DERIVATION OF PLANTATION TYPE MAPS

J. Chia, P. A. Caccetta, S.L. Furby and J.F. Wallace
CSIRO Mathematical and Information Sciences, Private Bag 5 PO Wembley Western Australia 6913
Joanne.Chia@csiro.au, Peter.Caccetta@csiro.au, Suzanne.Furby@csiro.au, Jeremy.Wallace@csiro.au

KEY WORDS: Forestry, Landsat, multi-temporal, classification

ABSTRACT:

The focus of the Australian Greenhouse Office Land Cover Change Program has been on the identification of the timing and extent of land cover change (to or from forest) as input to the carbon modelling process used to calculate emissions. In areas where new forest growth has been identified, the carbon modelling can be improved by using growth data specific to the new forest type. The forest categories of interest are hardwood and softwood plantations, environmental planting and native regrowth. Existing available GIS data for forest / plantation class is of variable quality, inconsistent within and between states, irregularly updated and often at inappropriate spatial scales for integrating with the forest cover mapping. Excellent data is held privately for some areas by forestry related businesses. Nationally consistent forest type classifications are simply not available at an appropriate spatial and temporal scale for use in the carbon modelling.

Research was undertaken to examine whether an operationally feasible methodology for labelling the type of new forest cover could be developed using the existing time series of Landsat imagery. At local scales, conventional classification of single date imagery can produce plantation type maps, however they do not extrapolate well across space and time. This paper describes how the time series was analysed to investigate the spectral growth patterns of different plantation types to identify the strategy to best separate them from each other and from native regrowth. At different times during growth the different forest types are separable. A compositing technique, incorporating change flags from the Land Cover Change mapping, is used to integrate age information into a single classifier that is feasible to apply operationally to derive a national product. The derivation of the methodology is discussed and some results from the national mapping program are presented.

1. INTRODUCTION

In the period 2000-2002, the Australian Greenhouse Office (AGO) created a project called *Land Cover Change Program (LCCP)* which aimed to quantify the change in forest cover for the Australian continent using time series remotely sensed Landsat satellite data. The project initially rectified, calibrated and classified eleven national (time) coverages of Landsat data to create a time series of land cover classifications having the classes *forest* and *non-forest*. There are now fourteen coverages in the ongoing program. The time series classifications form part of the National Carbon Accounting System (NCAS), which tracks emissions (sources) and removals (sinks) of greenhouse gases from Australian land based systems.

Models within the NCAS produce different carbon outcomes for a number of plantation types. In areas where new forest growth has been identified, the carbon modelling can be improved by using growth models specific to the new forest type (hardwood or softwood plantation, environmental planting or native regrowth). However, existing GIS data for forest / plantation class is of variable quality, is inconsistent within and between states, is irregularly updated and is often at inappropriate spatial scales for integrating with the forest cover mapping. That is, nationally consistent forest type classifications are not available at an appropriate spatial and temporal scale for use in the carbon modelling.

This paper describes research conducted to examine whether an operationally feasible methodology for labelling the new forest cover could be developed using the time series of Landsat imagery obtained as part of the LCCP. Section two describes the issues that arise from conventional classification of single date imagery. Section three describes the analysis of spectral growth patterns of the different forest types in the time series data. In sections four and five, the results of the investigations are translated into an operational mapping system. Section 6 show results achieved by the compositing method using same test area of study in section 1. Section 7 ends with discussion and conclusions.

2. SINGLE-DATE CLASSIFICATION

In this section, we describe a study conducted to investigate the spectral discrimination and classification of various plantation types of interest into classes using single date Landsat data. Details of this study can be found in (Caccetta and Chia, 2002). Here, we give a brief summary of the study, highlighting the main points.

The study area covers the 1:1million map sheet SI50 in Western Australia, and the plantation types of interests in this area are “karri”, “jarrah”, “hardwood” and “softwood” plantations. The overall methodology included the following steps

- Training sites from areas of karri, jarrah, softwood and hardwood were selected from a Landsat TM image based on ground-truth information;
- Canonical variate analysis was used to examine the spectral separation between the groups of training sites;
- A maximum likelihood classification was performed using results from the spectral analysis;
- The accuracy was assessed using independent validation data.

