The Potentiality of Remote Sensing in Biogeographical Research

Miska Luoto
Thule Institute, University of Oulu

Abstract: Spatial variability in species' geographical patterns is a crucial phenomenon and has formed the basis for numerous biogeographical studies. At present, human activities increasingly affect the terrestrial biosphere, resulting in habitat loss and degradation. Due to the rapidity of this process there is utmost need to detect and predict changes in the natural environment and assess the spatial distribution of valuable sites and habitats. Remote sensing (RS) produces valuable information for species mapping and monitoring. Due to their high spatial and temporal coverage, RS data have clear advantages compared to any other source of information. However, the benefits of remote sensing have not been fully utilized in biogeographical studies, and several methodological issues remain insufficiently explored. This paper presents novel approaches and techniques to integrate RS and geographic information (GI) data into biogeographical research. The issues discussed in this paper can have relevance in several fields of application of geographical data.

Introduction

Which factors explain the distribution and the abundance of a species has continued to be a central question in biogeographical research since the pioneering work of Andrewartha & Birch (1954). Nowadays, human activities increasingly affect the terrestrial biosphere, resulting in habitat loss and degradation which ultimately impair ecosystem function and ecosystem services (Matson & al. 1997). Due to the rapidity of this process, there is utmost need to detect and predict changes in the natural environment and assess the spatial distribution of valuable sites and habitats.

Current European and national environmental regulations often require assessments of the ecological effects of land use planning (Luoto & al. 2002b). Not only detailed knowledge of the present patterns of species richness and distributions (hereafter biodiversity), but also accurate spatial predictions for more poorly known regions are needed. The challenge in conservation and land use planning is that biodiversity data, as accurate as possible, is required within a limited amount of time and over considerably large areas. Unfortunately, such data are often sparse and expensive to acquire by traditional field inventories.

Alongside with biogeographers and landscape ecologists, environmental researchers, planners and policy-makers are increasingly turning to the discipline of remote sensing (RS) and geographic information (GI) science. RS-GI science is considered to be capable of providing the required techniques and data for developing cost-effective assessments, models and projections of biodiversity responses to environmental change over large areas. The expectations on RS-GI techniques have been boosted by the growing selection of remotely sensed data (often freely accessible) sources. At the moment, data archives provide increasing quantities of multi-temporal and multi-spectral remotely sensed imagery (Luoto & al. 2002b).

Remote sensing generates a remarkable array of biogeographically valuable measurements as well as the capacity to detect natural and human-induced land cover changes. Used in an integrative mode, RS-GI data can provide information about both historical and current land cover factors affecting biodiversity patterns. Relationships between species distributions and remotely sensed data, particularly in conjunction with topographic, soil and bedrock information, can be used to predict the distribution of species over large areas.

Remotely sensed imagery exposes land cover changes at spatial scales from local to continental, allowing one to monitor the pace of habitat conversion. These measurements of habitat loss can be used in providing quantitative estimates of biodiversity loss through the use of species-habitat modelling. The challenge is to go one step further and provide a more detailed understanding of which species are being lost and why. How can we best match existing and emerging RS and GI technologies to parameters that have clear implications for organisms and ecosystems?

Hitherto, the benefits of remote sensing have not been fully utilized in biodiversity studies. Moreover, there are several critical issues in applying RS-GI based biodiversity models which have been insufficiently explored. These include the applicability of different RS data sets in biodiversity assessments, and the effect of scale, statistical techniques and model complexity on RS-GI based biodiversity modelling. In this article, I present novel approaches and techniques to integrate RS and geographic information (GI) data into biogeographical research. I wish that the issues discussed in this paper can have relevance in several fields of application of spatial data in physical geography and related environmental sciences.

The potential of remote sensing for assessing biodiversity

Current approaches in using RS in biodiversity modelling can be categorized into two main types: 1) habitat mapping using RS data, and predictions of species distribution based on habitat requirements; and 2) direct relationships between spectral radiance values recorded from remote sensors and species' distributions recorded in field surveys. These approaches require the use of information collected from the field, as well as from the remote sensor. It is essential to establish the strength of the linkages between these different kinds of information collected on different scales. For example, the degree of correspondence between habitat maps and species distributions depends on the degree of habitat map generalization. Such relationships should be optimized to get maximum information on biodiversity. In comparison, a major limitation of spectral values-biodiversity models is that the relationship between spectral reflectance and species distributions depends not only on the species in question, but also the area. In practice, little guidance exists for investigators concerning the suitability of these methods for biodiversity assessments in different habitats. Thus, there is an utmost need to evaluate the applicability of these two main types of RS information in modelling biodiversity on different spatial scales in different landscape mosaics. I propose that such comparative work is greatly needed to produce guidelines for choosing optimal approaches in RS-GIS based biogeographical modelling.

