Implications of Scale Change on Native Vegetation Condition Mapping

Zerger, A. [1], G. Warren [1], M. Austin. [1], P. Gibbons [2] and J. Seddon [3]

[1] CSIRO Sustainable Ecosystems, Canberra, Australia
[2] Centre for Resource and Environmental Studies, Australian National University, Canberra, Australia
[3] Department of Environment and Climate Change (NSW), Canberra, Australia.
Email: andre.zerger@csiro.au

Keywords: *Vegetation condition, predictive modelling, GIS, remote sensing*

EXTENDED ABSTRACT

Native vegetation mapping has historically focussed on mapping the extent and composition (i.e. species and communities) of vegetation. However, policy makers have recently expressed a need for maps of native vegetation condition to assist with various aspects of native vegetation management (see Ecological Management and Restoration, *Special Issue on Native Vegetation Condition Mapping*, v.7, 2006). However a major challenge confronting researchers is to develop appropriate methods for extending site-based assessments of native vegetation condition to maps of native vegetation condition.

Zerger et al (2006) and Newell et al (2006) have proposed methods for creating maps of native vegetation condition by integrating GIS, remote sensing and predictive modelling. The methods rely on the use of spatially explicit models (e.g. statistical or neural networks) which assume a relationship between explanatory data available through GIS and remote sensing layers; and the response at the site or plot. Zerger et al. (2006) have argued that remote sensing has the potential to be an excellent predictor of native vegetation condition, in concert with traditional GIS-derived variables in a spatial model.

However, building remote sensing databases to represent such variables as vegetation cover, vegetation patch connectivity or the use of remote sensing indices such as the normalised difference vegetation index (NDVI) can be an expensive and time consuming component of a model building exercise. This becomes a particular problem for practitioners and natural resource managers who generally do not have access to high level image processing capabilities. For example, depending on the size of the study area, a key operational challenge may be to make a decision between high spatial but smaller swath width SPOT5 imagery, or multispectral Landsat TM data.

Via a case study predictive modelling experiment in the Murray Catchment of NSW Australia, this research compares the relative merits of regional/national scale Landsat satellite imagery and local scale SPOT5 satellite imagery for mapping native vegetation condition. As such the analysis provides an indication of the role of scale for native vegetation condition modelling.

The comparison of data scale is conducted by comparing the differences which emerge from predictive spatial models of native vegetation condition. Separate models are constructed for each scale of data. For the purposes of the comparison, modelling is conducted using Generalised Additive Modelling for 249 stratified vegetation condition plots acquired specifically for this study. Predictive models are evaluated using cross-validation methods.

Results show that there is limited difference between models developed using SPOT5 to those developed using Landsat TM data. We attribute this to issues of image seasonality and the challenges of developing image mosaics at regional scales, and the finer spectral resolutions inherent in Landsat TM data. Results from the evaluation have operational relevance to state and regional natural resource management bodies intending to develop regional maps of native vegetation condition to support their planning activities.

1. INTRODUCTION

Ongoing research by the authors (Zerger et al. 2006) has argued that remote sensing, in concert with GIS explanatory variables, should play an important role in any predictive spatial modelling of native vegetation condition. Remote sensing can complement the use of GIS-derived variables such as topographic position, land use, vegetation cover as it is a direct predictor of vegetation attributes. GIS variables on the other hand can be seen as indirect or surrogate variables. As it senses primary vegetation attributes, remote sensing also has the potential to contribute to the development of monitoring tools for native vegetation condition.

This paper examines the relative merits of two scales of remote sensing data (Landsat TM and SPOT5) in concert with GIS surrogates, as possible predictors of native vegetation condition. The assessment of scale is important for operational reasons as both satellite products are the most common operational platforms in Australia. Each platform provides a number of advantages and disadvantages including cost, ease of processing and spectral/spatial resolution. For example, the swath width of Landsat TM imagery means that complex edge-matching for regional scale mapping is not required. On the other hand, SPOT 5 provides spatial resolutions of approximately 10 metres ensuring that it is possible to map smaller vegetation remnants and scattered paddock trees.

This paper introduces the study area and discusses the predictive modelling methods with a particular emphasis on the remote sensing data and the respective indices and predictor variables derived from satellite imagery. Generalised additive modelling (GAM) results are presented for two structural vegetation attributes and model performance is evaluated using cross-validation methods. The relative contribution of SPOT5 and Landsat TM to the predictive models is examined and the discussion provides possible reasons for the observed differences.