2.1 Data

The 2000 Landsat TM data for map sheet SI50 south west quadrant from the AGO image archives was used. Ground data of plantation extent (for 2002) were made available to us by the Department of Environment and Conservation. Approximately 230 training sites were drawn from areas of karri, jarrah, hardwood and softwood plantations in the 2000 Landsat image. Some of the selected sites of each type were bare as the more recent plantations had not yet grown sufficiently to be detectable in the imagery.

2.2 Spectral Analysis

Canonical variate analysis (CVA) (Campbell and Atchley, 1981) was performed to investigate the spectral separability of the four plantation types under study, and the potential grouping into spectral classes for classification. A canonical variate analysis provides a transformation of the data that maximises between class separation relative to within class variability. Canonical roots provide a measure of separability. Full listings of the results achieved can be found in the appendix section of (Caccetta and Chia, 2002).

Several CVA analyses were performed. In the first run using all sites, canonical roots yielded were {10.65, 3.695, 2.869, 1.411, 0.6616, 0.3344}. This shows that most of the variation between these sites is contained in the first three dimensions and there is a reasonable overall level of separation. Figure 1 shows the plot of the first two canonical variates after the "bare" sites were removed. It shows a tendency of the softwood sites to cluster together. There is some clustering of the other sites but also a lot of overlap. From this analysis, a number of sites were grouped into classes belonging to "softwood mature", "karri green", bare (karri), bare (hardwood), bare (softwood) for classification. The aggregated sites were removed and a CVA was performed on the remaining sites. There was little spectral separation between the different plantation types, and no apparent natural grouping of sites into classes was visible. Thus, directed contrasts for each class pair combination were performed so as to have a closer examination of the separation between the different plantation types.

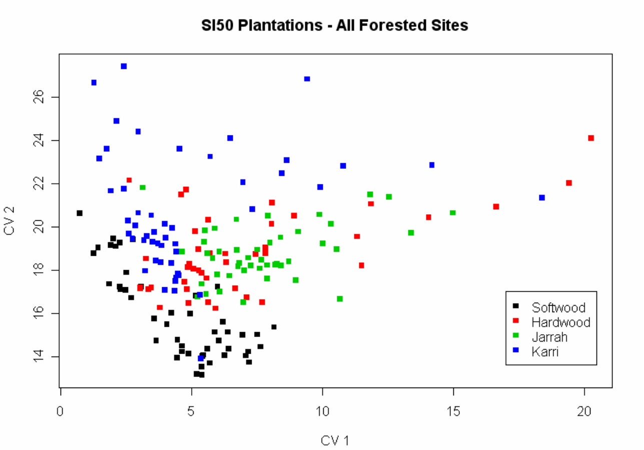

Figure 1: Canonical variate means plot for all the plantation training sites bare excluded). Softwood sites are shown in black, karri sites in blue, jarrah sites in green and other general hardwood sites in red.

2.3 Directed Contrasts

The aim of directed contrast analysis is to find a vector (linear combination of bands) that best separates two classes, while minimising the variation of sites within a class. Six contrasts analyses were conducted namely "karri vs softwood", "karri vs jarrah", karri vs hardwood", jarrah vs softwood", "jarrah vs hardwood", and "hardwood vs softwood".

The values of canonical roots yielded by all the contrasts range from 0.15 – 0.67 which suggest poor separation between the two classes in each case (typically a value greater than 1 indicates adequate separation for classification purposes). However, in each analysis, it is observed that there is relatively good clustering of like sites when the sites are ranked by canonical scores although there are some sites that are not separable. This means that classification is possible albeit with some errors. As an example, we show the rankings of CVA scores against the karri and softwood, and CVA scores against the karri and hardwood in Figure 2. The plots are in the form of boxplots. A clear general separation of the two plantation types is visible in each case, especially the contrast between "karri" and "softwood". As for the contrast between "karri" and "hardwood", the CVA scores are close to one another and there are more intermingled points, but a separation of the two classes is still evident. The canonical roots yielded for the karri and softwood, and karri and hardwood contrasts were 0.67 and 0.11 respectively.