The effect of scale on the biogeographical models

The problems and impacts of scale have long been a central issue in biogeographical (Rahbek & Graves 2001; Blackburn & Gaston 2002), geomorphological (Walsh & al. 1998; Luoto & Hjort 2006) and landscape ecological studies (Wiens 1989; Levin 1992). The spatial scale on which species distribution modelling is undertaken is of fundamental importance for the results and inferences of biogeographical studies. The concept of spatial scale consists of two important attributes: the unit of sampling and the geographical space covered. The first attribute is defined by ‘grain’ (or ‘resolution’) and ‘focus’, grain being the size of the common analytical unit and focus the area represented by each data point. The second attribute is ‘extent’, describing the geographical area over which comparisons are made (Scheiner 2003; Rahbek & Graves 2001).

Most importantly, the choice of spatial resolution in modelling exercises can directly affect the ecological interpretations derived from the results and the usability of RS-GI based models. If the response of a given species to environmental factors shows a high scale-sensitivity, comparative studies over a range of scales can indicate the resolutions at which the species-environment patterns are most accurately modelled. However, we lack the knowledge of how the extent of habitat information and resolution of input data affect the performance of RS-GI based habitat-relationship models on different spatial scales.

Researchers have come to different conclusions about the usefulness of habitat-relationship models for predicting species presence or absence. This difference very likely stems from the failure to recognize the effects of spatial scale. In order to increase the plausibility of the species distribution models, the analysis should be conducted on the scale on which the target species most strongly responds to the environmental variation (Wiens 1989; Carroll & al. 1999). Identification of the most suitable resolution for species distribution modelling often requires studies conducted over a range of spatial scales. Krawchuk & Taylor (2003) argued that studies examining species response to the environment performed at a single scale may result in limited conclusions and potentially harmful management recommendations.

The effect of modelling techniques and complexity on RS-GI based models

Recently, species distribution modelling has become one of the key issues in biogeography

(Luoto & al. 2002a; Thuiller 2003), conservation planning (Ferrier & al. 2002; Lehmann & al. 2002) as well as in landscape ecology (Fleishman & al. 2001; Luoto & al. 2002b; Lawler & al. 2004). This development is based on two trends: growth in the availability of remotely sensed (RS) data and development of GI techniques (Gould 2000; Nagendra 2001; Kerr & Ostrovsky 2003), and the development of novel statistical techniques that can be applied to a wide range of ecological systems (Guisan & Zimmermann 2000; Thuiller 2003).

RS-GI based biodiversity modelling studies have usually been conducted by employing only one modelling approach. Obvious variation is expected from using different techniques because different models use a variety of assumptions, algorithms and parameterizations. Thus, when studies use a single modelling technique there is no information of whether the selected method provides the best predictive accuracy for the particular data set used. Only a few attempts have been made to evaluate the ability of different methods to identify plausible and accurate species-environment responses in complex RS-GI multiple regression settings including many potentially important explanatory variables (Guisan & Zimmermann 2000; Thuiller 2003). Consequently, insufficient guidance exists for investigators concerning the suitability of different modelling approaches for species distribution modelling and other biodiversity assessments.

Currently, there are several different modelling approaches available for biogeographical modelling, ranging from generalized linear models (GLM) and generalized additive models (GAM) (Bustamante 1997; Seoane & al. 2003; Luoto & al. 2004), to rule-based techniques (artificial neural networks; ANN and classification tree analysis; CTA) (Moisen & Frescino 2002; Olden & al. 2004), and novel regression techniques, for example multivariate adaptive regression splines (MARS) (Venables & Ripley 2002). Generalized linear models are mathematical extensions of linear models which can handle non-linear relationships and different types of statistical error distributions such as Gaussian, Poisson, Binomial or Gamma (Crawley 1993). Generalized additive models are flexible data-driven nonparametric extensions of generalized linear models (Hastie & Tibshirani 1990) that allow both linear and complex additive response curves to be fitted. Classification tree analysis is a rule-based method which has rarely been applied in RS-GI based modelling studies. CTA does not rely on a priori hypotheses about the relationships between predictor and response variables. The method generates a binary tree through binary recursive partitioning, a process that splits a node based on true/false answers about the values of predictors (Venables & Ripley 2002). The rule generated at each step maximizes the class purity within each of the two resulting subsets. Artificial neural networks are complex non-parametric techniques that have been shown to provide highly flexible function approximates for any data (Lek & Guegan 1999). Neural networks can provide a means to build accurate models when the functional form of the underlying equations is unknown (Ripley 1996). Multiple adaptive regression splines is a relatively new technique that combines classical linear regression, mathematical construction of splines and binary recursive partitioning to produce a local model in which relationships between responses and predictors are either linear or non-linear

