DERIVATION OF A PERENNIAL VEGETATION DENSITY MAP FOR THE SOUTH-WEST REGION OF WESTERN AUSTRALIA

Joanne Chia (1) **Graeme Behn** (2) **Dave Bebbington** (2) **Peter Caccetta** (1)

(1) Remote Sensing and Image Integration
CSIRO Mathematical and Information Sciences
65 Brockway Road, Floreat, W.A 6014
Phone: +61 8 93333 6138, Fax: +61 8 9333 6121
Joanne.Chia@csiro.au

(2) Remote Sensing
Department of Conservation and Land Management
65 Brockway Road, Floreat, W.A 6014
Phone: + 61 8 9333 6271, Fax: +61 8 9333 6121
graemeb@calm.wa.gov.au

Abstract

The Land Monitor Project has used sequences of Landsat imagery to provide maps of the extent and change in area of perennial vegetation for the south-west agricultural region of Western Australia (W.A). Maps of change over time within perennial vegetation are also produced using trends in a spectral index which has been shown to be sensitive to differences in vegetation density/condition over many vegetation types in the region. The data are widely used to inform land management, conservation and biodiversity issues. In this paper, we describe a recent study performed to estimate the vegetation density of the south-west region of W.A using ground data, and the year 2000 Land Monitor Landsat TM mosaic. The region was first stratified into broad zones reflecting variations in soil and vegetation. Density 'ground truth' information was derived from aerial photo interpretation acquired close in time to the Landsat TM mosaic. Interpretation was done by a single operator for 593 sites across the region; sites were selected as far as possible to represent a range of densities and vegetation types within each zone. The relationships between zones, 'ground' estimates of vegetation density, and the Landsat TM data were examined using regression tree techniques as implemented in the program known as 'Cubist', producing a map of vegetation density for the south-west region of W.A. This paper describes various aspects of the study.

1. Introduction

The Land Monitor project (Allen and Beetson, 1999) is an on-going collaborative project involving W.A state government and federal agencies. Initially, it was supported by the Natural Heritage Trust and state agencies, and its main aim was to provide information to farmers, land managers and administrators on extent and change of salt-affected lands, and perennial vegetation cover in the south-west agriculture region of Western Australia. The project has used digital elevation models, ground truth measurements and sequences of Landsat TM data to produce data and maps that allow the mapping and monitoring of both the salt-affected lands, and the extent and change of perennial vegetation cover over time. Here we consider the quantitative estimation of vegetation density for biodiversity and water management issues.

Various vegetation density (also known as tree canopy density) estimation techniques are available in the literature. A list of recent methods together with their strengths and weaknesses are noted in (Huang et al., 2001). Common methods include the use of linear models to approximate the relationships between spectral signal and canopy density. However, such relationships can get complex, especially if the estimation is performed over large areas. In such cases, a single linear model will not be adequate to perform the estimation. A study carried out by (Huang and Townshend, 2001) has described that using regression tree technique which involves a set of linear models produces more accurate estimates when approximating complex relationships. A number of studies have used regression tree techniques to successfully handle large data sets and model complex relationships (see e.g Huang et al., 2001 and Xian et al., 2002).

In this paper, we present a recent study carried out to estimate the vegetation density in the south-west region of W.A using regression tree technique. The estimation was performed using the year 2000 Land Monitor Landsat TM mosaic to produce a map of vegetation density for the south-west region of W.A.

In the next section, we describe the methodology adopted for the estimation process. This includes description of the strategies taken to choose training sites and to process the data for estimation. Then we show examples of the results taken from the final map, and discuss the results obtained.

2. Methodology

The overall approach proposed in this paper consists of three steps as follows.

1) The study area is stratified into zones reflecting variations due to factors including vegetation types, geology, soil and rainfall.

2) Ground measurements are selected using the stratification as a guide so that they represent a range of densities and vegetation types within each zone.

3) Apply regression tree technique to the training and Landsat TM data to estimate vegetation densities.

In the ensuing paragraphs, we describe each of these steps.

2.1 Area under study

The region covered by the datasets in the Land Monitor project is the south-west region of W.A as shown in Figure 1. The region extent is covered by 18 Landsat scenes. For this study, we used the Landsat 2000 mosaic compiled using 18 rectified and calibrated scenes drawn from the Land Monitor archive.

