Recent developments in spatial methods and data in biogeographical distribution modelling – advantages and pitfalls

MISKA LUOTO AND RISTO HEIKKINEN

Luoto, Miska & Risto Heikkinen (2003). Recent developments in spatial methods and data in biogeographical distribution modelling – advantages and pitfalls. *Fennia* 181: 1, pp. 35–48 Helsinki. ISSN 0015-0010.

Geography has a long tradition in studies of geographical distribution of flora and fauna. Detailed mappings of the distributions of biota over wide regions can produce highly valuable biogeographical data, but are extremely laborious. These challenges in biogeographical mapping, as well as the need for mitigation tools for the adverse impacts of human disturbance on the landscape and biodiversity, have stimulated the development of new approaches for assessing biogeographical patterns. Particularly, the ability to model distribution patterns of organisms and habitat types has recently increased along with the theoretical and methodological development of biogeography and spatial ecology, and modern spatial techniques and extensive data sets (provided e.g., by earth observation techniques). However, geographical data have characteristics which produce statistical problems and uncertainties in these modelling studies: 1) the data are almost always multivariate and intercorrelated, 2) the data are often spatially autocorrelated, and 3) biogeographical distribution patterns are affected by different factors operating on different spatial and temporal scales. Especially remote sensing and geographic information data provide powerful means for studies of environmental change, but also include pitfalls and may generate biased results. Quantitative analysis and modelling with correct and strict use of spatial statistics should also receive more attention. The issues discussed in this paper can have relevance in several fields of application of geographical data.

Miska Luoto & Risto Heikkinen, Finnish Environment Institute, Research Programme for Biodiversity, P.O. Box 140, FIN-00251 Helsinki, Finland. E-mail: miska.luoto@ymparisto.fi, risto.heikkinen@ymparisto.fi. MS submitted 8 November 2002.

Introduction

Spatial patterning and distribution of organisms has traditionally attracted much interest and has stimulated research in geography. Consequently, issues such as which environmental factors explain the distribution of various plants has continuously had a central role in biogeographical research since the pioneering work of Alexander von Humboldt in the early 19th century (von Humboldt 1807; Turner 1989).

Nowadays, the spatial distribution of organisms is also strongly affected by the adverse impacts of human disturbance, particularly habitat loss and fragmentation (Tilman et al. 1994; Enoksson et al. 1995; Huxel & Hastings 1999; Noss 2001; Fahrig 2002; Schmielgelow & Mönkkönen 2002; see also Watson 2002). This development has given rise to increasing concern about the potential loss of important natural values and has inspired a development of new techniques to map and monitor wide areas of land. Such techniques are clearly urgently required to analyse and model human-based impacts on landscape and biodiversity (Griffiths et al. 1993). The technical tools and theoretical framework needed in the modelling of spatial distribution of species in landscapes have actually improved due to the recent methodological developments in biogeography and spatial ecology, as well as in statistical methods and spatial data analysis (Scott et al. 1993; Stoms & Estes 1993; Hanski 1998; Debinski et al. 1999; Guisan & Zimmermann 2000; Roy & Tomar 2000).

However, the integration of geographical analysis and modelling and GI (geographic information) technology and spatial data from different sources requires transdisciplinary skills between geography, ecology, statistics and social sciences. Thus the pitfalls for the misuse of GIS technology with its high calculation capacity are very obvious. Several recent papers dealing with spatial data have highlighted the fact that the correct use of spatial statistics with GI and RS (remotely sensed) data is increasingly important (Stoms 1992; Luoto 2000a; Liebhold & Gurevitch 2002; Perry et al. 2002).

Geographical data sets have several characteristics which separate them from many other kinds of data sets. These features produce severe statistical problems and uncertainties in the modelling studies of biogeographical distribution data. First, spatial data are almost always multivariate, i.e. there are more than one variate or analyte of interest, which are correlated to some degree. Second, the spatial location of each data point can be described by its geographic coordinates. This positional association is often also manifested in another way, namely through some form of spatial correlation (Legendre 1993; Brito et al. 1999). Thirdly, distribution patterns and processes are often affected by different factors operating on different scales. Spatial systems generally show characteristic variability on a range of spatial, temporal and organizational scales and therefore, there is no single natural scale on which geographical phenomena should be studied (see Wiens 1989; Levin 1992; Stoms 1994).

Many of the above-mentioned problems are currently topical in geography, especially in studies with GI and RS data sets (Högmander & Møller 1995; Augustin et al. 1996). This paper does not aim at representing a fully comprehensive review covering all the relevant issues and their backgrounds in contemporary geographical data mining, analysis and modelling. Instead, we focus in this commentary paper on some selected key issues in the development of biogeography and landscape ecology, and particularly on the possibilities and potential pitfalls of analysing and modelling spatial data, which are attracting increasing attention. Many of the methodological issues and problems touched upon in this paper are those which researchers in biogeography and landscape ecology constantly face, and moreover, similar questions are also of importance in other fields of geography. Thus, the ideas presented here are applicable in several other fields of study where geographical data are applied.

Benchmarks in the development of biogeography and spatial ecology

In 1807, von Humboldt described the latitudinal and altitudinal distribution of vegetative zones. His work *'Ideen zu einer Geographie der Pflanzen nebst einem Naturgemälde der Tropenländer'* provided an inspiration to studies of the geographic distribution of plants and animals. Throughout the 19th century, botanists and zoologists described and explained the spatial distributions of various taxa mainly by macroclimatic factors such as temperature and precipitation (Turner 1989; Granö & Paasi 1997).

The emerging view was that strong interdependencies between climate, biota, and soil lead to long-term stability of the landscape in the absence of climatic changes. The early biogeographical studies also influenced Clements' theory (Clements 1936) of successional dynamics, in which the stable endpoint, the climax vegetation, was determined by macroclimate over a broad region. Clements stressed temporal dynamics but did not emphasise spatial patterning. The development of gradient analysis (Whittaker 1967) allowed description of the continuous distribution of species along environmental gradients. Abrupt discontinuities in vegetation patterns were believed to be associated with discontinuities in the physical environment.

Watt (1947) first linked space and time on a broader scale in biogeography. He described the distribution of the entire temporal progression of successional stages as a pattern of patches across a landscape. The complex spatial pattern across the landscape was constant, but this constancy in the pattern was maintained by temporal changes at each point. The modern concept of the shifting steady-state mosaic, which incorporates natural disturbance process, is related to Watt's conceptualisation (Turner 1989).

