A Regression Model Approach for Mapping Woody Foliage Projective Cover Using Landsat Imagery in Queensland, Australia.

Tim Danaher, John Armston and Lisa Collett
Queensland Department of Natural Resources, Mines and Energy
80 Meiers Rd Indooroopilly QLD Australia 4068
tim.danaher@nrm.qld.gov.au

Abstract—**This paper describes the development of a regression model for predicting foliage projective cover (FPC) using an extensive set of over 2000 field observations for Queensland, Australia. The model includes Landsat TM and ETM+ imagery and a climatological ancillary variable, vapour pressure deficit. The resulting model was validated using independent site data and preliminary validation against FPC estimates from airbourne laser scanner data is presented. Results suggest the model is robust and performing well over a range of soil types and vegetation communities. This regression-based methodology is currently included in the process of monitoring annual woody vegetation change over Queensland and will form the basis of new products for monitoring longer term trends in FPC.**

Keywords- Vegetation; Landsat; projective foliage cover

I. INTRODUCTION

The Statewide Landcover and Trees Study (SLATS) was initiated by the Queensland Government to map and monitor the change in woody vegetation cover over the State of Queensland, Australia, for vegetation management and greenhouse accounting objectives [1]. Since 1985, Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) satellite imagery have been analysed to regularly monitor the change in woody Foliage Projective Cover (FPC) over the 87 Landsat scenes covering the State and provide baseline landcover mapping.

The initial method used by SLATS to map 1991 FPC over Queensland was based on a polynomial relationship between woody FPC, the Normalised Difference Vegetation Index (NDVI), and Thematic Mapper Band 5 [2]. These models were developed and applied on a scene by scene basis using field site signatures from within and surrounding scenes. While the method worked well in most cases, it tended to be affected by the variation in soil brightness and colour, and in particular by areas of dark cracking clay soils. Fire scars also caused errors with woody FPC being overestimated. Image stratification approaches were trialed to overcome these variations, but the lack of stratification data at a scale consistent with Landsat TM data and the additional requirement for field data limited their use.

An empirical radiometric correction developed by [3] enabled production of a radiometrically matched time series of Landsat TM and ETM+ image mosaics. The application of this correction to the SLATS Landsat TM/ETM+ archive enabled signatures for over 1600 measured vegetation sites to be extracted and analysed as a statewide data set, providing a basis for developing an improved method. This improved method reported by [4] used a multiple regression approach incorporating Landsat TM bands and band interaction terms. It provided more accurate mapping of woody FPC without the need for image stratification.

Multiple regression is a common technique for estimating sub-pixel cover fractions in satellite imagery, however its application is often limited by a lack of field data for calibration, and radiometric, spatial and spectral uncertainties in remotely sensed imagery [5]. In the presence of representative calibration data, multiple regression has been shown to perform as well as more complex non-linear techniques such as regression trees and artificial neural networks [6] [7]. In this paper, the approach of [4] has been extended, and the limitations identified by [5] have been overcome, through improvements in image pre-processing and additional field data.

Independent validation is a critical component of any remote sensing project, however acquisition of independent field data to validate estimates of FPC is prohibitively expensive. Airbourne laser scanning (ALS) is a technology that has been shown to provide estimates of FPC with comparable levels of accuracy and bias to field measurements [8]. This study presents a preliminary analysis of recent ALS acquisitions for independent validation of FPC estimates over large areas.

II. DATA AND PRE-PROCESSING

The Landsat imagery was acquired with Australian Centre for Remote Sensing (ACRES) Level 5 processing and registered to an orthorectified base image mosaic based on differential GPS control and precision ephemeris Landsat 7 imagery. Approximately 530 Landsat TM and ETM+ scenes, acquired between 1987 and 2004 were used in this study. The Landsat TM and ETM+ imagery was standardised to enable the state-wide use of this site data, rather than use on a scene by scene basis. This standardization included the removal of onboard radiometric calibration for Landsat 5 TM and replacement with a vicarious calibration based on a model of the lifetime response of the sensor [9], and application of an empirical radiometric correction [3] to standardise for surface reflectance anisotropy and some atmospheric effects. Preflight calibration was used for Landsat 7 ETM+ [10].