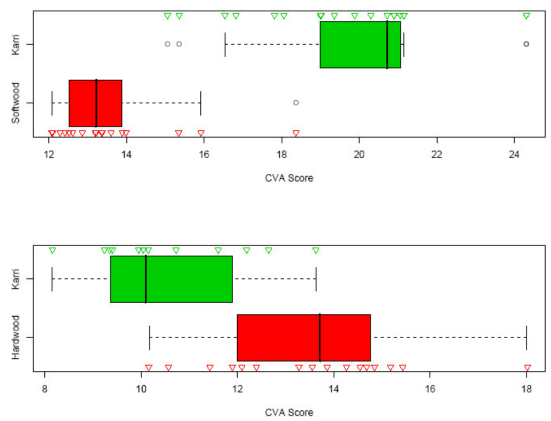

Figure 2: Top picture: Box plot of CVA scores against "karri" (green) and "softwood" (red) sites. Bottom picture: Box plot of CVA scores against "karri" (green) and "hardwood" (red) sites.

Using the results from the directed contrasts analysis, four more spectral classes, namely "hardwood", "jarrah", "softwood" and "karri" were defined for classification purposes.

2.4 Classification

From the results of the CVA analyses and the directed contrast analyses, the training sites were grouped into 9 classes for classification. These classes were then used in the maximum likelihood classifier (MLC) to produce a map. In (Caccetta and Chia, 2002), both the MLC and MLC – NBR (maximum likelihood classifier with neighbourhood context) were performed and results compared. The latter method was shown to have a higher accuracy rate.

Return to Program

2.5 Accuracy Assessment

To estimate the accuracy of the classification, we used 268 independent sites that include labels from jarrah, karri, softwood and hardwood. The counts of pixels for the true class labels (from the validation data) and the mapped class labels (from aggregation data) were used to generate a table shown below. The overall accuracy may be calculated as (100 x counts in agreement / total count) which gives 86.3%.

	True Class Labels				
Mapped	Soft	Karri	Hard	Jarrah	Total
SoftWoodMature	3876	178	213	292	4559
SoftWood	1216	893	379	86	2574
KarriGreen	15	5485	207	588	6295
Karri	242	19559	620	1882	22303
Hardwood	176	1397	6973	2223	10769
Jarrah	252	1458	1572	19833	23115
Bare (Karri)	140	614	343	841	1938
Bare (HardWood)	747	425	3913	161	5246
Bare (Softwood)	133	8	376	163	680
Total	6797	30017	14596	26069	77479

Table 1: Number of pixels tabulated by mapped and true classes for the single-date classification.

The classification was repeated using only the three Landsat MSS equivalent image bands with a significant loss of discrimination between the classes. For this reason the mapping is restricted to time periods for which Landsat TM data is available (1988 onwards in Australia).

3. TIME TRACES

This section presents a study conducted to investigate the temporal characteristics of different plantation types namely, "karri", "jarrah", "soft wood" and "hard wood", and how these characteristics may be of use for mapping and monitoring. Details of this study can be found in (Caccetta and Chia, 2002). Here, we give a brief summary of the study, highlighting the main points.

We concentrate on plantings and re-generation that start around 1990, and track their spectral response through time to 2002. This gives plots which clearing/planting at around 1990 and their spectral evolution for roughly a decade.

Using multi-channel image files derived from "brightness index" ((TM band 3 + TM band5)/2) and "greenness index" (TM band4 – TM band3 + 100) from Landsat data at approximately 2-yearly intervals from 1988 – 2002, and relevant training sites, a series of plots were obtained for each plantation type of interests. These plots include "brightness vs time", "greenness vs time" and "brightness vs greenness" where time represents the years from 1990 – 2002. The full set of plots for each plantation type can be found in (Caccetta and Chia, 2002).

The main observations made from the plots include

- For the four plantation types considered, most change in spectral response happens in the first four years of growth and or regeneration.
- Jarrah and karri age responses have distinctive shapes. The change in spectral response is more pronounced in karri than in jarrah
- Hardwood in rotation has a similar temporal response to jarrah and karri. New plantations have a step response, and some sites exhibit a longer bare response than for karri and jarrah.
- The four types have a similar trajectory through "brightness/greenness" space although they vary in absolute values.