The interest of biogeographers in spatial aspects increased after the introduction of the theory of island biogeography by MacArthur & Wilson (1967). The new theory explained how distance and area together regulate the balance between immigration and extinction in island populations. The three basic characteristics of insular biotas are: 1) the number of species increases with increasing island size, 2) the number of species decreases with increasing distance to the nearest continent or other source of species, and 3) a continual turnover in species composition occurs, owing to recurrent colonisations and extinctions, but the number of species remains approximately the same. MacArthur and Wilson (1967) proposed that the number of species inhabiting an island represents an equilibrium between opposing rates of colonisation and extinction.

The theory of island biogeography was based on simple mathematical models and looked for equilibria in species numbers using the data on species occurrences. The basic assumption of equilibrium in spatially defined ecological systems was later found to be inappropriate (Haila 2002). Since the 1980s, the theoretical presuppositions of island biogeography have been challenged, and empirical research has become multifaceted. Fragments of a particular habitat type are viewed as elements in a heterogeneous landscape rather than as ‘islands’ surrounded by a hostile ‘sea’. As the interest in island biogeography declined, it was replaced by metapopulation (Levins 1969) as the paradigm of spatial ecology (Hanski 1998, 1999).

Spatial dynamics has received increasing attention in many areas of biogeography and ecology during recent decades (Mooney & Godron 1983; Turner 1989; Wiens 1997; Hanski 1999). The role of spatial landscape pattern, i.e. the distribution and structure of different habitats, in influencing species distribution is also increasingly studied by landscape ecologists (Naveh & Lieberman 1984; Turner 1989; Forman 1995) and metapopulation ecologists (Verboom et al. 1991; Thomas et al. 1992; Hanski 1999). Finally, the influence of spatial locations of individuals, populations and communities on their dynamics has been demonstrated in a number of recent spatial ecological studies (Hanski & Gilpin 1997; Hanski 1999).

Nowadays, three different approaches in large-scale spatial ecology can be distinguished (Hanski 1998): 1) theoretical ecology, 2) landscape ecology and 3) metapopulation ecology. Theoretical ecologists have investigated a range of models depicting individuals with localized interactions and restricted movement range in uniform space, demonstrating how population dynamics can generate complex dynamics and spatial patterns without any environmental heterogeneity (Tilman & Kareiva 1997). By contrast, landscape ecologists have been occupied by descriptions of the generally complex physical structure of real environments, distribution of resources in landscapes, and the movements of individuals (Forman 1995; Wiens 1997). Metapopulation ecology makes the simplifying assumption that suitable habitat patches for the focal species occur as a network of idealised habitat patches varying in area, degree of isolation and quality and surrounded by uniformly unsuitable habitat (Hanski & Gilpin 1997; Hanski 1998).

Potential of remote sensing and GIS-based modelling

Along with the conceptual advances discussed in the previous section, the availability of modern computer software and hardware (e.g., geographical information systems, increased computer speed and memory) has recently expanded our abilities to address many of the most interesting and critical problems in biogeography. Prior to the availability of these tools, analysis of many of the important issues associated with spatial data was impossible because of the sheer magnitude of the data sets and the complexity of their analysis (Liebhold & Gurevitch 2002; Nagendra 2001).

Spatial data on the geographical distribution of habitats and species are often sparse, and factors affecting their distribution patterns are insufficiently known. For modelling and predicting species distribution and location of areas with considerable ecological and nature conservation values, accurate data would be desirable. In reality, such data covering extensive areas is often not available or it is too expensive to be acquired by research projects. As highlighted by several authors (e.g., Margules & Austin 1991; Cherril et al. 1995; Debinski et al. 1999; Nagendra & Gadgil 1999), it is necessary to develop spatial modelling methodologies for rapid and cost-effective mapping of large areas to assess their ecological value for nature conservation.

The ability to analyse, model and predict distribution patterns of habitats and species on the basis of landscape variables derived from RS and GI data could mitigate the damage caused by human land use and facilitate the preservation of biodiversity (Scott et al. 1993; Stoms & Estes 1993; Debinski & Humphrey 1997; Gould 2000;

Guisan & Zimmermann 2000; Roy & Tomar 2000; Nagendra 2001; Suárez-Seoane et al. 2002). The growth in the availability of remotely sensed data and the development of GIS techniques allows access to an extensive assortment of potential spatial covariates, so that analyses of factors affecting biogeographical distribution patterns can be adapted to different spatial applications. Moreover, they enable us to derive predictive models from relations within the data and to spatially extrapolate potential species distribution, abundance and/or habitat preferences from those models to wider areas (Stoms & Estes 1993; Brito et al. 1999).

In several studies species distribution patterns for selected taxonomic groups have been modelled using remotely sensed environmental data, for example birds, mammals, plants, reptiles and butterflies (Austin et al. 1990; Pereira & Itami 1991; Augustin et al. 1996; Brito et al. 1999; Gould 2000: Luoto et al. 2002a). Debinski et al. (1999) suggested that GI and RS data could be employed in modelling of species distribution, because species are often significantly correlated with one or more remotely sensed habitat types, particularly when they are highly specialized in their habitat utilisation. In order to build predictive models of species distribution using remotely sensed data, a species must either be common enough and/or habitat-specific enough to exhibit a significant relationship with remotely sensed data. Thus satellite imagery can provide one potential basis for deriving surrogates (see Gaston 1996) of species level biodiversity. However, as pointed out by Nagendra (2001), the modelling of species-RS relationships can include several pitfalls.

The inaccuracies of the prediction models highlight the need to be careful and to avoid applying the models rigidly and uncritically. Thus both good biogeographical and ecological knowledge of the predicted species and actual field checking are needed to evaluate the results of the predictive models in unknown terrain. In order to achieve complete assessment of the area concerned, landscape analysis and monitoring must be integrated with confirmatory field studies (Heikkinen 1998).

RS and GI data and techniques, if carefully applied, can also be used in monitoring short- or longer-term changes in different aspects of biodiversity and land cover (Stoms & Estes 1993; Johnston 1998; Nagendra & Gadgil 1999). For example, the conversion of forests to urban or intensively managed agricultural areas can be detected, and rates of change measured, by superimposing satellite images taken on different years (Iverson et al. 1989). Changes in habitat quality can be reflected by the changes in landscape element heterogeneity (Stoms & Estes 1993). For example, agricultural areas are usually characterised in remotely-sensed images by more regular shapes than natural landscapes.

