Data quality and the detection of woody vegetation

Implications for environmental and conservation applications.

Elizabeth Farmer*, Karin J. Reinke, Alex M. Lechner and **Simon D. Jones**
School of Mathematical and Geospatial Sciences
RMIT University
Melbourne, Australia
* elizabeth.farmer@rmit.edu.au

Abstract **- This paper presents preliminary work into the data quality of a series of woody vegetation presence maps of differing spatial resolutions. The correspondence between two differing map products is compared. The non-woody/woody map products considered were derived from medium resolution satellite imagery, including the SPOT and Landsat satellites, and produced by government agencies at a national and regional scale. These 'off-the-shelf' products are routinely utilised in environmental, conservation and landscape management. Whilst metadata and data quality statements exist for these map products, localised error in terms of the detection of small patches of woody vegetation remains largely unquantified and poorly understood by users. This work describes and quantifies differences in mapped woody vegetation extent as a consequence of remote sensing data product. Localised errors are compared to landscape scale error estimates, and the implications for the delineation of small woody vegetation patches are discussed.**

Small patches of remnant woody vegetation have recognised ecological and conservational importance in terms of the ecosystem services they offer. Consequently, there is a distinct need to map and monitor these critical ecological structures. It is in this context that remote sensing technologies, due to their spatial coverage and synoptic view of landscape structure, are increasingly being utilised to provide assessments of woody vegetation extent. The ability of remote sensing technologies coupled with standard processing methods to accurately map and characterise small patches of woody vegetation needs to be understood in order for users of this information to achieve best practice in environmental and conservation management.

Previous research (Farmer *et al.*, in review) has indicated that the minimum mapping unit of remote sensing derived map products is significantly greater than a single pixel when considering small, discrete patches of woody vegetation. As a result small patches of woody vegetation are under-represented by up to 40% in current methods of vegetation mapping. The magnitude of these localised errors and their implications for landscape scale estimates of woody vegetation cover are demonstrated to be a consequence of map product and landscape structure, in particular landscape fragmentation and the associated patch size distribution.

Keywords: Fitness for purpose, Landscape structure, Scale, Spatial resolution, Woody vegetation, Extent

I. INTRODUCTION

Remote sensing data sources are increasingly being utilised by governmental and non-governmental organisations to support the management of natural resources, in particular native vegetation. Monitoring natural resources in ways that are timely and cost-effective can result in compromises in data product resolution. As a consequence not all data products are suitable for all applications, and there is a need for users to understand the implications of data source selection upon the intended use. Of interest, in this study, was the ability of differing map products, derived from remotely sensed imagery, to detect woody vegetation in particular fine scale vegetation structures such as remnant patches and paddock trees.

Within highly modified landscapes, these fine scale vegetation structures often represent the only remaining woody cover and the dominant representation of particular vegetation communities (Gibbons and Boak, 2002; Manning et al., 2006). In addition to serving as biological legacies, paddock trees provide multiple ecosystem services and improvements to the landscape. Manning et al. (2006) concluded that native scattered trees [paddock trees] should be considered keystone structures given their disproportionate effect on the ecosystem relative to the small area, density and low biomass they occupy. Yet, the immediate and non-immediate threats to these structures are pressing with loss due to agricultural intensification and human clearance currently the most direct threat to paddock trees (Manning et al., 2006; Maron and Fitzsimons, 2007). Given the threats to these keystone structures there is a need for a rapid, sufficiently accurate product capable of mapping and supporting the management of these small vegetation structures at a landscape scale (Gibbons and Boak, 2002; Carruthers et al., 2004).

The accuracy of woody vegetation extent, as derived from remote sensing technologies, is known to be influenced by the interaction of sensor spectral and spatial resolution, and the size and spatial arrangement of elements within the landscape (Ozdogan and Woodcock, 2006). Differences in woody vegetation extent, including those represented by paddock trees, as a function of remote sensing data source and classification methodology, have been estimated at between 1.5 and 11% in South Eastern Australia (Gibbons and Boak, 2000; Carruthers et al., 2004).

Inaccuracies in vegetation extent mapping typically arise when the elements of landscape pattern being considered are scattered and similar or smaller in extent to the spatial resolution of the sensor (O'Neill et al., 1996). Multiple authors (Moody and Woodcock, 1995; Wu, 2004; Ozdogan and Woodcock, 2006) have demonstrated that the spatial characteristics of a land cover class significantly influence its response to changing spatial resolution. Moody and Woodcock (1995) conclude that initial class proportion, patch size distribution, spatial arrangement and the interaction of these factors significantly influence the response of a land cover class to spatial aggregation. That is, the magnitude of errors in deriving class area statements, with increasing pixel sizes, are a function of the spatial structure of the land cover. Results from these studies were confirmed, in a South East Australian context, via a small scale comparison of 0.15, 10 and 25 metre spatial resolution woody vegetation maps. Differences between the data sources were demonstrated to be a function of vegetation amount and structure, that is, greater data source errors were apparent in those study areas characterised by highly fragmented woody vegetation structures.

