Areal Sample Units for Accuracy Evaluation of Single-date and Multi-temporal Image Classifications

Kim Lowell[1], Alex Held[2], Tony Milne[3], Anthea Mitchell[3], Ian Tapley[3], Peter Caccetta[4], Eric Lehmann[4], Zheng-shu Zhou[4]

[1] Cooperative Research Centre for Spatial Information, Dept. of Infrastructure Engineering, University of Melbourne, Carlton, VIC 3053 AUSTRALIA klowell@crcsi.com.au
[2] CSIRO Canberra, ACT 2601 alex.held@csiro.au
[3] Cooperative Research Centre for Spatial Information, School of Biological, Earth, and Environmental Sciences, The University of New South Wales, Kensington, NSW 2052 t.milne@unsw.edu.au, a.mitchell@unsw.edu.au, hgciant@bigpond.net.au
[4] CSIRO Mathematics, Informatics & Statistics, Floreat, WA 6014 (peter.caccetta@csiro.au, eric.lehmann@csiro.au, zheng-shu.zhou@csiro.au)

Abstract

Areal sample units are explored as an alternative to conventional point-based samples for map accuracy assessment. Three analyses are examined: confidence limits on estimates of the total amount of each landcover, regression of map data versus reference data, and a two-part confusion matrix approach. It is concluded that areal sample units provide some advantages compared to point-based samples, but are nonetheless subject to problems associated with providing robust accuracy assessment metrics for rare classes.

Keywords: landcover change, rare event sampling, regression, modified confusion matrix

1. Introduction

Considerable effort on accuracy assessment of remote sensing products addresses image classification schemes for a single-date. This effort has produced a well documented and widely accepted methodology; a useful reference is Congalton and Green (2009).

As digital image archives have expanded over time, considerable effort is now being expended on multi-temporal land cover change mapping. This is also being driven by carbon accounting and international climate change activities.

Efforts to assess accuracy of landcover change maps has not kept pace. Congalton and Macleod (1994) were among the first to address this; they proposed an extension of the well-known confusion matrix approach to land cover change maps. A modification of this concept -- the Transition Error Matrix or TEM has also been explored (Liu and Zhou, 2004; Van Oort, 2007; Li and Zhou, 2009). This approach addresses the difficulty of obtaining reference data that match the period covered by the land cover change map. Correctness is inferred by examining the change classes or "trajectories." For example, in an area undergoing rapid urban expansion, any points falling in an urban-to-agriculture trajectory are assumed to be in-

correct. Lowell *et al.* (2005) also wrestled with the lack of temporally suitable reference data. They tabulated single-date landcover information against a point's "lineage" (i.e., trajectory) using a fuzzy interpretation.

The use of points as sample units presents difficulties for accuracy assessment of landcover change maps because of the lack of temporally suitable reference data, but also because the relative rarity of land cover change which is of greatest interest. To overcome these problems Barson *et al.* (2004) and Lowell (2001) used relatively large areas as sample units.

The references cited have focussed primarily on the accuracy of global estimates of land cover change – specifically deforestation and reforestation. Given the increased focus on land cover change, accuracy assessment methodology that addresses more than global land cover estimates would be useful. The purpose of this paper is to explore a number of alternatives that are relevant to sample units that are relatively large areas.

2. Study Area and Data

The study area is Tasmania – the island state located approximately 240 km south of the eastern portion of the Australian mainland. Tasmania is approximately 6.8 million hectares in size of which approximately two-thirds is forested.

Single-date and multi-temporal maps derived from two different sensors – thematic Mapper (TM; optical) and PALSAR (radar) – were used. For both sensors, image data for Tasmania had been acquired and processed to produce binary maps of forest and non-forest for 2007 and 2009; these were then used to produce a 2007-2009 landcover change map using map overlay.

A 33 km square sample grid was established over Tasmania. At each sample location, a 10 km square (10,000 ha) was established. The amount of forest and non-forest was extracted for each sample from the 2007 and 2009 maps, and the amount of deforestation, reforestation, and no change extracted from the change maps.