2. STUDY SITE

The New South Wales Murray Catchment spans approximately 35,362 square kilometres extending from east of Khancoban to some 50 km west of Swan Hill. The catchment is considered one of the most modified regions in Australia owing to a history of agricultural production resulting in extensive clearing of native vegetation. It is estimated that 22% of the catchments' woody native vegetation remains with half of this reserved in several major national parks (Miles 2001). Consequently, a large proportion of the remnant native vegetation occurs either on private land, roadside vegetation and travelling stock reserves. As with other agricultural regions, the landscape is highly fragmented with many small isolated patches of remnant vegetation not linked to any major conservation easements. The project study area is situated across two 1:100,000 scale map sheets (561,316 ha) in the Murray Catchment Area (CMA) of NSW (Figure 1).

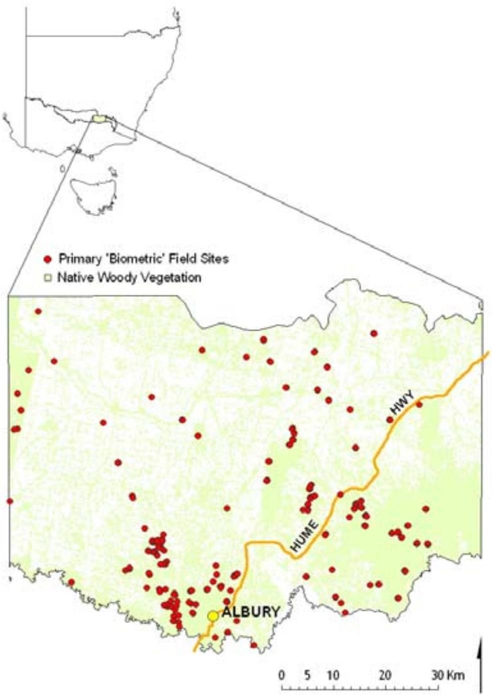

Figure 1. Study site map showing the location of *BioMetric* plots in the Murray CMA.

3. METHODS

There are a number of studies where statistical modelling, combined with GIS has been used to spatially predict vegetation species and community composition, rather than vegetation condition. Methods can include generalised linear and additive modelling, genetic algorithms (Newell et al. 2006) and classification and regression trees to name only a few. Elith et al. (2006) provide a detailed comparison of these methods from the perspective of model performance.

These methods all take advantage of regional-scale GIS databases such as high-resolution digital elevation models and their derived variables (slope, aspect, and topographic indices), soil and geology layers, climatic variables, land use and land tenure. Owing to the scale of generally available GIS data, modelling is typically conducted at regional or larger scales (catchments and sub catchments).

In this study, spatially explicit explanatory variables are explored as possible predictors of native vegetation condition. Spatially explicit predictors include remote sensing indices (Normalised Difference Vegetation Index, Perpendicular Vegetation Index, Greenness, Soil Adjusted Vegetation Index), and a suite of GIS predictors including digital elevation models, land use, landscape connectivity and vegetation cover to name a few. Landscape connectivity and vegetation cover layers are also derived from satellite imagery (SPOT5 and Landsat TM).

The relationship between explanatory variables and the vegetation condition attributes are explored and modelled using Generalised Additive Modelling (GAMs) under the GRASP framework (Lehman et al. 2003). Prior to model building, explanatory variables were tested for collinearity (r^2 > 0.8) and informed decisions were made regarding which variable should be removed from further analysis. Variables which are relatively simpler to derive or interpret, and ecologically intuitive variables were retained as candidates for modelling. Results are expressed in the scale of the additive predictor before transformation into the prediction scale by a link function. Variable selection was made using a stepwise (forward & backward) selection process using an Akaike Information Criteria (AIC) test.

Site data were collected using the '*BioMetric*' methodology which underpins the NSW Property Vegetation Planning (PVP) process (http://www.nationalparks.nsw.gov.au/npws.nsf/Content/BioMetric_tool Last Accessed: July 10, 2007). Although the *BioMetric* methodology contains 10 attributes (Table 2), this paper presents the results for only two of these. These include cover of native grasses and the volume of fallen logs at a site. These site condition attributes were chosen because they are important components of the composite site condition score in *BioMetric* and because it was thought they were likely to be described well by predictor variables derived from remote sensing and GIS data.