Such results contrast those of landscape scale studies which demonstrate that total woody vegetation extent, at a landscape scale, remains relatively stable with increasing pixel sizes (Lechner et al., 2007). This may, in part, be due to the total area of small patches being small in comparison to the overall class area at a landscape scale (O'Neill et al., 1996) or a result of smaller patches being incorporated into larger patches. Accordingly, the exclusion of land cover contained in these small patches does not significantly change the overall cover proportion. As a result, indices based on the relative proportion of land covers are insensitive to the bias introduced by larger pixel sizes (O'Neill et al., 1996).

Previous studies have demonstrated that data product influences the detection of woody vegetation. Importantly, land use has been identified as a key factor in determining the magnitude of errors between data products. This can often have a greater impact than that of the data product spatial resolution (Farmer et al., in review). There is however, a need to expand existing small scale studies to the entire landscape so to determine if general area statements are valid at multiple spatial resolutions, where landscapes are dominated by small patches of remnant vegetation.

II. Method

Two temporally coincident map products (1) the Australian National Carbon Accounting System Woody dataset and (2) the Victorian State Tree25 dataset were used to assess the impact of geographic extent and land use on the magnitude of errors in vegetation maps of differing spatial resolutions. In this study, woody vegetation was considered to be any vegetation greater than 2 metres in height with a canopy cover exceeding 10%.

The study area was located in regional Victoria, Australia and encompassed a mix of regional towns, agricultural or arable areas and large woody blocks across an area of approximately 4 million hectares. Large woody vegetation stands were generally located on higher areas, typically delineated as national parks, while arable areas were typical of low lying flat areas. Arable areas, including both cropping and pastoral systems, were typified by large fields and limited woody vegetation cover. Woody cover in arable areas tended to occur as paddock trees or isolated stands of woody vegetation, usually less than 1 hectare in extent.

A. Datasets

1) Tree25

Produced by the Department of Sustainability and Environment (DSE), Victoria, Australia, Tree25 aimed to map the presence/absence of tree cover, at a scale of 1:25,000. It was derived from a combination of manual and automated classification of SPOT panchromatic imagery to a spatial resolution of 10 metres.

2) NCAS

Developed by the Australian Department of Climate Change as part of the National Carbon Accounting System (NCAS), this continental map product was designed to enable a multi-temporal land cover analysis of forest extent (Furby, 2000). The NCAS dataset was derived from Landsat MSS, TM and ETM+ data re-sampled to a spatial resolution of 25 metres.

B. Sample Design and Comparison

The Tree25 and NCAS map products were compared in a series of sample plots. The study area was stratified according to land use and 30 random, non-overlapping samples were independently taken, for each stratum, at each of the nominated geographic extents between 25 and 10,000 hectares (Table 1). Land use was extracted from the catchment scale, Australian Land Use and Management Classification (ALUMC) map product and generalised to represent very broad land use types. These land use types were defined to be representative of the likely distribution of woody vegetation in the landscape. Strata included (a) arable, that is, pasture and cropping systems within which woody vegetation was sparse and highly fragmented; (b) plantation and conservation, that is, remnant native woodland and plantations; and (c) a mixed stratum containing those samples in which the arable and plantation/conservation strata constituted a mix of between 40% and 60% (for each stratum).

The magnitude of differences between the two data products was calculated as the difference in mapped woody vegetation as a percentage of the vegetated area in the Tree25 product. This magnitude of difference was utilised, rather than absolute difference, as within highly fragmented landscapes a small absolute difference had the potential to represent a significant proportion of the total woody vegetation amount. The Tree25 dataset was taken as the reference, to which the NCAS-Tree25 difference was compared, as this map product had a higher resolution and had previously been demonstrated to be most similar to a highly accurate (aerial photograph derived) reference data source.

III. Results and Discussion

Total woody vegetation extent, in the entire study area, was recorded as approximately 1.6 million and 1.5 million hectares for the Tree25 and NCAS data products, respectively. This accounted for approximately 40% and 38% of the entire study area.