Thirty of these areal sample units were semi-randomly chosen for subsequent analysis. For purposes of demonstration, the maps derived from TM data are considered to be the maps whose accuracy is being assessed, and the maps derived from PALSAR are treated as the reference data.

3. Methodology and Results

Three metrics are considered.

First, conventional 95% confidence limits were calculated for each landcover type on each map. This was done with progressively larger sample sizes to identify the number of samples required to reach a robust conclusion about map accuracy.

Second, the amount of each landcover type from the TM-based and PALSAR-based maps were regressed against each other.

Third, a two-part confusion matrix approach was employed. The first part is a conventional pixel-by-pixel analysis cross-tabulation of map data (TM) against reference data (PALSAR). The second part is an "areal confusion matrix." Map data are tabulated against reference data with respect to the total size of the sample unit. Table 1 illustrates this for a two-class map. (A more detailed discussion for

more than two map classes is provided by Pontius and Cheuk (2006).) For this fictitious 10,000 ha sample unit, the TM and PALSAR maps record 7,000 ha and 5,000 ha of forest, respectively, and 3,000 ha and 5,000 ha for non-forest, respectively. The diagonal (agreement) elements are obtained using the minimums of the forest and non-forest classes -- 5,000 ha and 3,000 ha respectively. The off-diagonal elements are determined by ensuring that the totals are correct for each landcover type for each sample. The confusion matrices for individual sample units are then summed to provide a global "areal confusion matrix".

Table 1: Areal confusion matrix for a single 10,000 ha sample unit having 7,000/3,000 ha of forest/non-forest on the optical map, and 5,000/5,000 ha of forest/non-forest on the radar map.

		PALSAR(Radar)		
		Forest	Non-forest	TOTAL
TM(Optical)	Forest	5,000	2,000	7,000
	Non-forest	0	3,000	5,000
	TOTAL	5,000	5,000	10,000

4. Results

Each of the three analyses described was applied to the 2007, 2009, and 2007-2009 landcover change maps. In the interest of space, results are presented for 2009 non-forest, 2007-2009 no change, and 2007-2009 deforestation.

Figure 1 shows 95% confidence limits for the total area of each class. In all cases, it can be concluded that there is no difference between the TM and PALSAR maps. This conclusion is reached after approximately 15 samples are examined as this is the point at which confidence intervals stabilise. This represents a sampling intensity across Tasmania of about 2%. However, the abrupt changes that occur for the change map confidence intervals (e.g., at n=26 in Figs. 1b and 1c) indicate difficulties associated with rare classes. Had the change maps been less similar, or change been more spatially concentrated, assessment of map differences may only have been conclusive with a much greater sampling intensity.