Table 2. '*BioMetric*' Attributes captured at each field site (variables marked * are those examined in this paper)

'BioMetric' Attribute
Native Plant Species Richness
Exotic Plants
* Native Ground Cover - Grasses
Native Ground Cover - Other
Native Ground Cover - Shrubs
Native Mid Storey Cover
Native Over Storey Cover
Organic Litter Cover
* Volume of Fallen Logs
Tree Hollows
Regeneration
Final *BioMetric* Score

Geo-referenced SPOT5 panchromatic and 4-band multi-spectral imagery and Landsat 5 TM 7-band multi-spectral imagery were acquired for the Murray CMA study area. Landsat 5 imagery was used rather than Landsat 7 owing to continued problems with the Landsat 7 platform (scan line correction failure). SPOT5 data consist of six scenes captured on four different dates in three seasons (spring, summer and autumn). The two western images were captured on 30/10/2004 and the two eastern images on 26/01/2005. The two central images were captured on 19/03/2005 (upper image) and 26/05/2005 (lower image).

The Landsat 5 TM data consist of two scenes; the western image was taken on 04/02/2007 and the eastern image on 25/12/2006. The spatial resolution of the SPOT5 multi-spectral data are 10 metres compared to the Landsat resolution of 30 meters, allowing the crowns of individual large trees to be distinguished. Landsat has more bands, particularly in the thermal and shortwave infrared. The increased spectral resolution of the Landsat sensor should compensate for the reduced spatial resolution when vegetation classifications are undertaken.

A key requirement for this research was to derive maps of woody vegetation cover to assist with landscape context analysis (patch size and landscape connectivity), and remotely sensed image indices which could help explain the condition of native vegetation through spectral analysis of pixel values. The following discussion describes the methods used to map vegetation configuration and density.

2.1. Vegetation Configuration and Density

Percentage crown cover was derived using a woody vegetation layer derived from SPOT5 and Landsat 7 multispectral satellite imagery using a supervised classification. Percentage cover was then derived by applying a moving window technique where the output is the sum of values found in the window. This is converted to per cent cover, or vegetation density. Three moving window dimensions were tested to account for any model sensitivity to analysis window size. Moving windows of 5 x 5, 10 x 10 and 20 x 20 cells were tested in the model (cell size of 10 metres for SPOT5 and 25 metres for Landsat TM). In addition to the mapping of per cent crown cover, the classified woody vegetation layer was also used to develop a 'habitat connectivity' layer using the methods of Drielsma *et al.* (2007). Remnant patch size was also calculated using the vegetation layer.

2.2. Remote Sensing Indices

Using Landsat and SPOT multi-spectral mosaics, Normalised Difference Vegetation Index (NDVI), Perpendicular Vegetation Index (PVI), Soil Adjusted Vegetation Index (SAVI) and Greenness indices were calculated by applying the relevant vegetation index function to the band-math tool of ENVI 4.3. The NDVI index is the ratio of near infrared to red fraction in the radiated or reflected spectrum. NDVI uses the band-math expression (NIR – red) / (NIR – red), calculating values for each pixel in an image. Values of zero indicate bare dry soil or water. NDVI values between zero and one indicate the presence of vegetation, the higher the NDVI value the greater the vegetation vigour.

PVI defines for each pixel in a dataset the distance that the vegetation radiance is located from the plane of soil reflectance. PVI is calculated using the expression *sin(a)NIR – cos(a)red* where the value *a* is the angle that lies between the soil line and the near infrared axis. The soil line is a hypothetical line in spectral space that describes the variation in the spectrum of dry bare soil in an image. PVI has poor dynamic range but it is sometimes better at discerning vegetation in scenes of low plant cover compared to NDVI.

SAVI minimises soil brightness induced variation that occurs in other vegetation indices, as such it is the best index where low plant cover is concerned. Somewhat of a hybrid between the perpendicular-based indices and the ratio-based indices SAVI is the ratio of near infrared to red fraction with the addition of a parameter to the red reflectance. The parameter *L* is empirically derived and ranges from zero for very high plant cover to a value of one for data with very low plant cover. Initially the Landsat and SPOT SAVI models included a constant value of 0.5 however it was found that increasing the constant to account for the low vegetation cover of the study area improved classifications. Greenness or Ratio Vegetation Index is the division of brightness values within the near infrared band by the corresponding red band values.

To ensure that remotely sensed vegetation indices were representative of the response observed across an entire *BioMetric* plot, the mean index value was calculated across the plot. To examine the role of spectral heterogeneity, the standard deviation of each index over a 1000 m^2 area (*BioMetric* plot) was also derived.