Probably the best widely applicable option for developing appropriate RS-GI based monitoring of land cover and biodiversity changes is to focus on landscape analysis on the habitat level (Nagendra 2001), and if possible, to identify changes in the cover and distribution in the ecologically most important habitat types, such as old-growth forests (Stoms & Estes 1993; Mladenoff et al. 1994; Pakkala et al. 2002). From a more applied perspective, the detection and assessment of long-term trends in land use changes can help in the formation of policy in anticipation of the problems, e.g., loss of biodiversity, that result from the changes (Campbell 1996). However, it must be stressed that in order to develop truly successful RS-GI based monitoring programmes it is imperative to have intensive ground truth data available, which can be used in identifying landscape elements or habitat types on the basis of supervised classification (Nagendra & Gadgil 1999; Gould 2000; Roy & Tomar 2000). Other critical factors include errors in georeferencing, i.e. even minor differences in the placement of two separate maps derived from imagery acquired on different years, differences in the interpretation techniques, or spectral differences between imagery caused by clouds, haze, or other degrading factors (Campbell 1996, p. 576; Johnston 1998, p. 121–123).

Spatial autocorrelation

The lack of spatial independence in biogeographical data has typically been viewed as a problem that can obscure the researcher's ability to understand the geographical patterns being studied. Spatial autocorrelation examines the degree of synchrony between variables observed across geographic space and is important for a wide variety of geographical and ecological phenomena (Legendre 1993). Consequently, spatial autocorrelation is nowadays increasingly incorporated into biogeographical models and analyses based on spatial data (see Högmander & Møller 1995; Koenig & Knops 1998; Guisan & Zimmermann 2000; Henebry & Merchant 2002).

A variable is said to be autocorrelated if a measure made at one point supplies information on another measure of that variable recorded at a point located a given distance apart (Rossi & Queneherve 1998; Ferguson & Bester 2002). In this case the values are not independent in a statistical sense. If spatial autocorrelation is present, assessing the relationships between variables is complicated by the ineffectiveness of most classical statistical tools such as ANOVA or correlation analysis (Legendre 1993). The presence of common patterns between two or more variables may lead to spurious correlations, i.e. variables are apparently related, although in fact they only independently display a common spatial pattern. In such cases, it is necessary to examine the relationships between variables while controlling the effect of the common spatial structure.

Luoto et al. (2001) studied the occurrence pattern of the Clouded apollo butterfly (*Parnassius mnemosyne*) using a spatial grid system in southwestern Finland (Fig. 1). Spatial autocorrelation was statistically highly significant ($p < 0.001$) in the Clouded apollo distribution data and caused some problems in the interpretation of the modelling results. This was because the regression analysis showed clear differences between the explanatory capacity of predictive variables when the modelling procedure was performed with and without an adjusting spatial autocorrelation variable. In a model with no spatial autocorrelation variable, five environmental-topographical variables were included in the logistic regression model. However, when a spatial autocorrelation variable was entered into the model only three of the environmental-topographical variables remained statistically significant. In this example, it appears that the two excluded variables reflected mainly the spatial structure of the data, without any clear significant ecological relevance to the distribution of Clouded apollo (see Legendre 1993; Luoto et al. 2001).

Various methods have been devised for eliminating or avoiding the effects of spatial dependence in measuring or analysing geographical responses (Legendre 1993). For example, sampling of spatial data has typically been carried out by stratifying across space and averaging to infer underlying processes and mechanisms. Recently, however, biogeographers and spatial ecologists have begun to acknowledge that there is much important biology and ecology in the spatial dependence of biotic responses, and have become increasingly interested in examining spatial relationships directly. Whereas earlier research ignored or sought to remove the effects of spatial patterns of the data, the current approach is explicitly to analyse and model spatial patterns of the data as a fundamental feature of the study (Liebhold & Gurevitch 2002).

Most straightforwardly, spatial autocorrelation from the grid square *i* can be calculated in a spatial grid system as an average of the number of occupied grids among a set of eight neighbour grid squares of the square *i* (Augustin et al. 1996). The significance of spatial dependence of the data can be estimated by entering the spatial autocor-

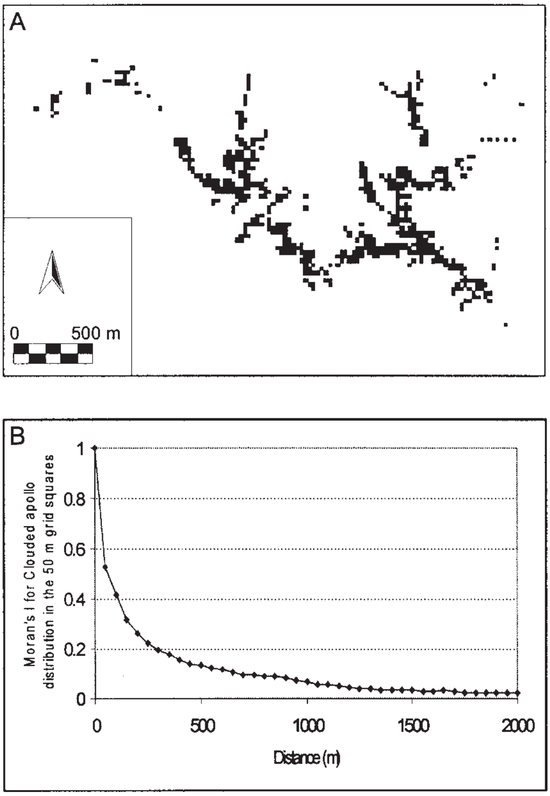

Fig. 1. (A) Distribution of the Clouded apollo (*Parnassius mnemosyne*) in the river Rekijoki area in southwestern Finland. (B) Spatial autocorrelation of the Clouded apollo observations, measured by Moran's I in relation to distance (see Legendre 1993; Brito et al. 1999).

relation variable as an additional explanatory variable in the final model. For more explicit methods see Koenig & Knops (1998) and Brito et al. (1999), in which various techniques to measure and analyse the spatial pattern of the data are described and reviewed.

Model building and verification

Several recent papers have criticized automatic stepwise procedures, as they do not necessarily select the most influential variables from a subset of variables (Bustamante 1997; Mac Nally 2000; Luoto et al. 2002a). Furthermore, the application of stepwise procedures in spatial data sets can give rise to statistically explicit, but ecologically irrelevant results. This may lead to models which agree closely with the observations in the study sites but which give poor predictions when extrapolated to unsurveyed areas (James & McCulloch 1990; Buckland & Elston 1993).

One pitfall in automatic stepwise model-building is the difficulty to produce ecologically and geographically plausible regression models, particularly when the number of candidate explanatory variables is large and the potential causal relationships between them and the response variable are not *a priori* well-known. Strong colinearity among the environmental variables may give rise to spurious regression models. In other words, the ecologically most important variables may well be excluded from the models when using automatic stepwise regression procedures ('statistically-focused modelling') (Flack & Chang 1987; Mac Nally 2000). Several recent papers argue that a more plausible regression model can be produced by the 'ecologically-focused' modelling approach, in which the biologically most important variables are forced to enter the model first or are given priority when selecting more or less equally important variables (Bustamante 1997; Mac Nally 2000).