In the following, we show the growth paths of karri and softwood plantations over the period 1990 – 2002 (note that each type was bare in 1990) in a "greenness vs brightness" plot as displayed in Figure 2. The trajectories of these two types are somewhat similar, but they vary in absolute values and in the rate of growth in the first few years. For example, mature softwoods tend to be darker, maturing karri tend to be greener and the slope from bare to peak greenness is steeper for the karri.

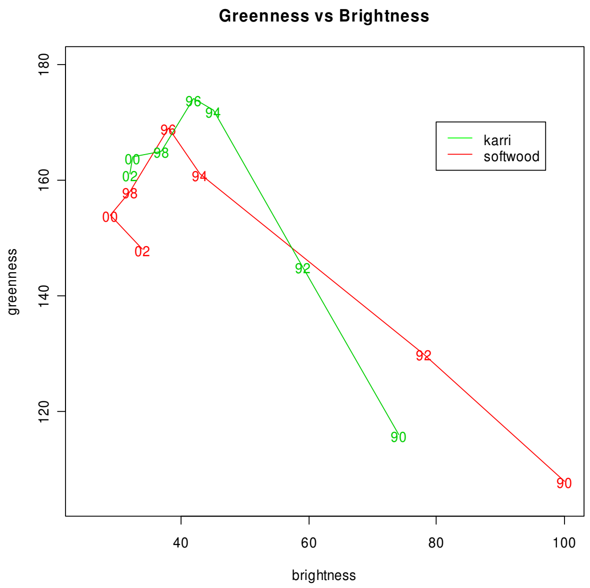

Figure 2: Plot of "greenness" index against "brightness" index for "karri" (red) and softwood (green). The data points in the plot are the overall mean values of both indices of the training sites over the period 1990 - 2002. Lines are drawn through data points to show the paths of their growth.

From this study we learn that the ability to distinguish different types based on their spectra alone varies as they grow. This suggests that multi-temporal classification models may be advantageous. A plantation would at some points in the time series be at the ages where it is separable from other classes. At other ages it may be spectrally not separable from some other vegetation types.

4. COMPOSITING

The investigations described in sections 2 and 3 show that the spectral response of a plantation depends on both the plantation type (softwood or hardwood) and the age of the trees. The separability of plantation types from each other and native regrowth and hence classification accuracy varies with the age of the plantation. The challenge is to use the temporal information to improve classification accuracy in a time and cost effective manner.

In theory, for each stratification zone, a single date classifier trained to recognise all the likely forest types (for example no new plantation types emerge through time) could be applied to each image in the time series to produce a temporal sequence of forest type labels. In practice this is not feasible due to limitations in calibration accuracy, and variability due to mismatches in acquisition dates and seasonal differences. The effect is most observable in classes that have low spectral separation (of the order of a few digital counts). The alternative is to optimise the classifier for each image date. This also allows new classes to be accommodated. At some point in the plantations lifecycle it enters the stage where it is most spectrally separable from other forest types and a temporal model can be used to increase the accuracy of the time series classification as a whole. This rationale is used successfully in the Land Monitor Project (http://www.landmonitor.wa.gov.au) for mapping and monitoring salinity. However, the approach is operator intensive and time consuming (a rough estimate, of the order of twenty person years for the plantation regions of the Australian continent assuming nine Landsat TM epochs in the time series).

Instead, we adopted a strategy known as *compositing* where a single scene composed of data drawn from all images in the time series is analysed. The aim of the compositing procedure is to combine the information on forest areas in the Landsat TM sequence into a single image for classification purposes. It does so, however, in a naïve way that does not use the full power of the temporal information. Compositing significantly reduces the computational effort (one classification instead of nine) as well as potentially increasing the separability of some classes by amalgamating data across growth stages. The following sections describe several options of the compositing procedure.