3. RESULTS & DISCUSSION

Figure 2 shows the frequency distribution for two of the *BioMetric* structural attributes for the Murray CMA. The histograms show patterns which can be expected in these fragmented and intensely modified landscapes. For example, some 50 sites have no observed fallen logs and the distribution is skewed. Similarly, the percentage of native grasses along a *BioMetric* transect are relatively small, perhaps reflecting the drought conditions in these regions. Alternatively the effectiveness of the stratification may drive these patterns.

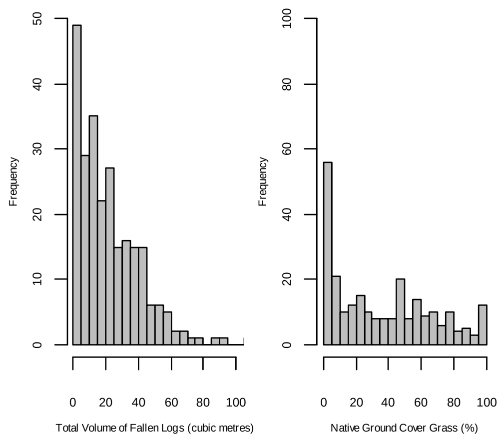

Figure 2. Frequency histogram showing distribution for two *BioMetric* variables including the volume of fallen logs and native ground cover (grasses) scores for 249 plots

Table 2. Candidate vegetation condition predictor variables – correlated variables ($r^2 > 0.8$) have already been removed

Predictor Variables	Name
Remote Sensing SPOT	
SPOT5 RSI (greenness) Focal Mean	grn_sp_fmrc25
SPOT5 NDVI Focal Mean	ndvi_sp_fmr25
SPOT5 RSI (greenness) Focal Std. Dev.	grn_sp_fstd
SPOT5 NDVI Focal Std. Dev.	ndvi_sp_fstd
SPOT5 PVI Focal Std. Dev.	pvi_sp_fstd
SPOT5 Vegetation Cover (10 x 10)	svcover10
SPOT5 Vegetation Patch Area	woodypatch_sp
SPOT5 Landscape Connectivity	sp_cba_07
Remote Sensing Landsat TM	
Landsat 5 TM PVI	pvi_ls5tm
Landsat 5 TM NDVI Focal Mean	ndvi_ls_fmr25
Landsat 5 TM NDVI Focal Standard Deviation	ndvi_ls_fstd
Landsat TM Vegetation Cover (4 x 4 window)	lcover4
Landsat TM Vegetation Cover (10 x 10 window)	lcover10
Landsat TM Landscape Connectivity	ls_cba_07
Landsat TM Vegetation Patch Area	woodypatch_ls
DEM Related	
Elevation (25 metre DEM)	dem25
Topographic Position Landform (4 classes)	landform
Topographic Position (Continuous)	tpos150
Topographic Roughness (2 x 3 window)	demfsd3
Other	
Land Use	landuse

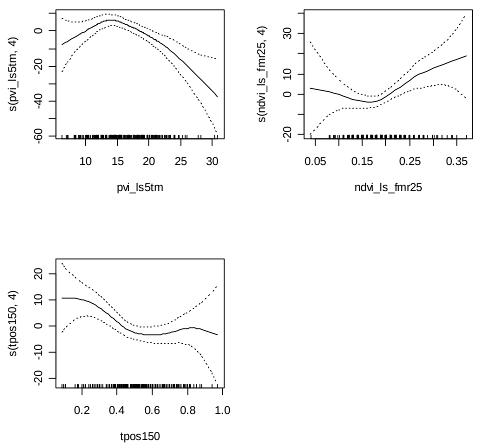

Figure 3. Partial response curves - Volume of Fallen Logs (Landsat TM) - dashed lines represent upper and lower point wise twice-standard-error curves.

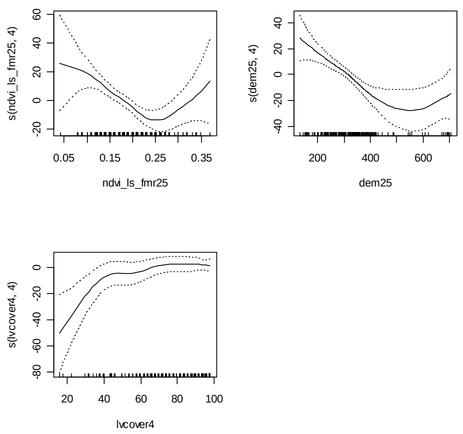

Figure 4. Partial response curves - Ground Cover Grasses (Landsat TM) – dashed lines represent upper and lower point wise twice-standard-error curves.

