University of Dundee

Mapping the species richness and composition of tropical forests from remotely sensed data with neural networks

Foody, Giles M. ; Cutler, Mark

Published in:
Ecological Modelling

DOI:
10.1016/j.ecolmodel.2005.11.007

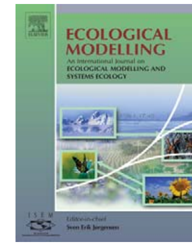
Publication date:
2006

Document Version
Early version, also known as pre-print

available at www.sciencedirect.com

SCIENCE DIRECT®

journal homepage: www.elsevier.com/locate/ecolmodel



Mapping the species richness and composition of tropical forests from remotely sensed data with neural networks

Giles M. Foody [a,*], *Mark E.J. Cutler* [b]

[a] *School of Geography, University of Southampton, Highfield, Southampton SO17 1BJ, UK*
[b] *Department of Geography, University of Dundee, Dundee DD1 4HN, UK*

ARTICLE INFO

Article history:
Published on line 19 December 2005

Keywords:
Remote sensing
Biodiversity
Neural network
Tropical forest

ABSTRACT

The understanding and management of biodiversity is often limited by a lack of data. Remote sensing has considerable potential as a source of data on biodiversity at spatial and temporal scales appropriate for biodiversity management. To-date, most remote sensing studies have focused on only one aspect of biodiversity, species richness, and have generally used conventional image analysis techniques that may not fully exploit the data's information content. Here, we report on a study that aimed to estimate biodiversity more fully from remotely sensed data with the aid of neural networks. Two neural network models, feedforward networks to estimate basic indices of biodiversity and Kohonen networks to provide information on species composition, were used. Biodiversity indices of species richness and evenness derived from the remotely sensed data were strongly correlated with those derived from field survey. For example, the predicted tree species richness was significantly correlated with that observed in the field ($r = 0.69$, significant at the 95% level of confidence). In addition, there was a high degree of correspondence (~83%) between the partitioning of the outputs from Kohonen networks applied to tree species and remotely sensed data sets that indicated the potential to map species composition. Combining the outputs of the two sets of neural network based analyses enabled a map of biodiversity to be produced.

1. Introduction

Arguments ranging from ethical to pragmatic have been proposed for the conservation of biodiversity (Hampicke, 1994). While there are great uncertainties associated with the current and future levels of biodiversity (Richards, 1994), human induced land cover change, notably the fragmentation and destruction of habitats, is known to be a major threat to biodiversity (Chapin et al., 2000). This is particularly evident in tropical forest environments, where competing pressures on the forests that are believed to support more than half of the total global biodiversity are of major concern. Understanding and managing forest biodiversity is, however, difficult. Although some forest sites may be designated as reserves for conservation, the level of protection may be incomplete and inevitably there is a desire to conserve larger areas that is incompatible with other competing demands on the forests. However, protected reserves are not the only means of conserving biodiversity. Unprotected and indeed actively used forests, such as those commercially logged, can represent an important resource in biodiversity conservation and for sustainable development (Cannon et al., 1998; Lugo, 1999). Moreover, it is increasingly recognised that there is a need to consider the conservation status of forests outside reserves and in particular to consider the landscape mosaic in biodiversity conservation. Thus, biodiversity conservation activities may be most appropriately undertaken at the scale of the landscape, that is over regions >100 km^2 (Innes and Koch, 1998;

* *Corresponding author*. Fax: +44 2380 593295.
E-mail address: g.m.foody@soton.ac.uk (G.M. Foody).
0304-3800/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2005.11.007

Gould, 2000; Margules and Pressey, 2000; Oindo and Skidmore, 2002). Unfortunately, most knowledge on forest biodiversity derives from small, typically <1 ha, plots (Innes and Koch, 1998) and the scaling of knowledge is extremely problematical (Loreau et al., 2001). The understanding and management of biodiversity would, therefore, be well served by the provision of information on biodiversity at the landscape scale.