This argument is supported, for example, by one modelling study of rare plant species richness in SW Finland (Luoto et al. 2002b). The overall fit of the ecologically-focused model developed in the study decreased clearly less (from 57.1% to 50.1%) when it was fitted to the test set of grid squares (i.e. a set of squares not used in developing the model), as compared with the corresponding decrease in the statistically-focused model (from 65.6% to 51.8%).

Another simple example can be considered: the study material includes topographically heterogeneous grid squares in a river valley and squares from gently sloping mountains some 300–500 metres higher. In this case the explanatory variables topographical heterogeneity and mean temperature (or some other energy-related factor) of a grid square would be intercorrelated. Most researchers would probably agree that mean temperature has a major impact on species richness in this example (see Currie 1991; Heikkinen 1998). However, it may well be excluded from a multiple regression model developed with typical automatic stepwise procedures due to colinearity, if simple topographical heterogeneity happens to have slightly better explanatory power in statistical terms. In this example it may be well justified to force more primary environmental variables to enter the model first, and only afterwards consider whether heterogeneity variables explain some further variation in species richness (cf. Begon et al. 1996).

Other examples where automatic stepwise modelling procedures may produce less desirable models are cases where climatic variables such as mean temperature or rainfall are highly correlated with altitude, latitude or longitude, particularly if the latter variables produce a somewhat better statistical fit. In such a case, it may be justified to select a biologically more meaningful variable first into the model, e.g., temperature instead of altitude (see Nicholls 1991; Bustamante 1997). These examples show that automatic regression model-building procedures can result in less causal relationships and consequently inaccurate predictions, and that the variable-selection process can be improved if the process is based on existing knowledge and theory (cf. Mac Nally 2000).

Several studies show that abiotic variables often have considerable statistical power, at least in the model building area. However, when the derived models are extrapolated to wider areas their predictive power can clearly decrease (Luoto et al. 2002a). Especially in extrapolative, predictive modelling, care should be taken to produce models that are ecologically more realistic than those derived from automatic stepwise regression procedures (Milsom et al. 2000, Mac Nally 2000). These ecologically-focused models may be less powerful than the statistically-focused models in model building, but can be still more appropriately applied over large areas with different topographic and landscape characteristics (Luoto et al. 2002a).

The importance of model verification is fundamental in predictive modelling (Boone and Krohn 1999). Not only should models be assessed with respect to their ability to explain observed variation, but they should also be validated. This can be done either using 'leave-one-out' jack-knife and bootstrap techniques (random sampling with replacement), or by evaluating the quality of the derived model by fitting it to an independent data set (the 'split-sample' or 'training-evaluation data sets' approach) (Guisan & Zimmermann 2000; Fleishman et al. 2001; Henebry & Merchant 2002; Suárez-Seoane 2002). Model predictions must be regarded as testable hypotheses. If the hypotheses are largely validated, then the model can be legitimately employed, for example for landscape management or conservation purposes (Fleishman et al. 2001). Moreover, the derived statistical models must also be tested for their ecological sensibility (Austin et al. 1990).

It is noteworthy that due to the dynamics and social factors affecting populations, not all suitable sites for a species are necessarily occupied at the same time. However, identification of unoccupied, but nevertheless suitable, sites using species-environment based modelling approaches can be highly important for long-term conservation planning. Johnson & Krohn (2002) gave examples of dynamically changing seabird colonies, for which carefully applied habitat occupancy models could be used in identifying features associated with suitable, but at a particular time unoccupied islands.

One additional problem in the biogeographical model building procedure is the spatial coverage of the model building area. The models should be based geographically and ecologically on an appropriate sample of the area, especially when they are used for spatial extrapolation. The models often produce somewhat inaccurate predictions, especially in those cases where the landscape pattern is different from that of the model building area (Luoto et al. 2002a).

Logistic regression analysis

The use of multivariate statistics to model biogeographical distribution patterns has increased in the past two decades and a wide variety of statistical techniques is now available (see Walker 1990; Mladenoff et al. 1995; Bustamante 1997; Brito et al. 1999). Probabilities of occurrences are generally assessed using the logistic regression methods. Logistic regression has been shown to be a powerful tool, capable of analysing the effects of one or several independent variables, discrete or continuous, over a dichotomic (presence/absence) variable (Pereira & Itami 1991; Augustin et al. 1996; Brito et al. 1999). Fitting a logistic regression model to distribution data is a straightforward task and algorithms are available in several statistical program packages.

Multiple logistic regression is an appropriate and widely used method for statistical analysis in different distribution problems in biological and ecological studies (see Pereira & Itami 1991; Carroll et al. 1999). However, logistic regression has not hitherto been very widely employed by geographers or landscape ecologists; rather it is preferred as a practical method for summarising species distributions along environmental gradients (see Peeters & Gardeniers 1998; Hill et al. 1999). A more technical and detailed review of logistic regression was presented by McCullagh & Nelder (1989) and Collett (1991). Logistic regression has the form:

$$\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}$$

where α is the constant and βx is the coefficient of the respective independent variables. The probability of presence π (ranging from 0 to 1) is given as a function of the vector of this model and becomes apparent after the logistic transformation, giving the form:

$$\ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \alpha + \beta x$$

where ln denotes the natural logarithm (Rita & Ranta 1993; Sokahl & Rohlf 1995).

In a model that attempts to explain the variation in distribution problems, the residuals cannot be normally distributed, as they should be in ordinary regression. This is because there are only two possible values for the response variable in data: 0 for absence and 1 for presence. Thus the statistical theory developed for ordinary regression models is not applicable to binomial distribution data. The use of ordinary regressions in probability analysis may lead to estimates without biological or even mathematical realism (Hosmer & Lemeshow 1989; Rita & Ranta 1993). In logistic regression, the binary nature of the response variable variation is the basis of parameter estimation and thus, the logistic regression models will not produce inappropriate values ($\pi(X) > 1$ or $\pi(X) < 0$) for the probability of presence (Rita & Ranta 1993).

As mentioned earlier, logistic regression – although the predominant method applied in species-environment modelling exercises – is not the only technique available for the modelling studies of species distribution patterns. Other statistical approaches include the following: Generalized additive models (GAM), environmental envelope techniques, Bayesian logistic-based modelling approach and neural networks. The discussion of these techniques is beyond the scope of this paper. However, information concerning these approaches can be found from, for example, Guisan & Zimmermann (2000), Mac Nally (2000), Fleishman et al. (2001) and Suárez-Seoane et al. (2002).