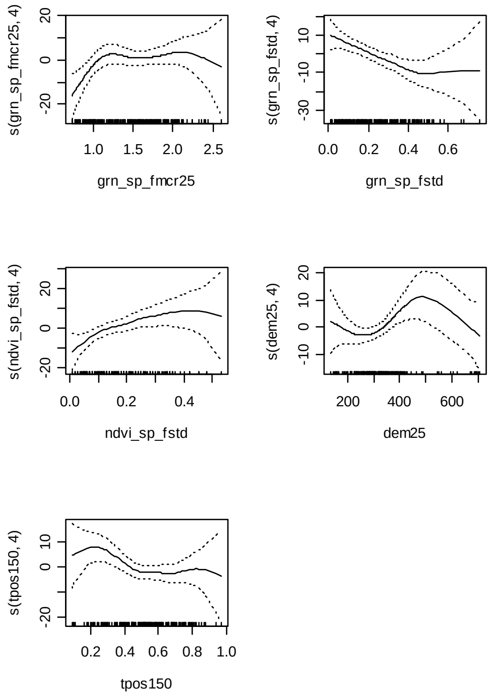

Figure 5. Partial response curves - Volume of Fallen Logs (SPOT5) - dashed lines represent upper and lower point wise twice-standard-error curves.

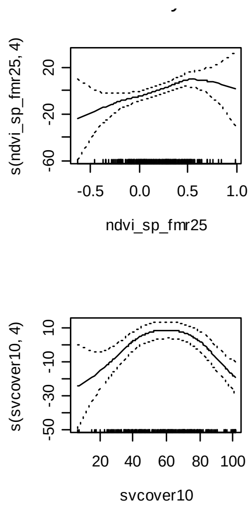

Figure 6. Partial response curves - Ground Cover Grasses (SPOT5) - dashed lines represent upper and lower point wise twice-standard-error curves.

Table 3. Final GAM model performance (r^2) for selected *BioMetric* attributes derived from cross validation (249 samples and 10 groups)

Vegetation Attribute	SPOT5	Landsat
Volume of Fallen Logs	0.21	0.27
Native Ground Cover Grasses	0.50	0.47

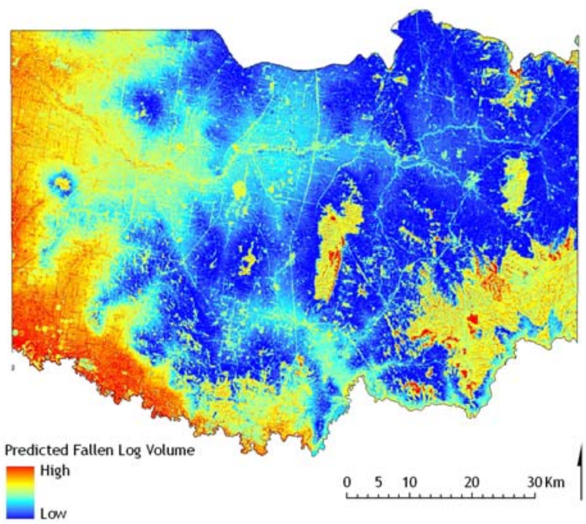

Figure 6. Example spatial prediction from GRASP modelling – Landsat TM derived prediction of volume of fallen logs .

Table 2 shows the final list of candidate variables after removal of variables correlated above an r^2 value of 0.8. This reduces a relatively large number of candidate modelling variables to 15 (SPOT5) and 14 (Landsat TM). Figures 3, 4, 5 and 6 show the response curves for selected predictor variables for both volume of fallen logs (SPOT5 and Landsat TM) and native grasses (SPOT5 and Landsat TM).

For all models, a remotely sensed predictor was selected in the final model. This ranged from a simple mean NDVI for the plot (native grasses SPOT5 model) to a combination of greenness and NDVI (volume of fallen logs SPOT5). If we look at the final models, results are consistent with what we would expect in these landscapes. For example, for the volume of fallen logs (Landsat TM), topographic position (tpos150) is selected as a predictor where we see a pattern of higher quality vegetation in low parts of the landscape (drainage channels or riparian corridors), decreasing vegetation quality on the flats and slopes where the primary intensive land uses (grazing and cropping) occur and then higher quality vegetation on ridges, albeit not as high quality as that observed along drainage channels. Of particular interest is Figure 6 and the contribution of vegetation cover (svcover10) to the final native grass model. Namely, we see lower densities of grasses where vegetation cover is less, increasing steadily as the cover increases but then witness a steady decrease as the cover passes a threshold. This is ecologically sensible as it summarises the relationship between overstorey competition and the impact of this on declining densities of native grasses.