The initial compositing scheme used only 'candidate' plantations where pixels with a clearing or harvesting or revegetation event were considered. The epoch having the most uniform appearance (most common acquisition dates and best between scenes calibration) was chosen as the base image. For candidate plantation pixels, the composite signal is the average of the image data for each epoch in which the pixel is labelled as forest. For non-candidate pixels (non-forest and native forest), the composite signal is the base image signal. Pixels are identified as forest based on the forest extent and change maps from the AGO Land Cover Change Program. Potential plantations or 'candidates' are identified based on the sequence of forest cover labels. Any areas that have been cleared or regrown and have at least one 'forest' label in the Landsat TM sequence are identified as 'candidates'. Candidates include all areas of forestry activity where a harvest or regrowth event has been detected as well as areas of native vegetation that have been significantly disturbed, for example fire in a national park.

Two issues arose with this compositing scheme. First, native forest areas are subject to variation caused by seasonal effects and minor disturbances that, although they do not result in an area being considered 'cleared', they do affect the variability of the spectral signal. In many areas, native forest is more accurately classified if the signal from all epochs is averaged in the same manner as for candidate plantations. Second, with gaps between the epochs of up to three years in the AGO image sequence (i.e, 1989, 1991, 1992, 1995,1998, 2000, 2002, 2004, …), it is possible that a harvest event will not be detected in an area of forestry operations if the new plantation has grown sufficiently to be classified as forest before the next image date. Although forest operations are not strictly 'Kyoto' forests, that is, plantations on cleared land in 1990 for carbon accounting purposes, the plantation type is still of interest for more general inventory purposes. An example from the 'Green Triangle' region along the Victorian / South Australian border is shown in Figure 5. Purple areas have been identified as candidate plantations based on a change event (clearing or regrowth) in the forest cover mapping. Green areas are mapped as forest in all image dates. However, many of green areas show the size, shape and internal boundaries typical of a managed forest system. Land use maps for the region (http://www.environment.sa.gov.au/reporting/land/landuse/current.html) show the whole area as plantations. Compositing was then extended to include all pixels mapped as forest, not just candidate plantations.

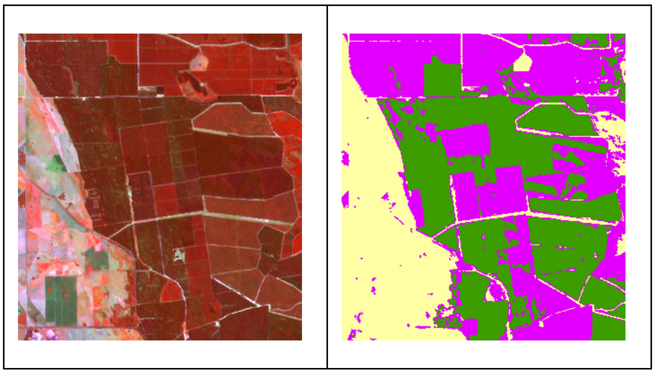

Figure5: Landsat TM image (left) and "candidate" plantations map (right). Area identified as candidate plantations are in purple. These are forest areas in which a change event (clearing or regrowth) has been identified in the forest cover mapping. Areas of forest with no change are shown in green. Yellow indicates non-forest areas (typically agricultural activities).

The limitation of these compositing schemes is that if an area of trees is cleared and then replanted with a different species, the composite signal is the average of two distinct forest cover types. Softwood to hardwood and native to plantation conversions are more than just a rare event.

In the final compositing scheme, if a revegetation event is detected, the average is over only those epochs following the most recent revegetation event, that is only over the new growth. As a special case, if the most recent epoch (2004, say) is non-forest but the preceding one (2002) is forest, then the signal is averaged from 2002 back to its corresponding revegetation event so that plantations harvested between the final epochs get a plantation type label.