A calibration dataset of approximately 1600 vegetation sites distributed over Queensland was assembled via an extensive field campaign from 1996 to 1999. Sample sites were chosen to represent the equivalent to a block of 3x3 TM pixels resampled to a resolution of 25m. The sites were typically located in relatively homogeneous, mature stands over a variety of vegetation communities on various soil types, and covering a range of tree basal areas from very sparse open woodland communities to dense rainforest. Each tree basal area was measured from an average of 5 recordings, using a calibrated optical wedge [11]. In addition to these basal area measurements, woody FPC was measured at a subset of sites, using gimballed crosshair sighting tube observations along 100m or longer transects. The position of all sites was measured using differential GPS [2].

Approximately 900 additional sites with zero woody FPC were chosen to represent the variation in background soil colour and recent fire scars not represented in the site data. This number of non-woody sites is consistent with the proportion of Queensland without woody vegetation cover [12]. The selection of these sites was based on either aerial photo interpretation or field observations.

To evaluate the accuracy of the FPC model a program of laser scanner data acquisition was commenced. To date, 10 transects have been acquired by AAM surveys using an Optech ALTM3025 laser scanner in woodland and sparse woodland communities. Each of the transects were 10-15km and approximately 300m wide with average sample spacing of 0.95m and a footprint of 0.22m. Future laser scanner data acquisitions will target forests where the FPC is higher.

Field measurements of basal area and FPC were taken within a month of the laser scanner data acquisition. Typically three sites consisting of 300m of FPC transect and seven basal area recordings were measured within each transect. The position of these sites was measured using sub-metre differential GPS.

III. DATA ANALYSIS

The 200 site measurements with coincident FPC and basal area measurements were used to calibrate a stand allometric equation between overstorey FPC and basal area similar to that developed by [12]. This relationship (r^2=0.824 s.e.=9.68 p=<0.001) is given in Eqn. 1.

$$FPC = \left(1 - e^{-0.03546578 * BA}\right) \times 100 \qquad (1)$$

Signatures for Landsat bands 1-7 were extracted for the 3x3 pixel mean surrounding the field site location. The 3x3 pixel mean was chosen to best represent the area sampled by field measurements and to reduce the effects of geometric mis-registration between imagery of different dates. Site signatures were extracted for 1991, 1995 and 1997 dates for Landsat 5, and 1999, 2000, 2001 and 2002 dates for Landsat 7. The signatures were filtered to remove those affected by cloud or clearing in any image date. It was assumed that any increase in basal area between the date of site measurement and the image acquisition date was less than the error associated with the basal area observations as sites were generally located in mature vegetation. The site data set was divided using a random stratified sampling approach, with 2440 observations used for model development and 260 observations or ~10 percent of the data set used for validation.

Goulevitch et al. [4] experimented with many combinations of published vegetation indices and bands but found that the combination of transformed Landsat band values and interaction terms e.g. band2×band3 provided superior results. As a result, vegetation indices were not considered in this study. Signatures for band 1 were not included in the analysis as the scene to scene differences in radiometrically corrected data were significantly higher than for all other bands [3]. A climatological ancillary variable, Vapour Pressure Deficit (VPD), was also incorporated into the regression model because the evaporative potential of the atmosphere strongly influences FPC [13]. The VPD data used were long term averages.

A regression model of percentage woody FPC was developed using best subsets multiple linear regression. Band 2-7 values, interaction terms, log, reciprocal and square root transformed band values, and interactions of transformed terms were considered. The models based on transformed band values produced similar results and all were better than those based on the original band values. The analysis tested the use of individual year, two and three year mean signatures for both TM and ETM+. Ultimately the three year means provided the best results, most likely due to the smoothing of phenological and atmospheric variations. Separate models were developed for Landsat 5 TM and 7 ETM+ due to known differences in bandwidth and spectral response of these sensors [10].

The criteria used for model selection included the adjusted r^2, standard error, the Mallows Cp statistic and the P value for individual model terms. The results for the selected models are summarized in Table 1. All terms in the models were significant ($p<0.001$). There was no significant relationship between the residuals and measured FPC.