Remote sensing has considerable potential as a source of information on biodiversity at the scale of the landscape. Recent studies have indicated that remote sensing may be able to provide useful information on biodiversity (Griffiths et al., 2000; Kerr et al., 2001; Nagendra, 2001). As with most of the literature on the topic, these studies have typically addressed biodiversity in terms of only the number of species or species richness. Biodiversity is, however, a broader issue than just a basic count of species present, and the composition of species and their relative abundance (evenness) are perhaps equally important but less well understood variables (Purvis and Hector, 2000). Additionally, many of the studies that have sought to estimate biodiversity from remotely sensed imagery have used basic vegetation indices such as the normalised vegetation index (NDVI) and standard statistical techniques such as regression (Gould, 2000; Oindo and Skidmore, 2002). The use of such approaches may not always be appropriate and often does not fully utilize the information content of the remotely sensed imagery. Here, we report on a pilot study to evaluate the potential of satellite remote sensing as a source of information on the richness, evenness and composition of tree species in a tropical forest using neural network techniques.

2. Test site

The test site was a 300 km^2 region of tropical forest surrounding the Danum Valley Field Centre in north eastern Borneo, Malaysia. The forests at this site have been used to varying degrees in the recent past. Part of the test site lies within an established conservation area that is believed to support primary rainforest while significant tracts within the site have been logged in the 1980s and 1990s. Consequently, the site contains regions that support forests of differing levels of biodiversity.

3. Data sets

In November and December 1997, field surveys were undertaken at the test site to acquire ground data on forest properties, including species composition. The ground data were acquired from 52 circular sample plots, each of 0.05 ha, established across a range of forest types. A stratified random sampling design was used, with each forest region (e.g. logging coupe) used containing five plots, a further two plots were acquired for one region to facilitate other research. In each plot, every tree with a diameter at breast height >20 cm was identified and, for commercially valuable species, its species recorded. Although the focus on only commercially valuable species is a limitation, information on their diversity is valuable, particularly in terms of sustainable development.

A search of remotely sensed data sets acquired within a year of the field survey was used to identify a data set of suitable quality, notably in terms of cloud contamination, spatial resolution and temporal coincidence, for relation to the data acquired by field survey. From this search, a Landsat Thematic Mapper (TM) image of the test site acquired in March 1997 was selected for the research. Here, the TM data acquired in the six non-thermal wavebands with a spatial resolution of 30 m were selected for the study (the data acquired in TM waveband 6 were not used). These data were rigorously pre-processed using conventional image analysis methods to reduce the impact of non-land cover variables on the remotely sensed response. The pre-processing of the data included a radiometric correction using pre-launch sensor coefficients (Mather, 1999), application of a simple image based atmospheric correction to reduce the effects of atmospheric attenuation (Chavez, 1996) and a topographic correction to reduce the effects of variation in terrain surface geometry and illumination on the remotely sensed response (Ekstrand, 1996).

Each plot surveyed in the field was identified in the image. To reduce the impact of image mis-registration errors (estimated to be in the order of a pixel), the remotely sensed response in the six TM wavebands was extracted for each plot from a 3 × 3 pixel window area centred upon its estimated location in the image. Unfortunately, cloud or cloud shadow impacted on the remotely sensed response of 22 plots and these were, therefore, excluded from analyses of the remotely sensed data.

4. Methods

Species richness and evenness were estimated from the data on forest composition acquired in the field. The two indices of biodiversity used in this research were the richness and evenness of the commercially valuable tree species observed at each plot. Specifically, the term richness in this paper relates to the number of commercially valuable species with a diameter at breast height >20 cm observed per-plot. The measure of evenness, E, used was the modified Hill's ratio that was computed for each plot from:

$$E = \frac{(1/\alpha) - 1}{e^H - 1}$$

where $\alpha = \sum_{i=1}^{s} p_i^2$, $H = -\sum_{i=1}^{s} p_i \ln p_i$, s is the number of species and p_i is the proportional abundance of each species (Ludwig and Reynolds, 1988).