For geographic extents greater than 2500 m2 the median difference between the data products, in the sampled areas, was relatively invariant (Table 1) and approached that of the entire study area, that is, 6% (Table 2). This suggested that geographic extent did not significantly influence class area statements at these scales. Conversely, when the geographic extent was less than 2500 m2, differences between the data products increased significantly, as indicated by a higher median difference (Table 1). Such a result, suggested that landscape structure had the potential to influence data product similarity when geographic extents were small and characterized by a single land use or landscape structure.

The influence of land use, a proxy for landscape structure, upon map product similarity was evident following the stratification of the entire study area (Table 2). As would be expected, mapped woody vegetation extents, in the NCAS and Tree25 products, were most similar in those land uses characterised by large blocks of woody vegetation. In contrast, within highly modified, arable, landscapes differences between the data products increased to 32% (Table 2).

When sampled across the landscape, the plantation and conservation stratum samples were characterised by medians of less than 1% with minimal variation across the different geographic extents. This was reflected in both the median values and the reported quartiles. Consequently, both map products reported similar woody vegetation area statements in this land use stratum. This has important implications for users of these types of data particularly in terms of the reduced computational load of processing a lower resolution data product.

TABLE I. THE INFLUENCE OF GEOGRAPHIC EXTENT EXPRESSED AS THE MEDIAN DIFFERENCE IN TOTAL WOODY AREA DETECTED BETWEEN THE TWO MAP PRODUCTS

	Magnitude of Difference (%)		
Extent (m^2)	*Median*	*First Quartile*	*Third Quartile*
500	31.26	0.92	96.67
1000	22.62	2.38	59.18
2500	8.07	0.55	46.15
5000	9.49	0.96	45.43
7500	6.72	0.40	81.64
10000	8.25	0.58	82.24

TABLE II. THE MAGNITUDE OF DIFFERENCE BETWEEN TREE25 AND NCAS FOR THE ENTIRE AREA FOR EACH OF THE PURE STRATA TYPES AND FOR THE STUDY AREA (ALL STRATA TYPES INCLUDED).

Land Use	**Magnitude of Difference (%)**
Plantation and conservation	3
Arable	32
All land uses	6

Arable samples, at each extent, had a significantly higher median difference (greater than 75%), when compared to remaining land use strata. Previous studies (Farmer et al., in review) have indicated that in such highly fragmented environments these differences are a consequence of small patches of remnant vegetation. This conclusion was further supported by the current analysis which demonstrated a significant underestimation of small patches of woody vegetation by the NCAS data product. For example, within the 1000 hectare arable samples the total woody vegetation amount was 143 hectares in the Tree25 product versus 46 hectares in the NCAS product. Therefore, in arable samples, the Tree25 product identified significantly more woody vegetation. Examination of the patch size distribution of woody vegetation in these samples demonstrated that a large proportion of patches, identified by Tree25, were below the stated minimum mapping unit of the NCAS data product (0.2 hectares). Differences between the map products, within the arable stratum, were therefore a consequence of the increased number of small (less than 1 hectare) patches identified by the Tree25 data product. These differences were evident at all sampling extents, as indicated by the high median difference between the map products at all extents.

In highly modified landscapes characterized by low woody vegetation amounts and dominated by paddock trees, small patches had the potential to represent a significant proportion (if not all) of the woody vegetation. Their exclusion from a data product, as a consequence of decreasing spatial resolutions, therefore, had the potential to significantly affect the reported amount of woody vegetation at both landscape and local scales.

Results indicate that the ability of class area statements to remain stable with increasing pixel size is a function of (a) the amount of woody vegetation and (b) its spatial configuration. Arable systems, it is proposed, contain insufficient woody vegetation to allow for errors of omission and commission to reach equilibrium. This is characteristic for the strata across all study area extents. However, as larger woody blocks become characteristic of the landscape, as exemplified by the mixed land use, it is suggested that errors of omission and commission move towards equilibrium. Equally, as landscapes become characterised by larger woody blocks, small patches (undetected with increasing pixel size) decreased in their contribution to the total woody vegetation amount. As such the exclusion of these structures from total woody area estimates is less influential on the reported woody vegetation amount.

IV. CONCLUSIONS

This paper presents preliminary results of ongoing research that aims to better understand the implications that application or problem context have in determining data fitness-for-purpose. Results indicate that land use, a proxy for landscape structure, has a significant influence on the ability of different map products of differing spatial resolutions, to support the monitoring and management of woody vegetation at a landscape scale. Global errors were shown to be non-representative of localised errors, particularly for given land use strata, highlighting a need to communicate the influence of localised errors and their implications to data users. While users are often presented with a global error statement, research is required to determine if localised errors have the potential to significantly influence the intended application.