(Friedman 1991). MARS has been advocated as an effective method to model complex spatial data that can include a relatively high number of variables.

My earlier results (Luoto & Hjort 2005) as well as the studies by Moisen and Frescino (2002), Thuiller (2003) and Segurado and Araújo (2004) suggest that the use of novel non-parametric techniques often results in a slightly better model performance than the parametric approaches. Non-parametric modelling techniques are powerful tools with limited statistical theory to support them, and investigators should be aware of their potential shortcomings. Non-parametric techniques have been criticized because they can produce a complex set of functions, whereas a corresponding parametric function may capture most of the same variation (Venables & Ripley 2002).

Downscaling species atlas distributions

One of the limitations in using biogeographical data sets, typically species' distribution atlases, in regional planning is their coarse resolution (typically from 10-km to 50-km) relative to the requirements of conservation and land use planning. In response to this, attempts have been made to downscale species distributions to finer resolutions (McPherson & al. 2006). The term 'downscaling' describes a procedure in which information about a process with a certain characteristic scale is derived from other processes with larger scales. The general limitations, theory and practise of downscaling methods are well described in the literature (McPherson & al. 2006). However, only little empirical work has been conducted in geography to use coarse-grained data in order to model biogeographical processes at finer spatial resolution, however see exceptions in Araújo & al. (2005) for species distribution modelling in U.K. Systematic studies employing downscaling methods for several different biogeographical processes utilizing independent model calibration and model test areas are largely lacking. By and large, the potentiality of statistical downscaling of the distribution of flora and fauna has been insufficiently investigated.

Downscaling has been done by combining remotely sensed data and expert opinion to assign species to habitats or land-cover classes considered suitable for species. One of the problems with this approach is that there may be insufficient expert knowledge of species-habitat relationships for many species and areas. An alternative is to use empirical modelling techniques that explore the correlation between species distribution and sets of predictor variables to downscale distributions of species to finer spatial resolution. One crucial assumption in this approach is that the main drivers of species distribution at coarse resolution are also important determinants of distributions at finer resolutions.

Recent studies have demonstrated significant variability in downscaled model predictions (Araújo & al. 2005; McPherson & al. 2006). These results highlight the need for further research to test and improve the approach so as to increase confidence in model predictions. Statistical downscaling has important limitations, and departures from model assumptions are likely to vary for different species and regions. In particular, it would be important to investigate which species are poorly modelled, where and why. Without such an assessment it is difficult thoroughly to assess the usability and accuracy of downscaled models in real-world situations.

Conclusions

Remote sensing and GI techniques provide an approximate cost-efficient method to estimate the biodiversity status of wide areas on a broad scale. I put forward that the RS-GIS based modelling approaches deserve further attention in biogeography and applied land use planning. However, there are also many potential shortcomings in using satellite imagery and other kinds of spatial data as a surrogate for species richness – how to appropriately collect source data for modelling, technical problems in processing satellite imagery, methodological pitfalls in regression models, etc. The integration of biogeographical modelling, RS and GIS technology requires often transdisciplinary skills between geography, ecology, statistics and geoinformatics. Thus the pitfalls for the misuse and abuse of the GIS technology with high calculation capacity are very obvious. We should be quite cautious and not apply the results of predictive models and extrapolations too uncritically, and not without a good geographical knowledge of the study area. In sum, broad-based understanding of geoinformatics and related issues is needed by spatial data producers and modellers.

References

Andrewartha, H.G. & L.C. Birch (1954). *The distribution and abundance of animals*. The University of Chicago Press.

Araújo, M.B., W. Thuiller, P.H. Williams & I. Reginster (2005). Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. *Global Ecology and Biogeography*, vol.14, 17-30.

Blackburn, T. & K. Gaston (2002). Scale in macroecology. *Global Ecology & Biogeography*, vol. 11, 185-189.