Two neural network models were used to derive biodiversity information from the remotely sensed imagery. First, standard feedforward neural networks were used to estimate the species richness and evenness of the sample plots from the remotely sensed response. Three types of feedforward neural network were evaluated for this application: multi-layer perceptron (MLP), radial basis function and generalised regression neural networks (GRNN). In each situation, these feedforward neural networks were used essentially as an alternative to regression analysis. Second, Kohonen self-organising feature map neural networks were used in mapping species composition. Using unsupervised learning methods, this network was used to produce a topologically ordered two-dimensional output space in which forest sample plots were located in relation to their similarity, here expressed in terms of species composition. In this situation, the neural network was used to effectively ordinate the data (Foody, 1999; Giraudel and Lek, 2001). A Kohonen network was also used to derive an unsupervised classification of the remotely sensed data.

For the estimation of species richness and evenness, the set of 30 sample plots for which there was paired ground and remotely sensed data (i.e. for the plots unaffected by cloud, etc.) was divided randomly into two independent samples. One sample was used to train the networks to develop an invertible relationship between the remotely sensed response and the selected biodiversity index and comprised the data for 20 sample plots. The other sample contained the data of the remaining 10 sample plots and was used to evaluate the accuracy of the species richness and evenness predictions derived from the neural networks. These training and testing sets were the same as that used in an earlier investigation focused on the estimation of forest biomass from the TM image (Foody et al., 2001). Here, the potential to predict biodiversity indices from a variety of networks of differing type, architecture and parameter settings was evaluated. For comparative purposes, estimates of the biodiversity indices were also derived using both a standard ordinary least squared regression approach utilising all six TM wavebands and from NDVI data.

A Kohonen neural network was used to summarize the information on tree species composition acquired by the field survey. This network was used to derive an unsupervised classification in which the classes were arranged within its two-dimensional output on the basis of their relative similarity. In this way, the network performed a task similar to a vegetation ordination (Foody, 1999). A variety of networks and associated parameter settings were investigated. Here, the presence/absence of data on tree species composition for each of the 52 sample plots was entered into a Kohonen neural network with an output layer that comprised 5×5 units. A further Kohonen neural network was used to classify the remotely sensed data. Here, the aim was to reduce the remotely sensed data set to a small number of classes. The network used had an output layer comprising 2×2 units and was used to classify the data for the 30 plots free from cloud contamination. A similar two-phase analytical approach was adopted with each Kohonen network. The first phase involved 200 iterations of the algorithm with the learning rate set initially at 0.6 and a neighbourhood of 1 while the second phase involved a further 200 iterations with the learning rate initially set at 0.1 and a neighbourhood of 0. These parameter settings were defined, arbitrarily, after experimentation.

5. Results and discussion

A series of several hundreds of candidate feedforward neural networks were evaluated to predict the biodiversity indices from the remotely sensed response. For brevity, the discussion is focused on only the results providing the strongest relationship with the biodiversity indices together, for comparative purposes, with those derived using conventional alternative methods.

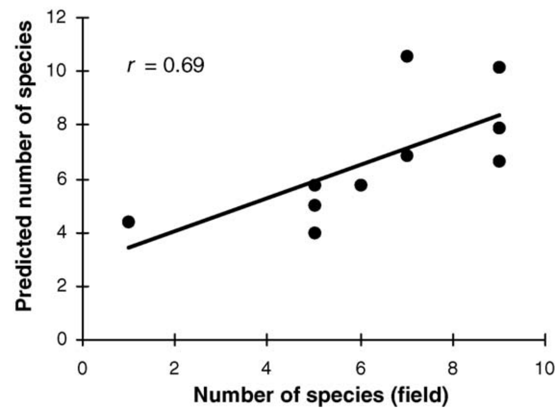


Fig. 1 – Relationship between the number of species predicted from a generalized regression neural network and that measured in the field. The species richness predictions derived from this network, which had 20 units in the first hidden layer, were strongly correlated to the richness observed in the field (training set $r = 0.88$, testing set $r = 0.69$).

The most accurate predictions of species richness were derived from a GRNN (Fig. 1; $r = 0.69$, significant at the 95% level of confidence).