5. PLANTATION MAPPING METHODLOGY – 2005 UPDATE

5.1 Product Specifications

The initial aim of the project was full forest inventory, which is to identify cover type (native forest, native regrowth, softwood, hardwood or environmental planting) for all areas of forest. As the project evolved the cover to be mapped became limited to

Return to Program

Kyoto plantations which are defined as areas of trees planted on non-forest land in 1990 (i.e. cleared for agriculture or other purposes). As native regrowth and environmental plantings are not spectrally separable, the final classes were:

0. No revegetation event after 1989 (includes native forest and plantations established before 1990 as well as non-forest areas)

1. Native regrowth on land that was not forest in 1990 (includes environmental plantings)

2. Softwood plantations on land that was not forest in 1990 (typically pines)

3. Hardwood plantations on land that was not forest in 1990 (typically eucalypts)

A single classification is performed on a composite image and aggregated into the above four classes. These classes are then inserted into the 0/1 revegetation layers from the LCCP forest cover mapping for each of the time intervals from 1989 onwards. A zero in the revegetation layer (no regrowth) is left unchanged and a one (regrowth in that time period) is replaced with one of the above class labels.

Only the AUSLIG 1:1,000,000 map sheets with significant areas of plantations were processed. The regions processed in the '2004 update' were:

Western Australia:	SH50, SI50 and SI51
South Australia:	SI53 and SI54
Victoria:	SJ54 and SJ55
Tasmania:	SK55
New South Wales:	SI55, SI56, SH56
Queensland:	SG56, SF55, SE55
Northern Territory:	SD52

5.2 Ground-truth Data

There were four sources of ground-truth data showing locations of hardwood and softwood plantations used to train the classifiers. Most of the data was supplied by MBAC – a consulting company with an extensive plantation / forest inventory database. For areas in Tasmania, more complete coverage is provided by the Tasmanian government plantation inventory mapping owned by DPIWE. In Western Australia plantation extent vectors (from 2002) from the Department of Environment and Conservation were the primary source of information. They identify hardwoods, softwoods, karri regeneration, jarrah regeneration and other species. Information for the Esperance region (SI51) was from a plantation resource mapping project (Ecowise Environmental, 2005). The data was digitised from 2002 imagery using national inventory information. There is a confidence measure for the assigned hardwood / softwood label (which is mostly 'very uncertain'). Most of this ground data covered areas of forestry operations, that is, not on the 'Kyoto plantations' of most interest to the AGO. As a result, the classification approach was to classify the whole area, then apply relevant masks to exclude non-forest, native forest and forestry areas.

5.3 Methodology

First, a composite image is formed using the method described in the preceding section. Cloud, haze and noise affected pixels are masked prior to compositing. The ground-truth data is overlaid on (if digital) or compared to the composite image. If there were spatial patterns in the distribution of the hardwood and softwood sites and / or spectral variation across the image, the region / map sheet was divided into stratification zones within which individual classifications were performed. Most of the map sheets required little, if any, stratification to separate one region of plantations from another. Most did, however, make use of a plantations extent boundary to limit the classification to just the regions where plantations are expected rather than the whole map sheet. Outside this boundary it is assumed that all regrowth is native rather than hardwood or softwood plantations. These plantation extent boundaries were set very conservatively.

Training sites were selected based on the ground-truth data to represent the cover classes of interest. Sites were selected from the full range of cover types in the composite image, not just from areas of new trees on cleared land in 1990. This is partly due to the initial forest inventory requirement, but in many regions few of the ground-truth sites are on 'new' land.

Next, canonical variate analysis was used to investigate the spectral separability of the training sites, the need for stratification and the potential grouping of training sites into spectral classes for the classification. The classes were aggregated into the final four classes listed in section 5.1.

The classification approach was to classify the whole composite image, then apply relevant masks to exclude non-forest, native forest and forestry areas. For classification, the neighbourhood-modified maximum likelihood classifier was used since plantations are usually planted in plots and it is reasonable to assume that, within a plot, the plantation type of a pixel will be the same as its neighbours. Various neighbourhood rules and numbers of iterations were evaluated in (Furby 2004). It is anticipated that in most areas this classifier (class means and covariances) can be used unchanged for subsequent updates (where another epoch is added to the existing time series), however there will be some exceptions.

5.4 Accuracy Assessment

Accuracy assessments were performed in the manner as described in section 2.5 in cases where independent data are available. In cases where such data are not available, output maps of specific important areas were sent for independent expert interpretations.

