Accuracy 2010 Symposium, July 20-23, Leicester, UK

The quantification and effect of data accuracy on spatial configuration and composition measurements of landscape pattern in ecology

Alex M. Lechner*, Karin J. Reinke, Elizabeth Farmer, Bill T. Langford and **Simon D. Jones**
RMIT University
Melbourne, Australia
*alex.lechner@rmit.edu.au

***Abstract*—In natural resource management, ecological models are increasingly used to investigate the relationship between environmental and ecological processes (e.g. quantifying the response of rare and threatened species to habitat fragmentation). This paper examines the potential effect of data accuracy upon the measurement of landscape pattern using spatial configuration and composition measures commonly used in ecological modelling. This was achieved via an accuracy assessment of two standard environmental metrics calculated from commonly used continental and regional vegetation datasets and compared to a higher spatial resolution dataset treated as if it were the truth. Three study areas were selected from within Victoria, Australia, that represented varying degrees of spatial heterogeneity. The two metrics calculated to describe the environment within each plot were vegetation extent and Nearest Neighbour (NN). The first metric, vegetation extent describes spatial composition, whilst the second metric, NN, describes spatial configuration. Assessments were made at the local scale, represented by 80 one hectare circular plots for each study area. The effect of transforming the data was also tested as in many cases the relationship between ecological phenomenon and land cover measurements may not be linear.**

Results confirm the expectation that the utilisation of lower accuracy spatial data results in the derivation of less accurate environmental metrics. Further findings showed that vegetation extent was less sensitive to map product than nearest neighbour suggesting that certain metrics may be more susceptible to error. Importantly, the magnitude of error was also influenced by the type of mathematical function used to transform the data. The overall magnitude of error as recorded for vegetation extent, for each of the three landscapes, were qualitatively shown to poorly predict the magnitude of error that would occur using the spatially explicit NN environmental metric. Consequently, it is likely that global error statements (e.g. confusion matrices) that are not spatially explicit, are potentially inadequate for describing error in map products that are to be used for modelling spatially explicit phenomenon.

Keywords: ecology; environmental metrics; error propagation; landscape pattern; remote sensing; spatial error

I. INTRODUCTION

In natural resource management, ecological models are increasingly used to investigate the relationship between environmental and ecological processes. Remote sensing data is commonly used to create thematic maps that describe landscape pattern in order to understand the effects of landscape composition and configuration on ecological processes (Antrop 2007). One such application of ecological models is assessing the response of rare and threatened species to habitat fragmentation. Two important factors determining the effect of landscape pattern on species distribution and diversity are the total area of habitat and the structure of that habitat. It is well documented that spatial uncertainty and error resulting from differences in the geographic representation of a landscape will affect the interpretation of landscape structure with respect to its ecological function.

Species-environment relationships are commonly derived through examination of the relationship between measurements of landscape composition and configuration derived from remotely sensed data sources, and ecological metrics (e.g. population abundance) measured in the field (e.g. Cushman and McGarigal 2004; Davis et al. 2007). This paper examined two common measurements of landscape structure used in landscape ecology at the plot level derived from remote sensing data sources: vegetation extent and nearest neighbour. These two metrics were chosen as the first is an example of spatial composition and the second is an example of spatial configuration. The key difference between these two metrics, and the reason for their selection in this study, is that error in the arrangement of the pixels should not affect the measurement of the spatial composition metric in comparison to the spatial configuration metric.

This study investigated the effect of data accuracy on environmental measurements and relationships between ecological phenomenon and land cover measurements that may not be linear. This was considered important as ecological relationships, such as species-environment relationships, can be nonlinear (Lindenmayer et al. 2005). When relationships between environmental data and ecological data are nonlinear there is a potential for classification error to magnify in nonlinear ways.