A similar, but slightly weaker relationship ($r = 0.68$), was derived with the use of a standard backpropagation trained MLP neural network. Weaker relationships were observed for the estimation of species evenness. The strongest relationship between predicted evenness and that measured in the field was derived from a GRNN ($r = 0.45$, insignificant at the 95% level of confidence). Although the focus upon only commercially valuable species and the smallness of the sample size, together with the possibility that the training and testing sets may have been sub-optimally defined in terms of the range of values represented in each, are limitations, the results indicated the potential to derive a strong predictive relationship between biodiversity indices and remotely sensed data, especially for species richness. Moreover, the relationships were stronger than those derived using conventional alternative methods. The standard multiple regression equations derived from the training set, in which the remotely sensed data were the independent variables, were, for example, used to derive predictions of the biodiversity indices for the testing set that were not as strongly related to those derived from field observation as the predictions made from the neural networks. The correlation coefficients for the relationship between the predicted index values and those derived from field observation were 0.41 and −0.05 for species richness and evenness, respectively (both insignificant at the 95% level of confidence). Similarly, the widely used NDVI was less strongly related to species richness ($r = 0.49$, insignificant at the 95% level of confidence) and evenness ($r = 0.07$, insignificant at the 95% level of confidence) than the neural network predictions.

The Kohonen networks provided unsupervised classifications of the data sets. The sample plots were located within the output spaces of these networks in a manner defined by their relative similarity. Thus, the Kohonen network with a 2×2 output layer was used to define four classes of forest

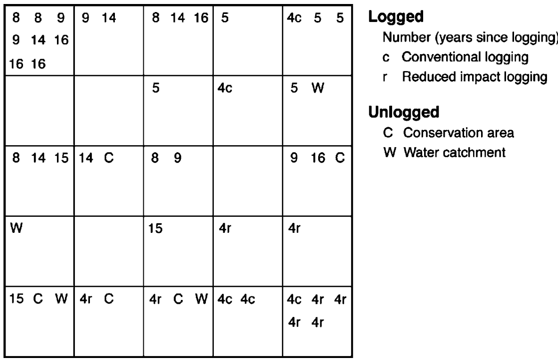


Fig. 2 – Location of sample plots in the output space of the Kohonen neural network applied to the tree species data. The symbol for each plot identifies the time (years) since the region in which it lies was last logged (and method of logging where appropriate) or the location of unlogged sites. Note there is a tendency for sites to cluster, with, for example, recently logged forests concentrated along the right hand side and older logged forests concentrated towards the upper left hand side of the output space.

that are highly separable spectrally. Similarly, the larger output space of the Kohonen network applied to the tree species data could be used to group together similar plots. Since plots were located in the output space by their relative similarity in terms of species composition, contiguous regions of the output space may represent similar forest types. This was evident, for example, in terms of past use of the forest (Fig. 2), with plots containing certain tree species located within limited parts of the output space (Fig. 3).

Fig. 3 – Location of a set of selected tree species within the output space of the Kohonen neural network applied to the tree species data. Note the confined space in which species were observed and that different species are often associated with dissimilar regions of the output space.

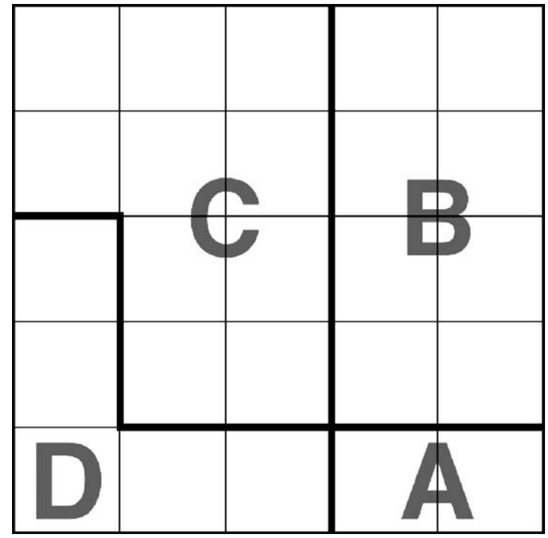


Fig. 4 – The partitioning of the output space of the Kohonen network applied to the tree species data (see Figs. 2 and 3) by that applied to the remotely sensed data. For illustrative purposes the four output units of the network applied to the remotely sensed data have been labelled A–D and their corresponding location in the output space derived from the network applied to the tree species data is illustrated.