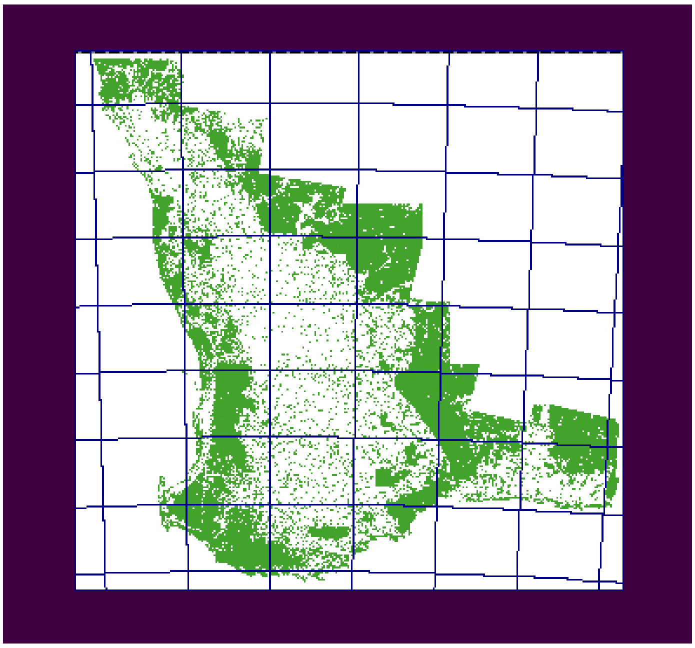

Figure 1. The Land Monitor project extents as depicted by the coverage of vegetation products for the south west region of WA. The blue grid represents standard 1:250000 mapsheets.

2.2 Stratification of study area

The stratification of the south-west region of W.A used in this study was based on that developed for the Australian Greenhouse Office Project (Furby, 2002). The region was divided into different zones reflecting variations in vegetation, geology, soil and rainfall. Figure 2 shows the stratification.

Figure 2. The south-west region of W.A with stratified zones overlay on it. The text within each zone represent the zone number. The TM image is in enhancement of bands 5,4 and 2 having colours red, green and blue respectively.

2.3 Training sites

In this study, we used aerial photography interpretation (Behn et al., 2001) to derive training data consisting of Canopy Cover estimates at selected sites.

Interpretation of sites was done by a single operator. Within each zone, sites were selected as far as possible to represent a range of densities and vegetation types. A total of 593 sites were chosen across the region. No training site was sampled in zones where aerial photo was not available. The zones that have no sites chosen were zones 3, 5, 7, 3, 13, 15, and 19. For analysis purposes, each of these zones was combined with the nearest zone where training sites were available.

For each sampled site, the zone number, vegetation type (non-forest, medium forest, low-forest or plantation), and the estimated vegetation density as a percentage were recorded.

2.4 Estimating vegetation density using regression trees

The theory of regression tree is developed by (Breiman et al. 1984). Briefly, regression trees can be thought of as a piecewise estimate of a regression function where the response variable is continuous. The tree is established by recursively partitioning the data and sample space. The regression tree technique is particularly useful when the predictors may be associated in some non-linear fashion. Some of the advantages of this technique include

(i) it is a non-parametric method and thus it requires no specification of the functional form

(ii) there is no requirement for assumption of a linear model or need to specify a prior probability distribution for the errors

In this study, we applied the regression tree method as implemented in a program known as “Cubist” to data, where the explanatory variable is the density estimates, and the predictors are the spectral signals and zone numbers. In short, Cubist generates models that are based on rules, where each rule is made up of conditions associated with a linear expression. A Cubist model can thus be thought of as a piecewise linear model. More information on Cubist can be found in (Rulequest 2004).

The program Cubist was run several times using data from different partitioning of zones in the study area. From the applications of Cubist, two models based on using data from zones 1, 2, 3, 10, 11, 25 and zones 6, 61, 4, 8, 9, 12, 81 were chosen. These two models will be known as model I and model II from this point onwards.

The performance of models I and II, and the results of the final vegetation density map illustrated using three examples taken from the map will be discussed in the next section.

3. Results and discussion

The performance of the two Cubist models I and II that were generated to estimate the vegetation densities could be measured by the average error magnitude, relative error magnitude and the correlation coefficient. The three quantities were calculated by Cubist.