TABLE I. RESULTS OF REGRESSION MODEL FITTING FOR LANDSAT 5 TM AND LANDSAT 7 ETM+ SENSORS

Sensor	n. vars	r^2	s.e.	Cp	Transformation	Variables
TM	13	0.82	7.9	15.2	square root	VPD, 2, 5, 7, 2x3, 2x4, 3x4, 3x5, 3x7, 4x5, 4x7, 5x7
ETM+	12	0.79	8.5	13.5	reciprocal	VPD, 3, 4, 5, 7, 2x4, 2x5, 3x4, 3x7, 4x5, 4x7

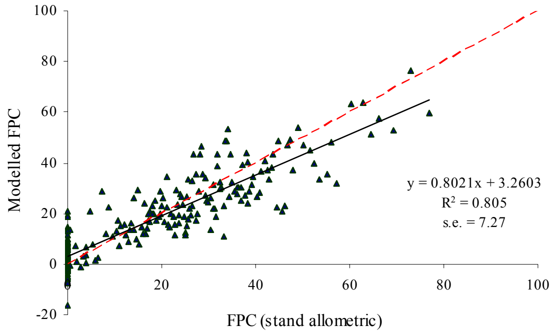


Figure 1. Stand allometric FPC vs. predicted woody FPC

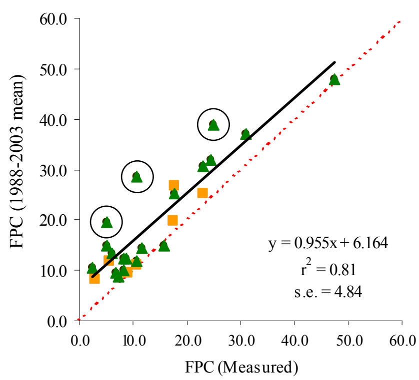


Figure 2. Measured FPC vs. modelled FPC

IV. VALIDATION

A. Site data validation

Initial validation of the model was done using the 260 site measurement not used in the model development. Two tests were done. Firstly, best subsets regression was used with the validation data to determine another regression model. This process selected the same terms selected as in the model development and the resulting model was not significantly different to the original model. The data was also used to compare FPC derived from the stand allometric equation (Eqn. 1) with FPC predicted using the model developed with 90 percent of the site data (Fig. 1). This comparison showed that the model is robust with correlation coefficient and standard error similar to the original model fit. However, Fig. 1 indicates that the model slightly over-predicts FPC near zero and under-predicts it at high FPC.

An independent set of field FPC measurements taken to calibrate the laser scanner data, provided another validation data set. In Fig. 2, the modeled FPC was compared to these field measurements. The sites located on predominately red soil are shown as orange squares and green triangles indicate tan soil colour. The three circled points in Fig. 2 were sites with observed *Triodia spp.* (spinifex) understorey.

There is a high correlation between observed and predicted data, however, the model is again shown to overpredict FPC. There seems to be no significant bias in modeled FPC for red and tan coloured soil sites. The over-prediction of FPC for the spinifex sites is expected due to its perennial green nature.

B. Effects of fire

A preliminary qualitative analysis aimed to determine the sensitivity of the FPC model to fire events of varying intensity. The analysis used coincident field measurements collected over a Landsat-5 scene (23 Oct 1999) for another study [14]. Field measurements included ground transects across fire scars where differential GPS position, burnt/un-burnt occurrence, fallen tree litter, scorch height, tree basal area, and other miscellaneous data were recorded at 1m intervals. These observations were clumped into 100m sites and the corresponding woody FPC model values were extracted for 3x3 pixels around the site centre on the coincident Landsat 5 image. Measured percent burnt, average scorch height, percent fallen scorched litter, and fire affected tree basal area, together with a qualitative assessment of the site photos, were used as a proxy for intensity. These measurements are summarised in Table 2.

Fig. 3 shows the variation in modelled FPC over the period 1995 to 2002. For sites 3 and 4, the leaf litter cover, fire affected tree basal area and scorch height indicated that the fire intensity was low and although the pasture was completely burnt, there was relatively little impact on the FPC. The fires at sites 2 and 5 show increasing impacts on the modeled FPC due to increasing disturbance of the canopy as measured in the field. At site 1, the field observations indicate presence of a more intense fire with greater canopy reduction. This change is very visible in the modelled FPC time series for this site.