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

Carroll, C., W.J. Zielinski & R.F. 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*, vol. 13, 1344-1359.

Crawley, M.J. (1993). *GLIM for ecologists*. Blackwell Scientific Publications, Oxford.

Ferrier, S., G. Watson, J. Pearce & M. Drielsma (2002). Extended statistical approaches to modeling spatial pattern in biodiversity: the north-east New South Wales experience. I. Species-level modeling. *Biodiversity and Conservation*, vol. 11, 2275-2307.

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

Friedman, J. (1991). Multivariate adaptive regression splines. Annals of Statistics, vol. 19, 1-141.

Gould, W.A. (2000). Remote sensing of vegetation, plant species richness, and regional diversity hotspots. *Ecological Applications*, vol. 10, 1861-1870.

Guisan, A. & N.E. Zimmermann (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, vol. 135, 147-186.

Hastie, T. & R. Tibshirani (1990). *Generalized additive models*. Chapman and Hall, London.

Kerr, J. & M. Ostrovsky (2003). From space to species: ecological applications for remote sensing. *Trends in Ecology and Evolution*, vol. 18, 299-305.

Krawchuk, M. & P. Taylor (2003). Changing importance of habitat structure across multiple spatial scales for three species of insects. *Oikos*, vol. 103:1, 153-161.

Lawler, J., R. O'Connor, C. Hunsaker, K. Jones, T. Loveland & D. White (2004). The effects of habitat resolution on models of avian diversity and distributions: a comparison of two land-cover classifications. *Landscape Ecology*, vol. 19, 515-530.

Lehmann, A., J. Overton & J. Leathwick (2002). GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling*, vol. 157, 189-207.

Lek, S. & J. Guegan (1999). Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling*, vol. 120, 65-73.

Levin, S.A. (1992). The problem of pattern and scale in ecology. *Ecology*, vol. 73, 1943-1967.

Luoto, M. & J. Hjort (2005). Evaluation of current statistical approaches for predictive geomorphological mapping. *Geomorphology*, vol. 67, 299-315.

Luoto, M. & J. Hjort (2006). Scale matters – a multi-resolution study of the determinants of patterned ground activity in subarctic Finland. *Geomorphology*, vol. 80, 282-294.

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

Luoto, M., T. Toivonen & R.K. Heikkinen (2002b). Prediction of total and rare plant species richness in agricultural landscapes from satellite images and topographic data. *Landscape Ecology*, vol. 17: 3, 195-217.

Luoto, M., R. Virkkala, R.K. Heikkinen & K. Rainio (2004). Predicting bird species richness using remote sensing in boreal agricultural-forest mosaic. *Ecological Applications*, vol. 14, 1946-1962.

Matson, P.A., W.J. Parton, A.G. Power & M.J. Swift (1997). Agricultural intensification and ecosystem properties. *Science*, vol. 277, 504-509.

McPherson, J., W. Jetz & D. Rogers (2006). Using coarse-grained occurrence data to predict species distributions at finer spatial resolutions - possibilities and limitations. *Ecological Modelling*, vol. 192, 499-522.

Moisen, G. & T. Frescino (2002). Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling*, vol. 157, 209-225.

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

Olden, J.D., M.K Joy & R.G. Death (2004). An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling*, vol. 178, 389-397.

Rahbek, C. & G.R. Graves (2001). Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences, USA*, vol. 98, 4534-4539.

Ripley, B. (1996). *Pattern recognition and neural networks*. Cambridge University Press, New York.

Scheiner, S.M. (2003). Six types of species-area curves. *Global Ecology and Biogeography*, vol. 12, 441-447.

Segurado, P. & M. Araújo (2004). An evaluation of methods for modelling species distributions. *Journal of Biogeography*, vol. 31, 1555-1569.

Seoane, J., J. Vinuela, R. Díaz-Delgado & J. Bustamante (2003). The effects of land use and climate on red kite distribution in the Iberian peninsula. *Biological Conservation*, vol. 11, 401-414.

Thuiller, W. (2003). BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology*, vol. 9, 1353-1362.

Walsh, S.J., D.R. Butler & G.P. Malanson (1998). An overview of scale, pattern, process relationships in geomorphology: a remote sensing and GIS perspective. *Geomorphology*, vol. 21, 183-205.

Venables, W.N. & B.D. Ripley (2002). *Modern applied statistics with S*. Springer-Verlag, Berlin.

Wiens, J.A. (1989). Spatial scaling in ecology. *Functional Ecology*, vol. 3, 385-397.