. and Scott, R. M., 1984. Review and evaluation of integrated surveys for conservation. Pp. 1–17 in *Survey methods for nature conservation* ed by K. Myers, C. R. Margules and I. Musto. CSIRO: Canberra.

Margules, C. R. and Stein, J. L., 1989. Patterns in the distribution of species and selection of nature reserves: an example from eucalyptus forests in south-eastern New South Wales. *Biological Conservation* **50:** 219–38.

MYERS, K., THACKWAY, R. M. AND READ, V. T., 1984. Faunal survey — principles and problems. Pp. 217–66 in *Survey methods for nature conservation* ed by K. Myers, C. R. Margules and I. Musto. CSIRO: Canberra.

NEAVE, H. M. AND NORTON, T. W., 1991. Integrated management of forest wildlife: comments on new ways to research habitat. Pp. 229–36 in *Conservation of Australian forest fauna* ed by D. Lunney. Roy. Zool. Soc. NSW, Mosman.

NICHOLLS, A. O., 1989. How to make biological surveys go further with generalized linear models. *Biological Conservation* **50:** 51–76.

NICHOLLS, A. O., 1991. Examples of the use of generalized linear models in analysis of survey data for conservation evaluation. Pp. 54–63 in *Nature conservation: cost effective biological surveys and data analysis* ed by C. R. Margules and M. P. Austin. CSIRO: Canberra.

NIX, H. A., 1986. A biogeographic analysis of Australian elapid snakes. Pp. 4–15 in *Atlas of elapid snakes of Australia* ed by R. Longmore. AGPS: Canberra.

NIX, H. A. AND GILLISON, A. N., 1985. Towards an operational framework for habitat and wildlife management. Pp. 39–55 in *Wildlife management in the forests and forestry-controlled lands in the tropics and the southern hemisphere* ed by J. Kikkawa. Queensland University Press.

SKIDMORE, A. K. AND TURNER, B. J., 1988. Forest mapping accuracies are improved using a supervised nonparametric classifier with SPOT data. *Photogrammetric Engineering and Remote Sensing* **54:** 1415–21.

SOKAL, R. R. AND ODEN, N. L., 1978. Spatial autocorrelation in biology. *Biological Journal of the Linnean Society* **10:** 199—228.

STOCKWELL, D. R. B., 1989. Modelling of qualitative data via machine learning. Pp. 98–103 in *Proceedings, SSA-IAMCS 1989 Biennial Conference.*

STOCKWELL, D. R. B., DAVEY, S. M., DAVIS, J. R. AND NOBLE, I. R., 1990. Using induction of decision trees to predict Greater Glider density. *AI Applications* **4:** 33–43.

WALKER, P. A. AND MOORE, D. M., 1988. SIMPLE, an inductive modelling and mapping tool for spatially-oriented data. *International Journal of Geographic Information Systems* **2:** 347–363.