II. METHOD

A. Study Area

Three study areas, each covering 614 hectares, were selected northwest of Melbourne, Australia. The study areas were chosen to represent three different landscapes of varying spatial heterogeneity. These were also areas that have experienced little or no significant landscape modification in the last decade. Study area 1 contained large, distinct areas of closed canopy woody vegetation. Study area 2 also contained large areas of woody vegetation, however these were characterised by a more open canopy structure and increasingly fragmented vegetation boundaries. In contrast, Study area 3 was a highly modified agricultural landscape in which woody vegetation was predominantly characterised by single or small groupings of trees.

B. Datasets

Two commonly used map products (1) the Australian National Carbon Accounting System woody vegetation dataset and (2) the Victorian State Tree25 dataset, were assessed against a reference dataset to investigate the impact of data accuracy on the measurement of landscape pattern. All products mapped the presence/absence of woody vegetation which was considered to be any woody vegetation greater than two metres in height.

1) National Carbon Accounting System (NCAS)

Developed by the Australian Department of Climate Change as part of the National Carbon Accounting System, this continental map product was derived from Landsat MSS, TM and ETM+ data re-sampled to a spatial resolution of 25 metres (Furby 2002).

2) Tree25 (T25)

Produced by the Department of Sustainability and Environment (DSE), this is a state-wide map product of tree cover across Victoria, Australia, at a scale of 1:25,000. (DSE 2006). It was derived from a combination of manual and automated classification of SPOT panchromatic imagery to a spatial resolution of 10 metres.

3) Reference Dataset

A high resolution reference dataset was generated on the basis of 0.15 metre near-infrared aerial photography. Woody/non-woody objects in this dataset were delineated via an automated image segmentation routine and subsequent manual Aerial Photograph Interpretation (API) process.

C. Spatial Configuration and Composition Measurements

The landscape pattern of each study area was measured using 80 randomly located, one hectare circular sample plots. Landscape measurements derived for each data source, within these sample plots were (a) the woody vegetation extent and (b) the nearest woody neighbour (NN)(where NN is calculated as the straight-line distance between the plot center and the nearest woody landscape element inside the plot). It should be noted that the maximum NN distance is limited by the size of the plot.

This kind of sampling design is commonly used in ecology to test influence of the surrounding landscape characteristics on ecological variables (e.g. species diversity) calculated at the center. The landscape measurements were calculated using the medium resolution NCAS and T25 map products and compared against measurements calculated from the reference dataset. The raw data was further transformed using two basic nonlinear functions commonly applied in ecology: a square $f(x)=x^2$ and log $f(x)=\log(x+1)$ as many ecological phenomenon may behave nonlinearly. Plot scale differences were subsequently summarised using the Normalized Root Mean Square Error (NRMSE)(1). NRMSE is a standardized measure which allowed for the comparison of error in vegetation extent and NN metrics despite different units of measurement being used.

$$NRMSE = \frac{\sqrt{(\sum_{i=1}^{n}(x_{1,i} - x_{2,i})^2)/n}}{x_{max} - x_{min}} \qquad (1)$$

III. RESULTS

A comparison of woody vegetation mapped via each product, qualitatively demonstrated that the API based approach characterised the landscape as having greater spatial heterogeneity than the T25 or NCAS map products. Fine