The average error magnitude of the two models are 11.8% and 12.4% respectively. The relative error magnitude is the ratio between the average error magnitude and the error magnitude results from always predicting the mean value. This value should be less than 1 for models to be useful for prediction. The relative error magnitude for Cubist models I and II are 0.59 and 0.66 respectively. The correlation coefficient is a measure of the agreement between the actual values of the explanatory variable and those predicted by the model. That is, a correlation coefficient of value 1 indicate total agreement. The correlation coefficients of the two models are 0.76 and 0.72 respectively, indicating that the models explain roughly half the variability of the ground estimates.

In the program Cubist, there is a function known as *f*-fold cross-validation that can be used to get a more reliable estimate of the predictive accuracy of the model constructed (Rulequest, 2004). In short, the method involves partitioning the training data into *f* number of blocks of roughly equal size. For each block *i* (*i* = 1, …, *f*), a model is built from the data in the remaining blocks. The model is then applied to the data in block *i*, and the value of the average error magnitude is computed. The mean value of the average error magnitudes from all the *f* blocks gives an estimate of the predictive accuracy of the model constructed by all the data.

It is also noted in (Rulequest, 2004) that the above process can be repeated for a specified number of times. The average error from the distinct cross-validation trials then gives an even more reliable estimate of predictive accuracy of the model built by all the training data.

In our case, we applied 10-fold cross-validation trials with 10 repetitions to assess the predictive accuracy of models I and II. The average error magnitudes for both models are 14.3% and 13.8% respectively, relative errors are 0.73 and 0.72 respectively. The correlation coefficients are 0.64 and 0.62 respectively which means the r-squared values are 0.41 and 0.38, indicating that about 40% of the variation of the ground estimates is explained by the models.

The final vegetation density map (1st pass) is shown in Figure 4. The ground density estimates and the corresponding predicted densities are plotted and the graph is displayed in Figure 5. The plot shows that much variation in the predicted site values remains. The source(s) of this error could include sensor limitations, photo interpretation error, insufficient stratification and the fitting method.

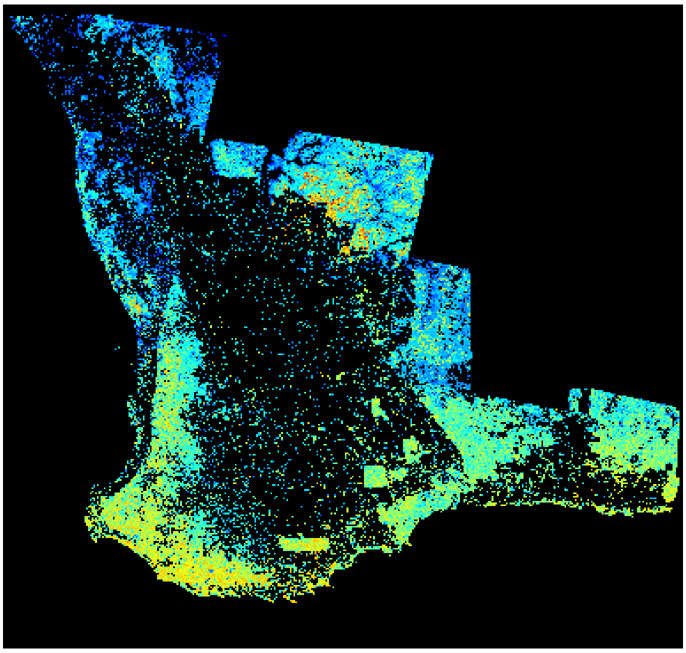

Figure 4. Vegetation density map (1st pass) of the south-west region of Western Australia. The vegetation densities are represented by colours where blue to red indicate low to high densities. Areas in black represent non-vegetation area.