Other critical issues in biogeographical modelling

Comprehensive species distribution data over large areas and regions rarely exist. Frequently, the only data available for spatial modelling studies are herbarium records or museum specimens (Margules & Austin 1994; Austin 2002; Johnson & Sargeant 2002). However, these records have usually been collected in an opportunistic manner. This has resulted in incomplete and often biased data sets with regard to both the geographical and the taxonomical coverage (Margules & Austin 1994). Thus, regional data sets or atlases based on herbarium and other sources often provide only a limited basis for modelling exercises. Such presence-only data sets are hampered by false negatives – cells with no record of a species that really is present (Johnson & Sargeant 2002). There are empirical methods, such as BIOCLIM (Busby 1991), for estimating distribution patterns of species from presence-only types of species data. However, these methods will only provide an overall climatic envelope within which a species occurs, and will tell nothing about where it will be absent within the climatic limits of the envelope (Austin 2002). Thus modelling studies should preferably be based on true presence/absence records of species derived from geocoded plots of specified size.

On the other hand, comprehensive field surveys of species distribution patterns over wide areas are generally too expensive or logistically impossible to carry out. The best solution is to define cost-effective survey designs that will yield unbiased and sufficiently representative species distribution data sets. It is important to ensure that a survey samples the full range of vegetation composition and environmental space defined by the major environmental gradients in the region. In a similar vein, more accurate predictions of species occurrence patterns can generally be attained if the model-building grid squares are located all over the area, covering effectively all biotopes and environmental gradients. More information on the appropriate survey designs and the subsequent statistical modelling of species-environment relationships can be found from Walker (1990), Austin and Heyligers (1991), Margules & Austin (1994), Wessels et al. (1998) and Austin (2002).

When carrying out the actual modelling exercise, it is imperative to realise that the relationships between species and their environments are often nonlinear and should thus be modelled as such (Austin et al. 1990; Heglund 2002). One simple way of taking this into account is to incorporate squared terms of the predictor environmental variables into the modelling procedure (i.e. second order polynomial regressions; see Bustamante 1997; Guisan & Zimmermann 2000; Fleishman et al. 2001).

Problems of remote sensing data

Remote sensing provides an extensive source of relatively cheap, reliable data. However, the use of satellite images and digital aerial photographs in biogeographical and landscape ecological studies includes many potential pitfalls (Kalliola & Syrjänen 1991; Nagendra 2001). Ecologically and conservationally important habitat patches, such as deciduous forests and semi-natural grasslands, are often missed in satellite imagery classification (Stoms 1992; Luoto et al. 2002a). When Landsat-TM images are used, small habitat patches inevitably remain below the level of resolution, because only patches larger than one pixel (900 m^2) can be discriminated from the image. However, it is possible that even some larger patches are excluded from the classification due to the sensor properties or the patch shape, elongation or location in relation to the pixel boundaries (Hyppänen 1996; Fisher 1997; Cracknell 1998). The problem is even more pronounced in undulating topography with small-scale habitat pattern and often with corridor-like patches (cf. Guisan & Zimmermann 2000: 175). A small patch on a steep slope appears smaller than it really is and may therefore be indistinguishable (Lillesand & Kiefer 1994).

Another bias in the classification originates from the fact that the spectral reflectance of a pixel is influenced by the reflectance of its neighbourhood, caused by the movement of the sensor (Fisher 1997). Moreover, when using multispectral and multitemporal data, the blending between adjacent pixels is pronounced. This is a result of the fact that pixels of different bands of an image do not always overlap and pixels of images from different dates seldom overlap (Cracknell 1998). Fisher (1997) and Cracknell (1998) discussed the problem of rectangular spatial units, because pixels seldom match the true shape or size of natural objects. This is not a problem in large, homogenous areas such as coniferous forests or fields, but in the case of linear habitats, e.g., semi-natural grasslands or riverside forests, it undoubtedly affects the size and detection of patches (cf. Nagendra 2001).

When using satellite imagery as the source data for the habitat map, some uncertainties must be expected. It would be feasible, however, to improve the habitat classification by using aerial photographs or new high-resolution satellite imagery (e.g., IKONOS with 4 m resolution). The areas requiring more detailed data could be selected on the basis of topography and the fragmentation of habitats.

Scale

The problem of pattern and scale is one of the central problems in biogeography and spatial ecology. Biogeographical study problems require interfacing of phenomena that occur on very different scales of space, time and organization and therefore, there is no single natural scale on which geographical phenomenon should be studied (see Wiens 1989; Levin 1992; Stoms 1994). The observer imposes a perceptual bias, a filter through which the system is viewed. Furthermore, every organism is an 'observer' of the environment, and has its own perceptual spatial and temporal scale. This has fundamental significance for the study of biogeographical systems, since the distribution patterns and processes that are unique to any range of scales will have unique causes and ecological consequences (Levin 1992; Heglund 2002).

The pattern detected in any biogeographical mosaic is a function of scale, and the ecological concept of spatial scale encompasses both extent and grain (Turner et al. 1989; Wiens 1989; Forman 1995). Extent is the overall area encompassed by an investigation or the area included within the landscape boundary. Grain is the size of the individual units of observation. For example, a fine-grained map might structure information into 1 m^2 units, whereas a map with a coarser resolution would have information structured into 1 ha units (Turner et al. 1989).

Extent and grain define the upper and lower limits of resolution of study and any inferences about scale-dependence in a system are constrained by the extent and grain of investigation (Wiens 1989). From a statistical perspective, it is not reasonable to extrapolate beyond the population sampled or to infer differences between objects smaller than the experimental units. Similarly, in the assessment of landscape structure, it is not possible to detect pattern beyond the extent of the landscape or below the resolution of the grain (Wiens 1989).

As with the concept of landscape and patch, it may be ecologically more meaningful to define the scale from the perspective of the organism or ecological phenomenon under consideration. For example, from an organism-centred perspective, grain and extent may be defined as the degree of acuity of a stationary organism with respect to short- and long-range perceptual ability (Kolasa & Rollo 1991). Thus, grain is the finest component of the environment that can be differentiated close to the organism, whereas extent is the range at which a relevant object can be distinguished from a fixed vantage point by the organism.

It has been suggested that information can be transferred across scales if both grain and extent are specified (Allen et al. 1987; Kunin 1998). However, it is partially unclear how observed landscape patterns vary in response to changes in grain and extent, and whether landscape metrics obtained on different scales can be compared. The limited work on this topic suggests that landscape metrics vary in their sensitivity to changes in scale and that quantitative and qualitative changes in measurements across spatial scales will differ depending on how scale is defined (Turner et al. 1989). According to Wickham & Riitters (1995), identical classifications for the same area could be arrived at from sensors with different spatial resolving powers, and the resultant landscape metric values should not be dramatically affected by the difference in spatial resolution.