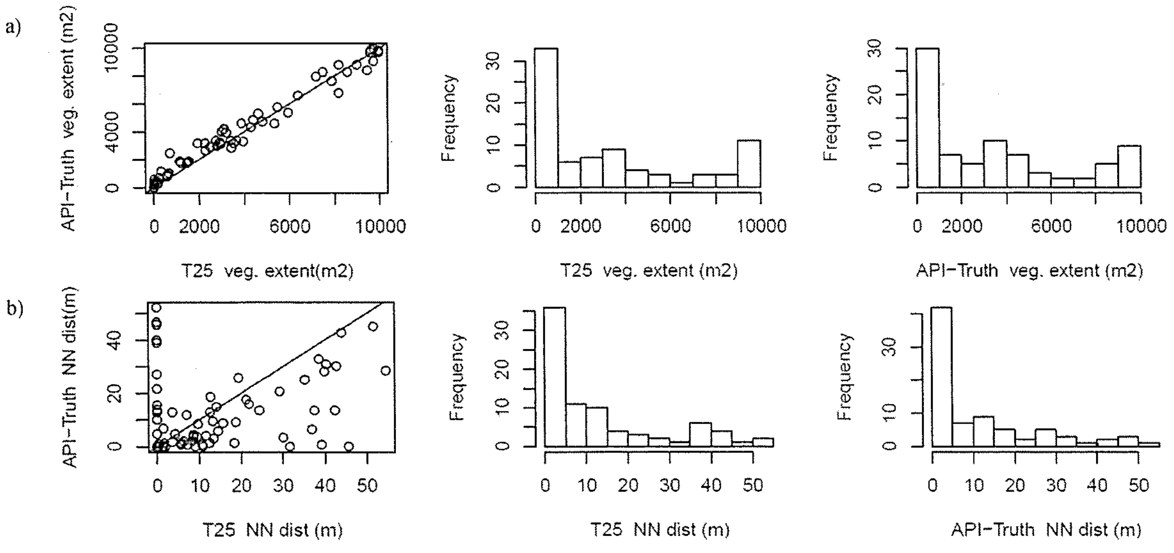

Figure 1. Comparison of measurments for (a) vegetation extent and (b) NN measurements between API (truth) and T25 data for study area 1.

scaled vegetation structures were under represented in the courser spatial resolution T25 or NCAS map products. For the majority of sample plots, woody vegetation occupied only a small percentage of the plot area (Fig. 1 and Table I). The distribution of vegetation extent was either skewed towards zero or non-normal for all study areas with a high frequency of zeroes. Study areas two and three showed similar distributions as represented by Fig. 1. The mean vegetation extent was greater for study area 1 than the other two study areas (Table I). In general, study area 3 had a much higher number of plots with no vegetation, and a lower mean vegetation extent in comparison to study area 1 and two (Table I) This trend was observed for all map products but with the API dataset reporting the least amount of variation between study areas. The distribution of values of NN were similar to those for vegetation extent; non-normal with many zero values (Fig. 1). The high number of plots with NN values of zero corresponded to plots with no vegetation recorded.

A comparison between vegetation extent and NN raw error values for the different map products demonstrates that the spatial composition measurements were in general more accurate than those for the spatial configuration metric (Fig. 1 and Fig. 2). This was true across all study areas with the exception of the NCAS dataset in study area 3 which recorded the opposite trend. The NRMSE recorded for the vegetation extent and NN metrics was always lower for T25 map product in comparison to the NCAS map product.

TABLE I. PLOT CHARACTERISTICS FOR EACH STUDY AREA (N=80 PER STUDY AREA)

Dataset	Study area	% plots with vegetation	Mean % vegetation extent
API	1	82	37
	2	84	17
	3	59	4
T25	1	72	28
	2	75	14
	3	15	1
NCAS	1	54	22
	2	24	7
	3	5	< 0.1

Transforming the data demonstrated NRMSE values to be much higher for the log transformed data than the raw or x^2 transformed data for all study area, landscape metric and data product combinations. The exception to this was for the vegetation extent metric as observed for study area 3 using the NCAS map product which showed results to the contrary. Transforming the data using $f(x)=x^2$ resulted in less error in comparison to the raw or the log transformed data for each study area, environmental metric and map product combination. Again, the exception to this was for vegetation extent in study area 3 using the NCAS dataset which resulted in completely different results to other combinations of study areas, data products and measurements methods. For this combination the data transformed with the $f(x)=\log(x+1)$ method had the lowest NRMSE and data transformed with $f(x)=x^2$ had the highest.