Interesting patterns are observed when we look at elevation (dem25) as a surrogate for vegetation condition for the SPOT5 models (Figure 5). Namely we witness a steady increase in the volume of fallen logs when we move to higher positions in the landscape. Where a remotely sensed variable has been selected as a predictor of the structural attribute, modelled responses are also sensible. For example, in the case of volume of fallen logs and NDVI we see an increase in NDVI as the volume of fallen logs increases. As NDVI typically provides some indication of vegetation 'vigour' we would expect this to lead to a greater density of fallen logs at the plot.

Table 3 summarises the results from modelling by cross validation using a *k*-folds process with ten groups. Results are expressed in terms of r^2 values. Results highlight that there is little difference between the results obtained for the volume of fallen logs and native ground cover (grasses) when modelling with either Landsat TM or SPOT5 data.

4. CONCLUSION

Results have shown that for some vegetation condition attributes (volume of fallen logs and native grasses) it is reasonably difficult to obtain predictive accuracies in excess of an r^2 of 0.5 when building regional scale spatial predictions. For the models tested in this study, spectral remote sensing indices were always selected as significant predictors of native vegetation condition. This is important as remote sensing data senses primary attributes of native vegetation rather than acting as surrogates for disturbance as is the case with most GIS derived variables (e.g. topographic position). This provides some important opportunities for developing monitoring systems.

The results have also shown that there is very little statistical difference between outcomes when different scale remote sensing data is used in the GAM modelling (SPOT5 versus Landsat 5). We attribute the relatively limited predictive power of SPOT5 to image seasonality issues and the challenges of developing appropriately corrected mosaics at regional scales. Landsat TM data do not have the same limitations as it has a significantly larger swath width. This has important operational implications as generating SPOT5 mosaics at regional scales is a challenging and expensive process owing primarily to image seasonality. In addition, when using spectral indices derived from remote sensing, Landsat TM data also provides greater spectral resolution which may overcome its relatively poorer spatial resolutions (30 metres versus 10 metres for SPOT5).

This study has also found that relying on archival satellite imagery such as SPOT5 and Landsat has inherent limitations as it is difficult to obtain neighbouring scenes of similar seasonality and which are cloud free at regional scales. Image seasonality plays an important role in determining the effectiveness of the modelling as was seen by the limited utility of the SPOT5 imagery compared to Landsat TM. Consequently for both operational purposes and for optimal model performance it is more effective to utilise Landsat TM data for regional scale native vegetation condition mapping. Although not discussed in this paper, modelling results for other *BioMetric* variables are consistent with these conclusions.

5. ACKNOWLEDGEMENTS

The authors acknowledge the contribution of the Murray CMA and its staff (Emmo Willinck, Alexandra Knight and Jack Chubb), the NSW Environmental Trust and the Australian Government Department of Environment and Water Resources (Peter Lyon) for their ongoing support of the project.

6. REFERENCES

Drielsma, M., G. Manion and S. Ferrier (2007). The spatial links tool: Automated mapping of habitat linkages in variegated landscapes. *Ecological Modelling* 200(3-4): 403-411.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. McC. M. Overton, A. Townsend Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberón, S. Williams, M. S. Wisz and N. E. Zimmermann (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29(2): 129-151.

Lehman, A., Overton, J. McC. and Leathwick, J. R., (2003) GRASP: generalized regression analysis and spatial prediction, *Ecological Modelling* 160: 165-183.

Miles, C. (2001) NSW Murray Catchment Biodiversity Action Plan, Nature Conservation Working Group, pp. 86.

Newell, G. R., M. D. White, P. Griffioen and Conroy, M. (2006). Vegetation condition mapping at a landscape-scale across Victoria. *Ecological Management & Restoration* 7(s1): S65-S68.

Zerger, A., P. Gibbons, S. Jones, S. Doyle, J. Seddon, S. V. Briggs and D. Freudenberger (2006) Spatially modelling native vegetation condition. *Ecological Management & Restoration* 7(s1): S37-S4