Although the firescar at site 1 is very evident on the Landsat TM image (Fig. 4(b)) it has not had a significant impact on the modelled FPC shown in Fig. 4(a). A photo of site 1 (Fig. 4(c)) shows that there has been a significant decrease in green foliage due to leaf scorching but much of the canopy is still intact, accurately reflecting the FPC image in Fig. 4(a).

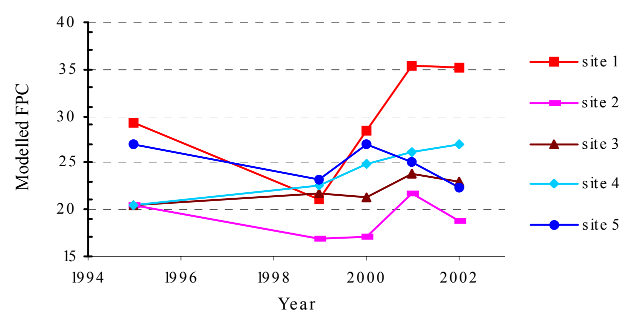
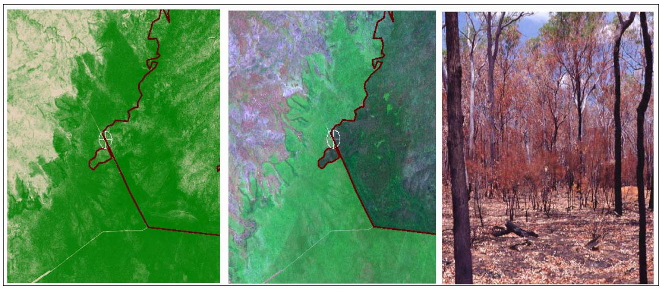


Figure 3. Modelled FPC time series for firescar field sites



(a) (b) (c)

Figure 4. Sensitivity of Woody FPC model to fire scars (a) Woody FPC image (b) Landsat TM composite image 2,4,5 BGR showing ground photo location in centre (c) ground photo close to overpass date at site 1.

TABLE II. FIRE SCAR SITE OBSERVATIONS

Site	Live FPC (%) (allometric)	Fire Aff. FPC (%) (allometric)	Intensity class
1	Not measured	Not measured	High
2	21.0	9.5	Mod
3	26.0	6.8	Low
4	23.9	0	Very Low
5	26.7	9.5	Mod - High

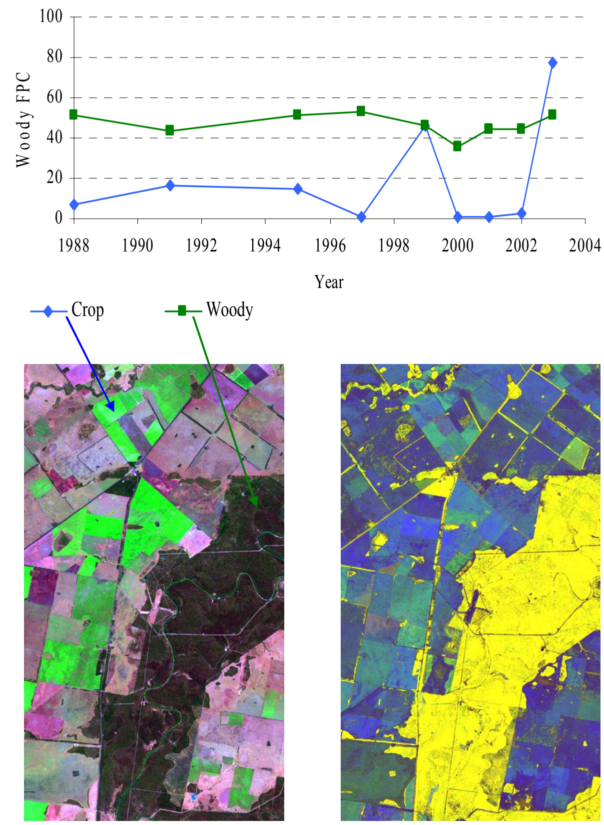
C. *Effects of crop and green pasture on model*