Table 2 shows the regression statistics that describe the relationship between the TM and PALSAR data. These describe the differences between the maps well. Though the relationship for the 2009 non-forest class is weak ($r^2 = 0.36$) yet statistically significant, the root mean square error (RMSE) is large considering that the sample unit is 10,000 ha. The intercept being above 0.0 (albeit not significantly) and the slope being significantly below 1.0 indicate that at lower amounts of non-forest (as estimated on the PALSAR map), TM estimates for non-forest will be larger than PALSAR estimates, but at higher amounts of non-forest, PALSAR estimates of non-forest will be larger. The information for the change maps indicate no agreement between maps: r^2 values are not statistically significant, and slopes and intercepts are generally significantly different from the ideal values. Figure 2 suggests that this is due to the rareness of the change classes. Most of the sample units have little deforestation or regeneration meaning there is a cluster of points around the point x,y (PALSAR,TM) of (10000, 10000). Any points with a large amount of deforestation and/or regeneration therefore have high leverage rendering regression results of limited value.
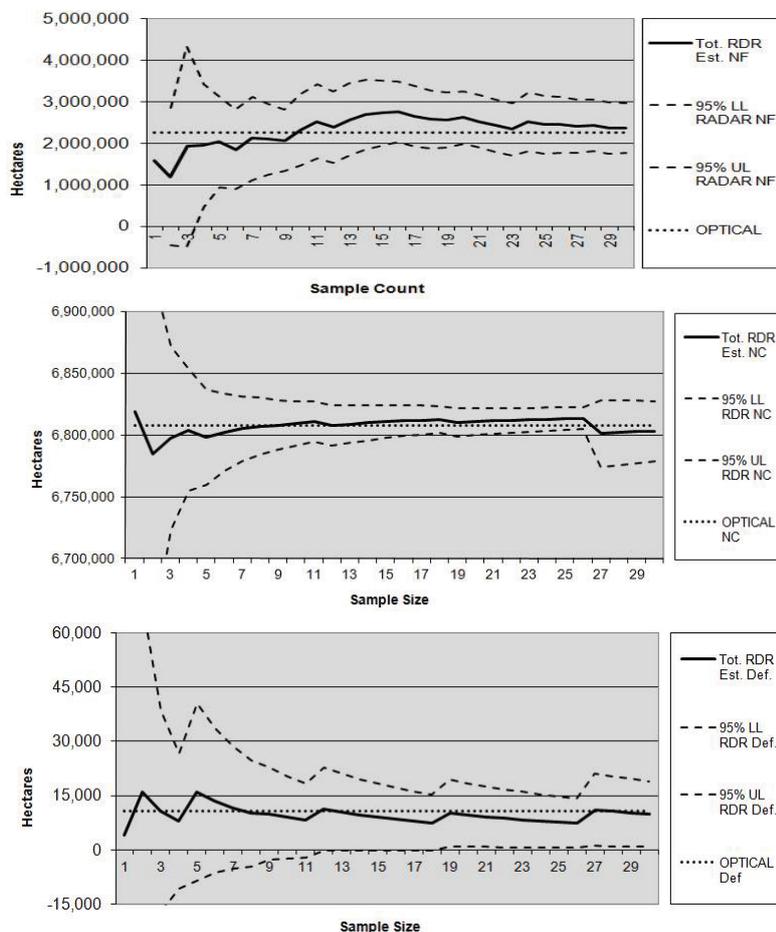
Figure 1: Confidence limits (95%) on Tasmania-wide amount of different land cover types with increasing sample size. 1a(top): 2009 Non-forest. 1b(middle): 2007-9 No change. 1c(bottom): 2007-9 Deforestation.

Table 3 shows the results of confusion matrix analysis. Recall that the areal confusion matrix relaxes the requirement for exact pixel matching. Consequently, it is not surprising that the summary statistics for the area-based confusion matrices are better than those for the pixel-based confusion matrices. For the 2009 map, the differences for all summary statistics are relatively small. However, for the 2007-9 change map, the difference depends on the statistic. The overall accuracy and user and producer accuracy for no change is (rounds to) 100% for both matrices. However, *kappa* is low for both. Taken together, it is clear that the summary statistics for the change map are dominated by the dominant no change class. Comparing summary statistics for pixel-based versus area-based confusion matrices provides insight into regional versus local map accuracy. Increasingly large sample units could also be considered to assess global accuracy. However, as with all confusion matrices but especially for areal confusion matrices, local accuracy cannot be inferred from global accuracy.

Table 2: Regression statistics for the amount of a given land cover on each areal sample for optical and radar maps. (See also Figure 2.)

Map	Adj. r^2	p^1	RMSE (ha)	Intercept[2]	Slope[3]
2009 Non-forest	0.36	0.00	1888.	922.	0.62*
2007-9 No Change	0.15	0.02	21.	8965.**	0.10**
2007-9 Deforestation	0.10	0.05	17.	7.	0.19**

[1] The p value is the overall significance of the relationship – i.e., a p value of 0.05 indicates a statistically significant relationship with 95% confidence.

[2] *(**) signifies that the intercept is significantly different from 0.0 with $p = 0.05$ (0.01).