As previously identified, there is a critical need to manage small patches of remnant vegetation and better understand their contribution to ecological functions and climate change mitigation (Manning et al., 2006). Continuing studies have demonstrated the substantial influence of data product and scale of comparison upon woody vegetation mapping, in particular small remnant patches. This requires the development of coherent and strategic methods to manage these small scale structures across landscape scales, a task which is currently beyond the effectiveness of localised aerial photograph based mapping.

TABLE III. THE INFLUENCE OF LAND USE AND EXTENT EXPRESSED AS THE MEDIAN DIFFERENCE IN TOTAL WOODY AREA DETECTED BETWEEN THE TWO MAP PRODUCTS.

		Magnitude of Difference (%)		
	Extent (m^2)	*Median*	*First Quartile*	*Third Quartile*
Arable	500	100.00	47.92	100.00
	1000	79.36	49.35	100.00
	2500	75.10	37.79	93.98
	5000	75.72	44.27	101.45
	7500	80.88	59.61	137.03
	10000	82.07	49.40	123.80
Mixed	500	16.30	4.22	57.80
	1000	12.99	5.57	36.91
	2500	8.95	3.88	18.78
	5000	10.77	3.90	15.42
Plantation and Conservation	500	0.00	0.00	2.38
	1000	0.35	0.06	2.38
	2500	0.27	0.08	1.51
	5000	0.51	0.30	1.39
	7500	0.45	0.20	1.19
	10000	0.65	0.33	1.21

For each extent, n = 30, extents where n was less than 30 the strata was excluded from the analysis.

Two research directions are underway from this work. The first looks at the ability of data sources to support the management of fine scale vegetation structures at a landscape scale. Further investigation is needed to compare remotely sensed map products to the "ground truth" in order to determine the optimum product for reporting the relevant spatial and aspatial metrics so to describe the current woody vegetation resource. The second investigates the impact of the data product within the context of decision frameworks and applications. This research proposes that quantifying the differences between these products is not sufficient in itself, arguing that there is a need to understand data error propagation within commonly performed geoprocessing techniques and models. Fundamental to this research is the consequence of underestimating small remnant landscape structures within decision making models.

ACKNOWLEDGEMENTS

Financial assistance was provided by the Australian Commonwealth Environment Research Fund 'Landscape Logic' research hub and is gratefully acknowledged.

REFERENCES

Carruthers, S., Bickerton, H., Carpenter, G., Brook, A. and Hodder M. (2004). *A landscape approach to determine the ecological value of paddock trees.* Summary Report Years 1 and 2. Biodiversity Assessment Services, South Australian Department of Water, land and Biodiversity Conservation.

Gibbons, P. and Boak, M. (2000). *The importance of paddock trees for regional conservation in agricultural landscapes: A discussion paper for consideration by the Riverina Highlands Regional Vegetation Committee.* New South Wales National Parks and Wildlife Services, Southern Directorate.

Farmer, E., Reinke, K. J. and Jones, S. D. (in review). Quantifying fine-scale vegetation structures: Evaluating accuracy and corrsependance at local scales.

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

Gibbons, P. and Boak, M. (2002). The value of paddock trees for regional conservation in an agricultural landscape. *Ecological Management and Restoration.* 3 (3), 205-210.

Lechner, A. M., Jones, S. D. and Bekessy, S. A. (2007). A study on the impact of scale dependent factors on the classification of landcover map. In Stein, A., Shi, J. and Wietske, B. (Eds). *Quality aspects in spatial data mining.* (pp. 315-328). Boca Raton: Chapman and Hall/CRC Press.

Manning, A. D., Fischer, J. and Lidenmayer, D. B. (2006). Scattered trees are keystone structures – Implications for conservation. *Biological Conservation.* 132, 311-321.

Maron, M. and Fitzsimons, J. A. (2007). Agricultural intensification and loss of matrix habitat over 23 years in the West Wimmera, south-eastern Australia. *Biological Conservation.* 135, 587-593.

Moody, A. and Woodcock, C.E. (1995). The influence of scale and the spatial characteristics of landscape on land-cover mapping using remote sensing. *Landscape Ecology.* 10 (6), 363-379.

O'Neill, R.V., Hunsaker, C.T., Timmins, S.P., Jackson, B.L., Jones, K.B., Ritters, K.H. and Wickham, J.D. (1996). Scale problems in reporting landscape pattern at the regional scale. *Landscape Ecology.* 11, 3, 169-180.

Ozdogan, M. and Woodcock, C.E. (2006). Resolution dependent errors in remote sensing of cultivated areas. *Remote Sensing of Environment.* 103, 203-217

Wu, J. (2004). Effects of changing scale on landscape pattern analysis: Scaling relations, *Landscape Ecology*, 19, 125-138.