An obvious limitation of the FPC model was the inability to distinguish between crop or green pasture and the woody vegetation components of total FPC, for every individual image. However, this limitation was overcome by calculating the FPC for the entire time series of imagery and classifying woody vegetation based on the coefficient of variation (CV) (Eqn. 2) of the FPC time series.

$$CV = \sigma / \mu \qquad (2)$$

σ is standard deviation and μ the mean of the FPC time series. As most of the SLATS imagery was acquired during dry seasons, intermittently green areas exhibit high CV in woody FPC, whereas constantly woody vegetation has lower CV in woody FPC. This is demonstrated in Fig. 5 where the woody areas with low variation in FPC in are shown as yellow, and do not include the crop areas shown as woody in 2003, in Fig. 5(c).

(a)



(b) (c)

Figure 5. (a) Landsat ETM+ image for 2003 (b) Woody FPC model time series image for 1988-2003 showing coefficient of variation of woody FPC, mean woody FPC and minimum woody FPC as BGR, with undistrubed woody vegetation showing as yellow.

D. *Laser scanner validation*

The resulting classification was validated using independent woody FPC estimates derived from ALS. These data were processed using a progressive morphological filter [15]. Due to the sensors footprint size (0.22m), ALS tends to overestimate true FPC [8]. For the purpose of comparison in this study, a simple calibration approach was applied where the slope of the linear regression line forced through the origin for the field-ALS FPC relationship was adjusted to 1:1 (r^2 = 0.7815, s.e. = 5.548) (Eqn 3).

$$FPC_{measured} = 0.5365 \times FPC_{ALS} \qquad (3)$$

FPC was calculated as the proportion of ALS data points 2m above the ground [8] at 75m spatial resolution. This was the spatial resolution used for the extraction of Landsat signatures. An example comparison of ALS and Landsat modelled FPC for the Dirranbandi region in southern Queensland is shown in Fig. 6.

There is a high correlation between ALS and Landsat modelled FPC (r^2 0.80, s.e. 2.94) for this site as shown in Fig. 6. The calibration equation (Eqn. 2) may need further refinement to enable use in regions with different vegetation communities. Furthermore, the 2m height threshold may have excluded a small proportion of woody FPC.

This may, in part, explain the deviation from a 1:1 line for this example. Current results are encouraging, however further research is required to link the independent ALS and Landsat estimates of FPC.

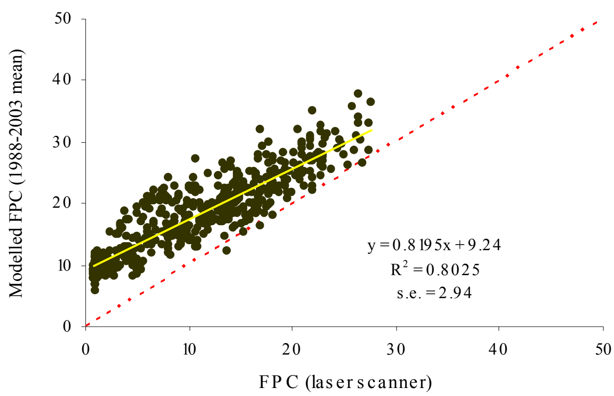


Figure 6. Comparison of ALS and modelled FPC

V. Conclusions and Future Research

Preliminary validation of the FPC model suggests multiple regression is a robust technique for this application. Further acquisition and analysis of laser scanner data is underway and will further add to the validation effort and other studies of vegetation structure conducted by SLATS. Results suggest that the regression model is performing well in areas with varying soil colour and fire-scar backgrounds, and that image stratification is not necessary as the regression model appears to compensate for these factors. The results highlight the importance of using multi-date imagery when mapping woody FPC.

Several improvements to this method are being considered, including; the use of alternative regression techniques such as reduced major axis regression [16], improvements to radiometric correction to enable use of band 1 data and correction of topographic effects in all bands.

The FPC model is currently being used in the process of mapping annual FPC change over Queensland. New products incorporating the FPC model are being developed for monitoring longer term trends in vegetation such as thinning, dieback and thickening.