[3] *(**) signifies that the slope is significantly different from 1.0 with $p = 0.05$ (0.01).
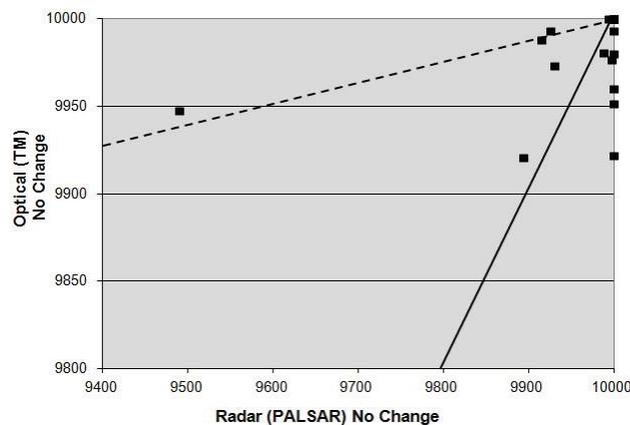

Figure 2: Amount of 2007-9 No Change for TM vs. Radar data. The dashed line is the ideal 1:1 line. The solid line is the actual regression line.

Table 3: Summary statistics for pixel- and area-based confusion matrices.

Map	Ma-trix[1]	Overall Accuracy (%)	Kappa	User Accuracy (%)			Producer Accu. (%)		
				Non-forest	Forest		Non-forest	Forest	
2009	**Pixel**	79	0.52	72	82		63	87	
	Area	85	0.66	82	86		68	85	
				NC[1]	Def[1]	Ref[1]	NC[1]	Def[1]	Ref[1]
2007-9	**Pixel**	100	0.02	100	4	0	100	2	0
	Area	100	0.29	100	39	53	100	25	16

[1]NC – No Change, Def – Deforestation; Ref – Reforestaiton.

5. Conclusion

Land cover change maps present many challenges for accuracy assessment. An area-based sample was employed here to overcome difficulties in obtaining conventional point-based data whose temporality matches that of the change map, and addressing issues related to the rareness of land cover change. As the archive for high resolution imagery expands, data availability may become less of an issue. Relative rareness of land cover change will remain problematic, however.

The associated difficulties will be affected by a number of factors. If land cover change is spatially concentrated, research will have to be undertaken to develop statistically adequate yet areally representative samples and associated analytical procedures that properly weight each observation; this is not a trivial task. Change maps that cover longer time periods may decrease the problems associated with rare classes because on such maps land cover change may not be rare. However, maps that cover longer periods may also have difficulty with land that changed from one land cover to another and then back to the original landcover.

Acknowledgments

References

Barson, M., Bordas, V., Lowell, K., Malanfant, K. (2004), "Independent reliability assessment for the Australian Agricultural Land Cover Change Project 1990/91-1995". *In: Lunetta, R., Lyon, J. (eds.). Remote Sensing and GIS Accuracy Assessment*, CRC Press, London, U.K.

Congalton, R., Green, K. (2009), *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. 2nd Edition*. CRC Press (Taylor and Francis Group), London, 183 pp.

Congalton, R., Macleod, R. (1994), "Change detection accuracy assessment on the NOAA Chesapeake Bay pilot study". *Proceedings: 1st International Symposium on the Spatial Accuracy of Natural Resource Data Bases*, Williamsburg, Virginia. pp. 78-87.

Li, B., Zhou, Q. (2009), "Accuracy analysis of multi-temporal land-ciover change detection using a trajectory error matrix". *International Journal of Remote Sensing,* Vol. 30(5):1283-1296.

Liu, H., Zhou, Q. (2004), "Accuracy analysis of remote sensing change detection by rule-based rationality evaluation with post-classification comparison". *International Journal of Remote Sensing* Vol. 25(5):1037-1050.

Lowell, K. (2001), "An area-based accuracy assessment methodology for digital change maps". *International Journal of Remote Sensing,* Vol. 22(17):3571-3596.

Lowell, K., Richards, G., Woodgate, P., Jones, S., Buxton, L. (2005), "Fuzzy reliability assessment of multi-period land-cover change maps". *Photogrammetric Engineering and Remote Sensing,* Vol. 71(8):939-945.

Pontius, G., Cheuk, M. (2006), "A generalized cross-tabulation matrix to compare soft-classified aps at multiple resolutions". *International Journal of Remote Sensing* Vol. 32(15):4407-4429.