5.5 Attribution

Some regrowth classes cannot be separated spectrally, for example human induced conservation plantings from natural recovery of native systems, green flushing river systems from young plantations or older hardwood plantations from native species. Thus, a final attribution step where contextual information is used to excluded or relabel such areas is applied independently by the AGO.

6. RESULTS

The methodology described in section 5 to classify plantation types using time series images was applied to the same SI50 test area as described in section 1. First the time series images are composited into a single image, and using the same set of

Return to Program

training data, spectral analyses were performed. The MLC – NBR classification method was then used to produce a classified map of the different plantation types.

An accuracy assessment has been performed on the product derived from the composite image. The counts of pixels for the true class labels (from the validation data) and the mapped class labels (from the classification of the composite image) were used to generate the data shown in Table 2. The overall accuracy, calculated as (100 x counts in agreement / total count), is 71.6%. However, the focus of the classification using the composite image was correct allocation to hardwood and softwood categories for the AGO, not to the jarrah and karri sub-classes (both considered hardwood for the carbon modelling). Treating 'karri', 'jarrah' and 'hardwood' as a single category gives an overall accuracy of 88.1% for the classification from the composite image.

	True Class Labels				
Mapped	Soft	Karri	Hard	Jarrah	Total
SoftWoodMature	3579	102	7	58	3746
SoftWood	1987	4006	742	350	7085
KarriGreen	19	6133	305	72	6529
Karri	17	12932	133	274	13356
Hardwood	420	3752	8404	1717	14293
Jarrah	387	2906	2727	22192	28212
Bare (Karri)	8	82	389	979	1458
Bare (HardWood)	4	27	190	5	226
Bare (Softwood)	15	1	15	7	38
Bare	361	76	1684	2	2123
Total	6797	30017	14596	26069	77479

Table 2 – Number of pixels tabulated by mapped and true classes for the classification of the composite image from the WA test area described in section 2.

7. DISCUSSION AND CONCLUSIONS

An initial approach has been developed using existing time series Landsat imagery to derive plantation type maps. The spectral growth patterns of different plantation types were investigated to identify the strategy that best separate them from each other and from native regrowth. A compositing technique was then developed and used to integrate age information into a single classifier that is feasible to apply to derive a national product. The development of the compositing technique was due to the prohibitive amount of effort required for multiple classifications and temporal modelling and the need to produce a product for carbon modelling purposes within a short time frame.

The accuracy of the plantation maps could be improved by:

- more ground-truth data in areas of 'Kyoto plantations';
- more sophisticated compositing techniques that merge data only according to plantation age ranges; and/or
- pursuing the multiple single-date classifications and temporal modelling approach.

The quality and extent of ground-truth data on plantation type is variable. In some regions, very detailed 'type' information exists in private databases. In other areas such databases, public or private, are yet to be developed. Appropriate blending of ground survey and remotely sensed information may provide a hybrid product superior to that from either source alone.

A simple compositing technique that averages over all of the new growth was used is this application. Alternatives include averaging only over key ages or weighting the contribution to the average by plantation age. Such a strategy is complicated by the larger gaps in the time sequence during the 1990s meaning that the age at which new plantations are first detected is quite variable. However, this approach would not be expected to be as accurate as performing single-date classifications and developing appropriate temporal models, if the time frame and resources permit.

ACKNOWLEDGEMENTS

This work was supported by AGO Land Cover Change Program.

REFERENCES

Caccetta P.A. and Chia, J. (2002). *Remote Sensing Methods for Plantation Attribution – Experiments and results for map sheet SI50*. CMIS Technical Report 04/31.

Caccetta, P.A. and Chia, J. (2003). *Plantation Classification for sheet SI50sw*. CMIS Technical Report 04/30.

Campbell, N.A. and Atchley, W.R. (1981). T*he geometry of canonical variate analysis*. Syst. Zool., 30, 268-280.

Ecowise Environmental (2005), *Western Australian Plantation Resource Mapping (Project AG100019)*, Report to the Australian Greenhouse Office.

Furby, S.L (2004). *Plantation Mapping – South Australia / Victoria Border - SJ54 (Green Triangle)***.** CMIS Technical Report 04/85.