IV. DISCUSSION

There is a complex relationship between map product, landscape spatial heterogeneity, transformation function and the impact these may have on the accuracy of environmental metrics. Study areas 1 and 2 exhibited similar trends with regard to the effect of transformation function and map product on the accuracy of both environmental metrics. However these trends were not always observed in study area 3. The differences reported in study area 3 may be the result of it being the most modified landscape in the study, suggesting that spatial heterogeneity may be the cause of this observation. In particular for the NCAS dataset study area 3 had an extreme number of plots with no vegetation (95%) and a mean vegetation extent of less than 0.1%. A greater sample size of landscapes of various levels of heterogeneity is required in order to investigate what appears to be an important effect on the outcome of the results of this study. Thus the results of this study may only be representative of data with a particular distribution and spatial heterogeneity and not generalisable outside of the study area and map products used.

In most cases NN had overall higher error values than compared to vegetation extent. This is possibly due to higher accuracy requirements for NN spatial configuration measures than vegetation extent composition measures. In other words, it is possible for vegetation extent to be reported as correct regardless of spatial agreement with the reference data, unlike NN.

An expected driver, yet to be tested, for the higher errors observed in NN measures is the relationship between landscape spatial heterogeneity and the spatial distribution of classification error. For example, it can be considered that the error that exists between the boundaries of two classes (i.e. vegetation/non-vegetation) will have a smaller effect on NN measurements due to the high probability of the boundary being located nearby. As a result, the NN will only be incorrect by a short distance such as a single or few pixels. However, there are no standard methods for reporting or quantifying the spatial distribution of error (McGwire and Fisher 2001) thereby limiting the ability to measure the effect of spatial distribution of errors on environmental metrics and ecological models. Furthermore there is little understanding of how spatially heterogeneous landscapes with a high proportion effect classification error spatial distribution.

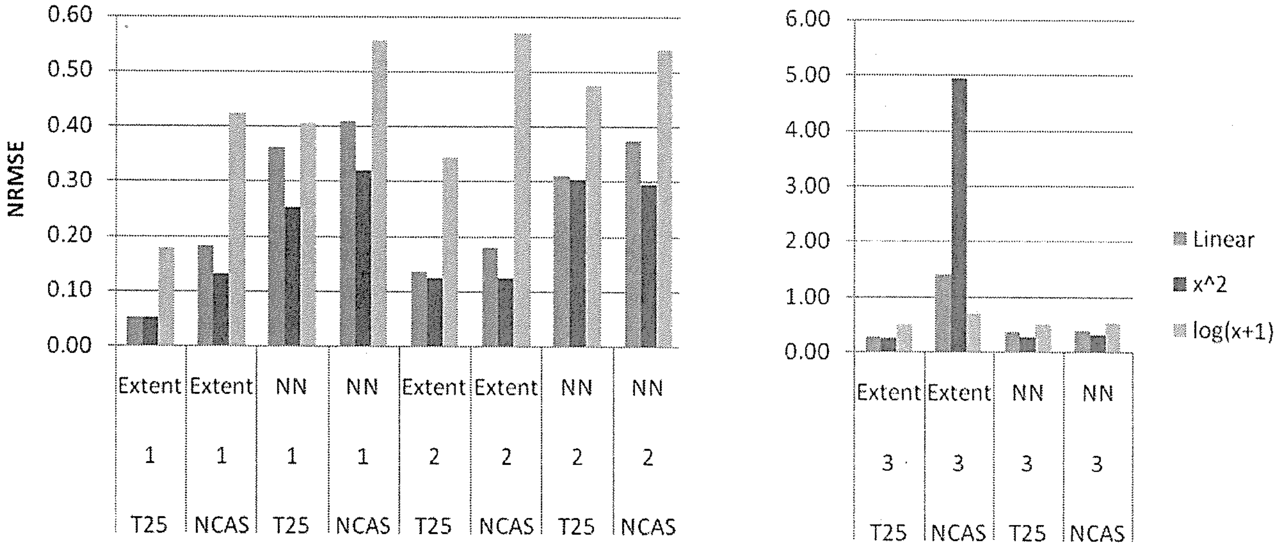

Figure 2. NRMSE for all combinations of environemntal metric, study area, map product and transformation.