The outputs from the two Kohonen networks were related to each other with the aim of identifying similarities in the groupings derived. Specifically, the aim was to determine if the output space of the network defined on remotely sensed data could be used to partition the output space from the network defined on the vegetation data and thereby indicate the ability to identify classes of tree species composition from the imagery. The plots allocated to each of the four output units of the network used with the remotely sensed data were located in the output space of the network used to ordinate the tree species data acquired in the field. The plots grouped by the network applied to the remotely sensed data were found to be clustered close together in the output space defined on tree species composition data. It was, therefore, possible to partition the output space of the Kohonen network defined by tree species composition into four spectrally separable groupings (Fig. 4). The partitioning was not perfect, but only 5 of the 30 plots (~17%) were linked inappropriately between the two Kohonen network outputs. Nonetheless, the simple partitioning of the Kohonen network output space depicting information on tree species composition by the outputs of the Kohonen network grouping the remotely sensed responses indicated the potential to map species composition classes from remotely sensed data.

Combining the results of the analyses based upon the feed-forward and Kohonen neural networks allowed a map depicting the spatial variation in species number, evenness and compositional attributes to be derived. An example is shown in Fig. 5, which indicates the variation in species richness for one of the four classes of species composition defined from

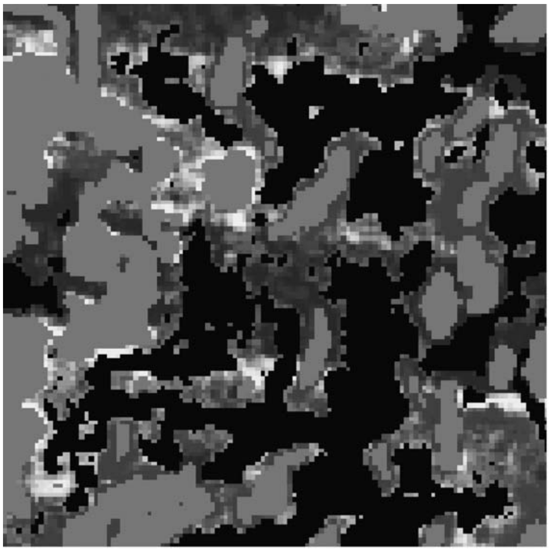


Fig. 5 – A sample representation of the biodiversity information extracted from the neural network based analyses. The variation in image tone mapped represents the spatial distribution of species richness for one of the four classes output from the Kohonen neural network applied to the remotely sensed data.

the output of the Kohonen network applied to the remotely sensed data.

6. Conclusions

There are great uncertainties associated with biodiversity but it is recognised that biodiversity conservation and management is probably most effectively undertaken at the landscape scale. Remote sensing has been identified as a major potential source of information on biodiversity at such scales. Most studies have generally only considered one aspect of biodiversity, that of species richness or the number of species present. Here, it has been shown that both species richness and species composition information can be usefully and accurately derived from remotely sensed data of a tropical forest site. Although the sample size and range of species considered in this study are limitations, the potential to extract useful biodiversity information from remotely sensed data is apparent. This potential is driving current research that aims to provide maps of biodiversity that may be used for monitoring purposes. This work aims to use a larger sample, which may be more appropriately divided into training and testing sets for biodiversity estimation. Additionally, it is intended to use the results of a major ground survey of the site to more precisely locate the plots in the remotely sensed imagery. The new locational information should enable the remotely sensed response of the individual pixel containing a specific plot to be extracted. This may enhance the ability to relate the ground and remotely sensed data sets as well as enable the useful derivation of image textural information. The latter should be of value in indicating forest heterogeneity that may further enhance the estimation of indices of biodiversity from remotely sensed data.