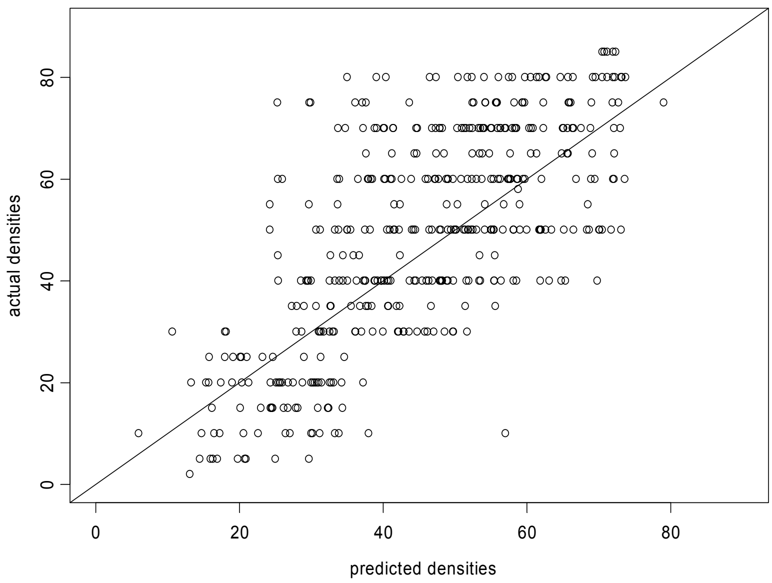

Figure 5. Plot of ground density estimates (“actual densities”) against predicted densities with the line y=x superimposed on it. Muxh variation in the predicted values remain.

In Figures 6(a), (b) and (c), we show three examples taken from the final vegetation density map. Each example is displayed with its corresponding Landsat TM data for comparison purposes.

In example 1 (Figure 6(a)), there is a fire scar located near the top left hand corner of the picture. It can be observed that the scar has been mapped appropriately with low density estimates of about 10% - 20%. The rest of the area in this example consists mainly of rather dense vegetation, and it can be seen that the vegetation has been mapped with density estimates with value up to 70%.

Example 2 (Figure 5(b)) shows an area of where vegetation is thin. It is observed in the map that the vegetated areas have all been mapped appropriately with low density estimates between 10% - 50%.

Example 3 (Figure 5(c)) displays an area that is mainly made up of blocks of vegetated areas of varying densities. The vegetation densities in this region have been predicted to be between 10% - 65%. It is observed that areas of dense and less dense vegetation have been mapped correctly.

Example 1

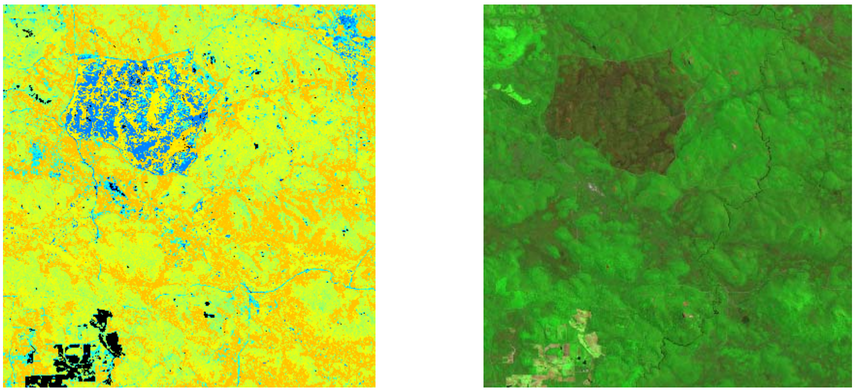

Figure 6(a) The picture on the left shows the estimated vegetation density map in the area where blue – orange colours represent low – high densities while black colour indicate non-bush, and has value zero. The estimated vegetation densities in this area range between 10% - 70%. The picture on the right is the corresponding TM image in the enhancement of bands 5, 4 and 2 having colours red, green and blue respectively.

Example 2

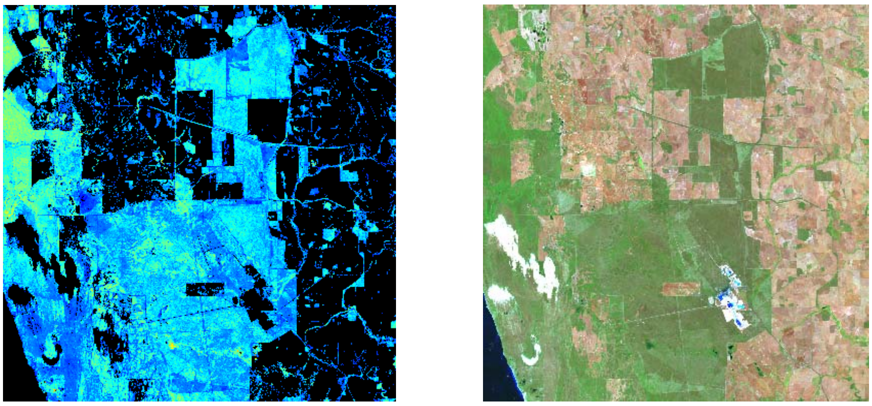

Figure 6(b) Left hand side picture shows the estimated vegetation density map in the area where blue – yellow colours represent low – high densities densities while black colour indicate non-bush, and has value zero. The estimated vegetation densities in this area range between 10% - 50%. The right hand side picture is the corresponding TM image in the enhancement of bands 5, 4 and 2 having colours red, green and blue respectively.