The change in accuracy of NRMSE after applying a transformation is affected by the shape of the distribution of the original NN and vegetation extent measurements. For example, the nature of the log transformation is that it maximises error for lower values. The data for all study areas had many values that were zero or close to zero, thus the log transformation results in higher NRMSE values. The opposite is the case when applying the x2 transformation; error is maximised for data with high values. Furthermore as the NRMSE is calculated by dividing the root mean square error with the range of values the x2 transformation increases the range thus further decreasing the effect of the error occurring for lower values. Thus, for the data in this study with lots of low values, the x2 transformation results in decreasing NRMSE to less than recorded for the raw data.

Gergel et al. (2007) suggested that it is important to describe not only the magnitude of errors but also the implication of the errors when using spatial data in landscape ecology. The magnitude of error as recorded by vegetation extent, for each of the three landscapes, was much less than the spatially explicit NN measure by a factor of two or more. Langford et al (2006) found that under certain conditions classification error can cause a thousand-fold increase in error in the calculation of landscape metrics. Classification error, as is commonly measured for remote sensing data and reported using global error statements is aspatial, can be considered similar to the spatial composition (i.e. vegetation extent) measurement errors as measured in this study. Consequently, global error statements may be potentially inadequate for describing error in map products that are to be used to investigate spatially explicit phenomenon or non-linear relationships often modelled in ecology. Further work is needed to investigate whether this is also true for other metrics that describe spatial composition and configuration, and to establish the relationship with landscape heterogeneity.

ACKNOWLEDGEMENTS

This work was supported by the Australian Council Discovery Grant DP0450889 and by the Australian Commonwealth Environment Research Fund Landscape Logic research hub.

REFERENCES

Antrop, M. (2007). Reflecting upon 25 years of landscape ecology. *Landscape Ecology*. 22, 1441-1443.

Cushman, S.A., and McGarigal, K. (2004). Patterns in the species-environment relationship depend on both scale and choice of response variables. *Oikos*. 105, 117-124.

Davis, J.D., Debinski, D.M., and Danielson, B.J. (2007). Local and landscape effects on the butterfly community in fragmented Midwest USA prairie habitats. *Landscape Ecology*. 22, 1341-1354.

DSE (2006). *Tree Cover Grid (TREE25/) Metadata*. Victorian Spatial Data Directory: Department of Sustainability and Environment.

Furby, S. (2002). *Land cover change: specification for remote sensing analysis - technical report no. 9. CSIRO Mathematical and Information Sciences*, Australian Greenhouse Office.

Gergel, S.E. (2007). New Directions in Landscape Pattern Analysis and Linkages with Remote Sensing. In M.A. Wulder and S.E. Franklin (Eds.*), Understanding forest disturbance and spatial pattern: remote sensing and GIS Approaches* (pp. 173-208). Boca Raton, FL: Talyor and Fracis Group.

Langford, W.T., Gergel, S.E., Dietterich, T.G., and Cohen, W. (2006). Map misclassification can cause large errors in landscape pattern indices: Examples from habitat fragmentation. *Ecosystems*. 9, 474-488.

Lindenmayer, D.B., Fischer, J., and Cunningham, R.B. (2005). Native vegetation cover thresholds associated with species responses. *Biological Conservation*. 124, 311-316.

McGwire, K.C., and Fisher, P. (2001). Spatially variable thematic accuracy: beyond the confusion matrix. in C. Hunsaker, M. Goodchild, M. Friedl, and T. Case (eds) *Spatial uncertainty in ecology* (pp. 308-329). New York: Spinger-Verlag,