Acknowledgments

The data sets used were acquired as part of the EU funded INDFORSUS project (ERBIC18CT960102) and we are grateful to all project participants for their help. We are also grateful for the support provided by the Danum Valley Field Centre and, Royal Society's South-East Asia Rainforest Research Programme (this is programme publication A/373) and the three referees for their helpful comments.

REFERENCES

Cannon, C.H., Peart, D.R., Leighton, M., 1998. Tree species diversity in commercially logged Bornean rainforest. Science 281, 1366–1368.

Chapin, F.S., Zavaleta, E.S., Eviner, V.T., Naylor, R.L., Vitousek, P.M., Reynolds, H.L., Hooper, D.U., Lavorel, S., Sala, O.E., Hobbie, S.E., Mack, M.C., Diaz, S., 2000. Consequences of changing biodiversity. Nature 405, 234–242.

Chavez, P.S., 1996. Image-based atmospheric corrections—revisited and improved. Photogrammetric Eng. Remote Sensing 62, 1025–1036.

Ekstrand, S., 1996. Landsat TM-based forest damage assessment: correction for topographic effects. Photogrammetric Eng. Remote Sensing 62, 151–161.

Foody, G.M., Cutler, M.E., McMorrow, J., Pelz, D., Tangki, H., Boyd, D.S., Douglas, I., 2001. Mapping the biomass of Bornean tropical rain forest from remotely sensed data. Global Ecol. Biogeogr. 10, 379–387.

Foody, G.M., 1999. Applications of the self-organising feature map neural network in community data analysis. Ecol. Model. 120, 97–107.

Giraudel, J.L., Lek, S., 2001. A comparison of self-organising map algorithm and some conventional statistical methods for ecological community ordination. Ecol. Model. 146, 329–339.

Gould, W., 2000. Remote sensing of vegetation, plant species richness, and regional biodiversity hotspots. Ecol. Appl. 10, 1861–1870.

Griffiths, G.H., Lee, J., Eversham, B.C., 2000. Landscape pattern and species richness: regional scale analysis from remote sensing. Int. J. Remote Sensing 21, 2685–2704.

Hampicke, U., 1994. Ethics and economics of conservation. Biol. Conserv. 67, 219–231.

Innes, J.L., Koch, B., 1998. Forest biodiversity and its assessment by remote sensing. Global Ecol. Biogeogr. 7, 397–419.

Kerr, J.T., Southwood, T.R.E., Cihlar, J., 2001. Remotely sensed habitat diversity predicts butterfly species richness and community similarity in Canada. Proc. Natl. Acad. Sci. U.S.A. 98, 11365–11370.

Loreau, M., Naeem, S., Inchausti, P., Bengtsson, J., Grime, J.P., Hector, A., Hooper, D.U., Huston, M.A., Raffaelli, D., Schmid, B., Tilman, D., Wardle, D.A., 2001. Ecology—biodiversity and ecosystem functioning: current knowledge and future challenges. Science 294, 804–808.

Ludwig, J.A., Reynolds, J.F., 1988. Statistical Ecology. Wiley, New York.

Lugo, A.E., 1999. Will concern for biodiversity spell doom to tropical forest management? Sci. Total Environ. 240, 123–131.

Margules, C.R., Pressey, R.L., 2000. Systematic conservation planning. Nature 405, 243–253.

Mather, P.M., 1999. Computer Processing of Remotely Sensed Images, second ed. Wiley, Chichester.
Nagendra, H., 2001. Using remote sensing to assess biodiversity. Int. J. Remote Sensing 22, 2377–2400.
Oindo, B.O., Skidmore, A.K., 2002. Interannual variability of NDVI and species richness in Kenya. Int. J. Remote Sensing 23, 285–298.
Purvis, A., Hector, A., 2000. Getting the measure of biodiversity. Nature 405, 212–219.
Richards, M., 1994. Towards valuation of forest conservation benefits in developing countries. Environ. Conserv. 21, 308–319.