The key to modelling and understanding of biogeographical issues lies in elucidation of the mechanisms underlying the observed patterns (Wiens 1989; Noss 1992). The difficulties embedded in these attempts are pronounced in the studies using GI or RS data, because spatial data provide information between fine-scale ecological variation and large-scale geographical–spatial gradient, overlapping both. This can lead to the situation described by Levin (1992), where the mechanisms underlying the biogeographical patterns operate on different scales from those on which they are observed, producing rather poor fit of the models (see also Heglund 2002).

Recently, GI-based approaches have been used on different scales to analyse and model biogeographical distribution patterns (Kunin 1998). However, our understanding of the factors influencing on different scales is limited, and we lack the knowledge of how the different spatial modelling scales affect the performance of biogeographical distribution models. There is an increasing need to evaluate how the analysis and modelling results behave on different spatial scales.

Conclusions

The applicability and employment of spatial data derived from remote sensing and geographic databases to model and monitor biogeographical distribution problems has increased considerably in recent years. If remotely sensed data are to be used effectively for biogeographical research, techniques to integrate observations of landscape pattern and habitat quality with data on biogeographical distribution patterns need to be developed further.

There are many national vegetation and land cover maps available with information on potential sites of certain species or of high biodiversity. Remote sensing and geographic information systems are uniquely poised to use these data, in conjunction with spatial analysis and modelling, to map and monitor species distribution and biodiversity patterns. Furthermore, predictive RS and GI-based modelling can provide a basis for focusing field assessment and allocating conservation resources in areas where the distribution of species is not well known (Gould 2000; Luoto 2000b).

Biogeographers and landscape ecologists typically view landscape as a mosaic of land cover elements (habitats, biotopes and ecosystems) and believe that their spatial arrangement controls or affects the ecological processes operating within them. A more holistic perspective in landscape studies, which also takes into account the geomorphological, hydrological and climatological aspects of the landscape, is needed for a comprehensive analysis and modelling of a certain area.

ACKNOWLEDGEMENTS

We thank R. Kalliola for valuable comments on an earlier draft of the manuscript. M. Bailey improved the English of the manuscript. This study is part of the project ‘Use of remote sensing in modelling biodiversity in agricultural landscapes’ (SA80784) funded by the Academy of Finland.

REFERENCES

Allen TFH, RVO O’Neill & TW Hoekstra (1987). Interlevel relations in ecological research and management: some working principles from hierarchy theory. *Journal of Applied System Analysis* 14, 63–79.

Augustin N, MA Mugglestone & ST Buckland (1996). An autologistic model for spatial distribution of wildlife. *Journal of Applied Ecology* 33, 339–347.

Austin MP (2002). Case studies of the use of environmental gradients in vegetation and fauna modelling: theory and practice in Australia and New Zealand. In Scott JM, PJ Heglund, ML Morrison, JB Haufler, MG Raphael, WA Wall & FB Samson (eds). *Predicting species occurrences. Issues of accuracy and scale*, 73–82. Island Press, Washington.

Austin MP & PC Heyligers (1991). New approach to vegetation survey design: gradsect sampling. In Margules CR & MP Austin (eds). *Nature Conservation: Cost effective biological surveys and data analysis,* 31–36. CSIRO, Australia.

Austin MP, AO Nicholls & CR Margules (1990). Measurement of the realized qualitative niche: environmental niches of five Eucalyptus species. *Ecological Monographs* 60, 161–177.

Begon M, JL Harper & CR Townsend (1996). *Ecology*. 3rd ed. 1068 p Blackwell Science, Oxford.

Boone RB & WB Krohn. (1999). Modeling the occurrence of bird species: Are the errors predictable? *Ecological Applications* 9, 835–848.

Brito JC, EG Crespo & OS Paulo (1999). Modelling wildlife distributions: logistic multiple regression vs overlap analysis. *Ecography* 22, 251–260.

Buckland ST & DA Elston (1993). Empirical models for spatial distribution of wildlife. *Journal of Applied Ecology* 30, 478–495.

Busby JR (1991). BIOCLIM – A bioclimate analysis and prediction system. In Margules CR & MP Austin (eds). *Nature conservation: Cost effective biological surveys and data analysis,* 64–68. CSIRO, Australia.

Bustamante J (1997). Predictive models for Lesser Kestrel *Falco naumanni* distribution, abundance and extinction in southern Spain. *Biological Conservation* 80, 153–160.

Campbell JB (1996). *Introduction to remote sensing.* 2nd ed. 622 p. Taylor & Francis, London.

Carroll C, WJ Zielinski & RF Noss (1999). Using presence-absence data to build and test spatial habitat models for the fisher in the Klamath Region, U.S.A. *Conservation Biology* 13, 1344–1359.

Cherril AJ, C McClean, P Watson, K Tucker, SP Rushton & R Sanderson (1995). Predicting the distributions of plant species at the regional scale: hierarchical matrix model. *Landscape Ecology* 4, 197–207.

Clements FE (1936). Nature and structure of the climax. *Journal of Ecology* 24, 252–284.

Collett D (1991). *Modelling binary data.* 369 p. Chapman & Hall, London.

Cracknell AP (1998). Synergy in remote sensing – What's in a pixel? *International Journal of Remote Sensing* 19, 2025–2047.

Currie DJ (1991). Energy and large-scale patterns of animal- and plant-species richness. *American Naturalist* 137, 27–49.

Debinski DM & P Humphrey (1997). An integrated approach to biological diversity assessment. *Natural Areas Journal* 17, 355–365.

Debinski DM, K Kindscher & ME Jakubauskas (1999). A remote sensing and GIS-based model of habitats and biodiversity in the Greater Yellowstone Ecosystem. *International Journal of Remote Sensing* 17, 3281–3291.

Enoksson B, P Angelstam & K Larsson (1995). Deciduous forest and resident birds: the problem of fragmentation within a coniferous forest landscape. *Landscape Ecology* 10, 267–275.

Fahrig L (2002). Effect of habitat fragmentation on the extinction threshold: a synthesis. *Ecological Applications* 12, 346–353.

Ferguson JWH & MN Bester (2002). The treatment of spatial autocorrelation in biological surveys: the case of line transect surveys. *Antarctic Science* 14, 115–122.

Fisher P (1997). The pixel: a snare and a delusion. *International Journal of Remote Sensing* 18, 679–685.

Flack VF & PC Chang 1987. Frequency of selecting noise variables in subset regression analysis: a simulation study. *The American Statistician* 41, 84–86.

Fleishman E, R Mac Nally, JP Fay & DD Murphy (2001). Modelling and predicting species occurrence using broad-scale environmental variables: an example with butterflies of the Great Basin. *Conservation Biology* 15, 1674–1685.