Example 3

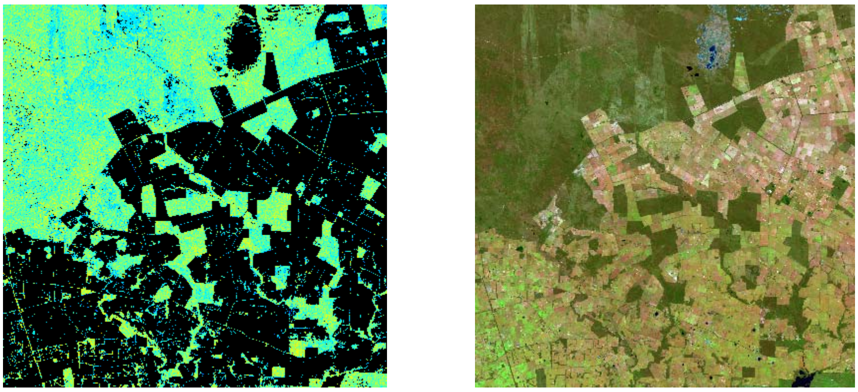

Figure 6(c) Left hand side picture shows the estimated vegetation density map in the area where blue – yellow colours represent low – high densities densities while black colour indicate non-bush, and has value zero. The estimated vegetation densities in this area range between 10% - 65%. The right hand side picture is the corresponding TM image in the enhancement of bands 5, 4 and 2 having colours red, green and blue respectively.

4. Summary

An initial approach has been developed to estimate the vegetation cover density in the south-west agricultural region of W.A. This approach uses the regression tree technique as implemented in the program known as "Cubist" to estimate the vegetation densities. As Cubist has no spatial knowledge of the area under study, it relies on good stratification of the region, in particular, the stratification should reflect the different types of vegetation, and good ground measurements that include a range of densities and vegetation types within each stratified zones in order to produce accurate estimates.

Using the supplied ground density estimates, Landsat data and zone numbers as data, two Cubist models were chosen to estimate the vegetation densities of the south-west region of W.A. The average error magnitude of the models lies between 13% - 14%, and the r-square value is between 0.38 – 0.41, indicating that the models describes about 40% of the variability of the ground estimates.

5. References

Allen, A., and Beetson, B. (1999) The Land Monitor Project; A Multi-agency project of the Western Australian Salinity Action Plan supported by the Natural Heritage Trust. Proceedings of WALIS Forum 1999, Perth, WA, March 1999, pp. 74-77.

Behn, G.A., McKinnell, F.H., Caccetta, P., Vernes, T. (2001) Mapping forest cover, Kimberley region of Western Australia. *Australian Forestry* Vol. 64, 80-87.

Breiman, L., Friedman, J.H., Olsen, R.A., and Stone, C.J. (1984) Classification and Regression Trees. Wadsworth, Belmont.

Furby, S., (2002) Land Cover Change: Specification for remote sensing analysis, technical report no. 9, Australian Greenhouse Office.

Huang, C. and Townshend, J.R.G. (2001) A stepwise regression tree for nonlinear approximation: applications to estimating subpixel land cover. International Journal of Remote Sensing, in press.

Huang, C., Yang, L., Wylie, B., and Homer, C. A. (2001) A strategy for estimating tree canopy density using Landsat 7ETM+ and high resolution images over large areas. Proceedings of the Third International Conference on Geospatial Information in Agriculture and Forestry, Denver, Colorado.

Rulequest (2004). “Data mining with Cubist”, http//www.rulequest.com/cubist-info.html

Xian, G., Zhu, Z., Hoppus, M. and Fleming, M. (2002) Application of decision-tree techniques to forest group and basal area mapping using satellite imagery and forest inventory data. Proceedings of I/FIEOS Conference, Percora 15/Land Satellite Information IV/ISPRS Commission.