Acknowledgments

The authors would like to thank David Mayer of DPIF for assistance with statistical analysis that made this work viable, Rob Hassett and Sel Counter for the measurement of field validation data and the entire SLATS project team who have all contributed to data collection and processing of imagery used in this work.

References

[1] SLATS, "Statewide landcover and trees study," http://www.nrme.qld.gov.au/slats/index.html, 2004.

[2] C. Kuhnell, B. Goulevitch, T. Danaher and D. Harris, "Mapping woody vegetation cover over the state of Queensland using Landsat TM imagery," *Proc. Ninth Australasian Remote Sensing and Photogrammetry Conference*, Sydney, Australia, 1998.

[3] T. Danaher, "An empirical BRDF correction for Landsat TM and ETM+ imagery," *Proc. Eleventh Australasian Remote Sensing and Photogrammetry Conference*, Brisbane, September 2002.

[4] B. Goulevitch, T. Danaher, A. Stewart, D. Harris and L. Lawrence, "Mapping woody vegetation cover over the State of Queensland using Landsat TM and ETM+ imagery," *Proc. Eleventh Australasian Remote Sensing and Photogrammetry Conference*, Brisbane, September 2002.

[5] R. Salvador and X. Pons, "On the reliability of Landsat TM for estimating forest variables by regression techniques: a methodological analysis," *IEEE Trans. Geosc. Rem. Sens.*, vol. 36, no 6, Nov 1998.

[6] R. DeFries, M. Hansen, M. Steininger, R. Dubayah, R. Sohlberg and J. Townshend, "Subpixel forest cover in Central Africa from multisensor, multitemporal data," *Rem. Sens. Env.*, vol. 60, pp. 228-246,1997.

[7] R. Fernandes, R. Fraser, R. Latifovic, J. Cihlar, J. Beaubien, and Y. Du, "Approaches to fractional land cover and continuous field mapping: A comparitive assessment over the BOREAS study region," *Rem. Sens. Env.*, vol. 89, pp. 234-251, 2004.

[8] D. Weller, R. Denham, C. Witte, C. Mackie, D. Smith, "The assessment and monitoring of foliage projected cover and canopy height across native vegetation in Queensland, Australia using laser profiler data," *Canadian Journal of Remote Sensing*, vol. 29, pp. 578-591, October 2003.

[9] C. deVries, T. Danaher and P. Scarth, "Calibration of multiple Landsat sensors based on pseudo-invariant target sites in Western Queensland, Australia," *Proc. IGARSS 2004*, Anchorage, 2004.

[10] NASA, "Landsat 7 Science data users handbook," http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_toc.html, 2004.

[11] J. Dilworth and J. Bell, *Variable probability sampling - variable plot and three P*, OSU Book Stores, Corvallis, Oregon, pp. 130, 1971.

[12] T. Danaher, J. Carter, K. Brook, A. Peacock and G. Dudgeon, "Broad-scale vegetation mapping using NOAA-AVHRR imagery," *Proc. Sixth Australasian Remote Sensing Conference*, Wellington, New Zealand, pp.128-137, 1992.

[13] R. Specht and A. Specht, *Australian plant communities dynamics of structure, growth and biodiversity*, Oxford University Press, pp. 249-262, 1999.

[14] L.Collett, R. Hassett, C. Taube, N. Flood and W. Hall, "Validation of NOAA firescar maps in the Laura Basin using Landsat TM imagery and field data, Report to Tropical Savannas CRC," Queensland Department of Natural Resources and Mines, 2001.

[15] K. Zhang., S. Chen, D. Whitman, M. Shyu., J.Yan & C. Zhang, "A progressive morphological filter for removing nonground measurements from airborne LIDAR data," *IEEE Trans. on Geosc. Rem. Sens.*, vol 41, no 4, pp. 872-882, 2003.

[16] W.B. Cohen, T.K. Maiersperger, S.T. Gower and D.P. Turner, "An improved strategy for regression of biophysical variables and Landsat ETM+ data", *Rem. Sens. Env.*, vol. 84, pp. 561-571,2003.