Forman RTT (1995). *Land mosaics. The Ecology of Landscapes and Regions.* 632 p. Cambridge University Press, Cambridge.

Gaston KJ (1996). Species richness: measure and measurement. In KJ Gaston (ed). *Biodiversity. A Biology of numbers and difference,* 77–113. Blackwell Science, Oxford.

Gould WA (2000). Remote sensing of vegetation, plant species richness, and regional diversity hotspots. *Ecological Applications* 10, 1861–1870.

Granö O & A Paasi (eds) (1997). *J.G. Granö: Pure Geography.* 192 p. Johns Hopkins University Press, Baltimore.

Griffiths GH, JM Smith, N Veitch & R Aspinall (1993). The ecological interpretation of satellite imagery with special reference to bird habitats. In Haines-Young R, DR Green & SH Cousins (eds). *Landscape Ecology and geographic information systems,* 255–272. Taylor & Francis Ltd, London.

Guisan A & NE Zimmermann (2000). Predictive habitat distribution models in ecology. *Ecological Modelling* 135, 147–186.

Haila Y (2002). A conceptual genealogy of fragmentation research: from island biogeography to landscape ecology. *Ecological Applications* 12, 321–334.

Hanski I (1998). Metapopulation dynamics. *Nature* 396, 41–49.

Hanski I (1999). *Metapopulation Ecology.* 324 p. Oxford Series in Ecology and Evolution. Oxford University Press, Oxford.

Hanski I & ME Gilpin (eds) (1997). *Metapopulation biology: ecology, genetics, and evolution.* Academic Press, San Diego.

Heglund PJ (2002) Foundations of species-environment relations. In Scott JM, PJ Heglund, ML Morrison, JB Haufler, MG Raphael, WA Wall & FB Samson (eds). *Predicting species occurrences. Issues of accuracy and scale,* 35–41. Island Press, Washington.

Heikkinen RK (1998). Can richness patterns of rarities be predicted from mesoscale atlas data? A case study of vascular plants in the Kevo Reserve. *Biological Conservation* 83, 133–143.

Henebry GM & JW Merchant (2002). Geospatial data in time: limits and prospects for predicting species occurrences. In Scott JM, PJ Heglund, ML Morrison, JB Haufler, MG Raphael, WA Wall & FB Samson (eds). *Predicting species occurrenc-*

es. Issues of accuracy and scale, 291–302. Island Press, Washington.

Hill JK, CD Thomas & B Huntley (1999). Climate and habitat availability determine 20th century changes in a butterfly's range margin. *Proceedings of Royal Society London.* B. 266, 1197–1206.

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

Hosmer DW Jr & S Lemeshow (1989) *Applied Logistic Regression.* John Wiley & Sons Inc., New York.

von Humboldt A (1807). *Ideen zu einer Geographie der Pflanzen nebst einem Naturgemälde der Tropenländer.* 182 p. Cotta, Tübingen.

Huxel GR & A Hastings (1999). Habitat loss, fragmentation and restoration. *Restoration Ecology* 7, 309–315.

Hyppänen H (1996). Spatial autocorrelation and optimal spatial resolution of optical remote sensing data in boreal forest environment. *International Journal of Remote Sensing* 17, 3441–52

Iverson LR, RL Graham & EA Cook (1989). Applications of satellite remote sensing to forested ecosystems. *Landscape Ecology* 3, 131–143.

James FC & CE McCulloch (1990). Multivariate analysis in ecology and systematics: Panacea or Pandora's box? *Annual Review of Ecology and Systematics* 21, 129–166.

Johnson CM & WB Krohn (2002). Dynamic patterns of associations between environmental factors and island use by breeding seabirds. In Scott JM, PJ Heglund, ML Morrison, JB Haufler, MG Raphael, WA Wall & FB Samson (eds). Predicting species occurrences. Issues of accuracy and scale, 171–181. Island Press, Washington.

Johnson DH & GA Sargeant (2002). Toward better atlases: improving presence-absence information. In Scott JM, PJ Heglund, ML Morrison, JB Haufler, MG Raphael, WA Wall & FB Samson (eds). *Predicting species occurrences. Issues of accuracy and scale,* 391–397. Island Press, Washington.

Johnston CA (1998). *Geographic information systems in ecology.* 239 p. Blackwell Science, Oxford.

Kalliola R & K Syrjänen (1991). To what extent are vegetation types visible in satellite imagery. *Annales Botanici Fennici* 28, 45–57.

Koenig WD & JMH Knops (1998). Testing for spatial autocorrelation in ecological studies. *Ecography* 21, 423–429.

Kolasa J & CD Rollo (1991). Introduction: The heterogeneity of heterogeneity: a glossary. In: Kolasa J & STA Pickett (eds). *Ecological heterogeneity,* 1–23. Springer-Verlag, New York.

Kunin WE (1998). Extrapolating species abundance across spatial scales. *Science* 281, 1513–1515.

Legendre P (1993). Spatial autocorrelation: trouble or new paradigm? *Ecology* 74, 1055–1067.

Levins R (1969). Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of Entomological Society of America* 15, 237–240.

Levin SA (1992). The problem of pattern and scale in ecology. *Ecology* 73, 1943–1967.

Liebhold AM & J Gurevitch (2002). Integrating the statistical analysis of spatial data in ecology. *Ecography* 25, 553–557.

Lillesand TM & RW Kiefer (1994). *Remote Sensing and Image Interpretation.* 650 p. John Wiley & Sons. Inc., USA.

Luoto M (2000a). Landscape ecological analysis and modelling of habitat and species diversity in agricultural landscapes using GIS (PhD thesis). *Annales Universitatis Turkuensis* A II, 141p.

Luoto M (2000b). Modelling of rare plant species richness by landscape variables in an agriculture area in Finland. *Plant Ecology* 149, 157–168.

Luoto M, M Kuussaari, H Rita, J Salminen & T von Bonsdorff (2001). Determinants of distribution and abundance in the Clouded apollo butterfly: a landscape ecological approach. *Ecography* 24, 601–617.

Luoto M, M Kuussaari & T Toivonen (2002a) Modelling butterfly distribution based on remote sensing data. *Journal of Biogeography* 29, 1027–1037.

Luoto M, T Toivonen & RK Heikkinen (2002b) Prediction of total and rare plant species richness from satellite images and GIS data in agricultural landscapes. *Landscape Ecology* 17, 195–217.

MacArthur RH & EO Wilson (1967). T*he theory of island biogeography.* 224 p. Princeton University Press.

Mac Nally R (2000). Regression and model-building in conservation biology, biogeography and ecology: The distinction between – and reconciliation of – 'predictive' and 'explanatory' models. *Biodiversity and Conservation* 9, 655–671.

Margules CR & MP Austin (1991). *Nature Conservation: Cost effective biological surveys and data analysis.* 207 p. CSIRO, Australia.

Margules CR & MP Austin (1994). Biological models for monitoring species decline: the construction and use of data bases. *Philosophical Transactions of the Royal Society, London* B 344, 69–75.

McCullagh P & JA Nelder (1989). *Generalized Linear Models.* 472 p. Chapman & Hall, New York.

Milsom TP, SD Langton, WK Parkin, S Peel, JD Bishop, JD Hart & NP Moore (2000). Habitat models of bird species' distribution: an aid to the management of coastal grazing marshes. *Journal of Applied Ecology* 37, 706–727.

Mladenoff DJ, MA White, TR Crow & J Pastor (1994). Applying principles of landscape design and management to integrate old-growth forest enhancement and commodity use. *Conservation Biology* 8, 752–762.

Mladenoff DJ, TA Sickley, RG Haight & AP Wydeven (1995). A regional landscape analysis and prediction of favorable gray wolf habitat in the northern Great Lakes region. *Conservation Biology* 9, 279–294.

Mooney A & M Godron (1983). *Disturbance and Ecosystems: Components of Response.* Springer-Verlag, New York.

Nagendra H (2001). Using remote sensing to assess biodiversity. *International Journal of Remote Sensing* 22, 2377–2400.

Nagendra H & M Gadgil (1999). Satellite imagery as a tool for monitoring species diversity: an assessment. *Journal of Applied Ecology* 36, 388–397.

Naveh Z & AS Lieberman (1984). *Landscape Ecology, Theory and Application.* Springer Verlag, New York.

Nicholls AO (1991). Examples of the use of Generalized linear models in analysis of survey data for conservation evaluation. In Margules CR & MP Austin (eds). *Nature Conservation: Cost effective biological surveys and data analysis,* 54–63. CSIRO, Australia.

Noss RF (1992). Issues of scale in conservation biology. In Fieldler PL & SK Jain (eds). *Conservation biology,* 239–250. Chapman & Hall, New York.

Noss RF (2001). Forest fragmentation in the Southern Rocky Mountains. *Landscape Ecology* 16, 371–372.

Pakkala T, I Hanski & E Tomppo (2002). Spatial ecology of the three-toed woodpecker in managed forest landscapes. *Silva Fennica* 36, 279–288.

Peeters ETHM & JJP Gardeniers (1998). Logistic regression as a tool for defining habitat requirements of two common gammarids. *Freshwater Biology* 39, 605–615.

Pereira JM & RM Itami (1991). GIS-based habitat modeling using logistic multiple regression: a study of the Mt. Graham red squirrel. *Photogrammetric Engineering and Remote Sensing* 57, 1475–1486.

Perry JN, AM Liebholt, MS Rosenberg, J Dungan, M Miriti, A Jakomulska & S Citron-Pousy (2002). Illustrations and guidelines for selecting statistical methods for quantifying spatial patterns in ecological data. *Ecography* 25, 578–600.

Rita H & H Ranta (1993). On analysing species incidence. Annales Zoologici Fennici 30, 173–176.

Rossi J-P & P Queneherve (1998). Relating species density to environmental variables in presence of spatial autocorrelation: a study on soil nematodes distribution. *Ecography* 21, 117–123.

Roy PS & S Tomar (2000). Biodiversity characterization at landscape level using geospatial modelling technique. *Biological Conservation* 95, 95–109.

Schmielgelow FKA & M Mönkkönen (2002). Habitat loss and fragmentation in dynamic landscapes: avian perspectives from the boreal forest. *Ecological Applications* 12, 375–389.

Scott JM, F Davis, B Csuti, R Noss, B Butterfield, C Groves, H Anderson, S Caicco, F D'Erchia, TC Edwards Jr, J Ulliman & RG Wright (1993). Gap analysis: a geographic approach to protection of biological diversity. *Wildlife Monographs* 57, 1–41.

Sokal RR & FJ Rohlf (1995). *Biometry*. 887 p. W. H. Freeman, New York.

Stoms DM (1992). Effects of Habitat Map Generalization in Biodiversity Assessment. *Photogrammetric Engineering and Remote Sensing* 58, 1587–1591.

Stoms DM (1994). Scale dependence of species richness maps. *Professional Geographer* 46, 346–358.

Stoms DM & JE Estes (1993). A remote sensing research agenda for mapping and monitoring biodiversity. *International Journal of Remote Sensing* 10, 1839–1860.

Suárez-Seoane S, PE Osborne & JC Alonso (2002). Large-scale habitat selection by agricultural steppe birds in Spain: identifying species-habitat responses using generalized additive models. *Journal of Applied Ecology* 39, 755–771.

Thomas CD, JA Thomas & MS Warren (1992). Distributions of occupied and vacant butterfly habitats in fragmented landscapes. *Oecologia* 92, 563–567.

Tilman D & P Kareiva (1997). *Spatial ecology.* Princeton University Press.

Tilman D, RM May, CL Lehman & MA Nowak (1994). Habitat destruction and the extinction debt. *Nature* 371, 65–66.

Turner MG (1989). Landscape Ecology: The Effect of Pattern on Process. *Annual Review of Ecology and Systematics* 20, 171–197.

Turner MG, RVO O'Neill, RH Gardner & BT Milne (1989). Effects of changing spatial scale on the analysis of landscape pattern. *Landscape Ecology* 3, 153–162.

Verboom J, A Schotman, P Opdam & J Metz (1991). European nuthatch metapopulations in a fragmented agricultural landscape. *Oikos* 61, 149–156.

Walker PA (1990). Modelling wildlife distribution using a geographic information system: kangaroos in relation to climate. *Journal of Biogeography* 17, 279–289.

Watson DM (2002). A conceptual framework for studying species composition in fragments, islands and other patchy environments. *Journal of Biogeography* 29, 823–834.

Watt AS (1947). Pattern and process in the plant community. *Journal of Ecology* 35, 1–22.

Wessels KJ, AS van Jaarsveld, JD Grimbeek & MJ van der Linde (1998). An evaluation of the gradsect biological survey method. *Biodiversity and Conservation* 7, 1093–1121.

Wickham JD & KH Riitters (1995). Sensitivity of landscape metrics to pixel size. *International Journal of Remote Sensing* 16, 3585–3594.

Wiens JA (1989). Spatial scaling in ecology. *Functional Ecology* 3, 385–397.

Wiens JA (1997). Metapopulation Dynamics and Landscape Ecology. In I Hanski & ME Gilpin (eds). *Metapopulation biology: ecology, genetics, and evolution,* 43–62. Academic Press, London.

Whittaker RH (1967). Gradient analysis of vegetation. *Biological Review* 49, 207–264.