Australian Government
Bureau of Rural Sciences

Land Use Mapping for the Murray-Darling Basin: 1993, 1996, 1998, 2000 maps

Baseline land use profiles to support the implementation of the basin Salinity Management Strategy

Postal address:
Bureau of Rural Sciences
GPO Box 858
Canberra, ACT 2601

Internet: http://www.affa.gov.au/brs

Foreword

Land use mapping has many applications relevant to national scale issues, such as supporting information on diversity of environment and economic use of lands, strategic industry-based planning, providing objective assessments on land use and land use change. Land use data also provides background information products to inform decision making and assist in the presentation of policy proposals.

The 1993 to 2000 time series Murray Darling Basin (MDB) land use maps provide an overview of dominant agricultural activities occurring within over 250 Statistical Local Areas (SLAs) of the Murray Darling Basin region. Outputs for the years 1993, 1996, 1998 and 2000 were achieved by spatially assigning statistical agricultural census data using the established Spatial Reallocation of Aggregated Data (SPREAD) method. The repeatable SPREAD approach allows for the cost effective and timely delivery of broad-scale agricultural activity across extensive areas of agricultural land. The outputs present a time-series profile of agricultural activity of 21 major agricultural commodity classes.

Dr Peter O'Brien
Executive Director
Bureau of Rural Sciences

Executive summary

Land use mapping produced in this project is to provide basin-wide representations of major commodity types for mapping and display, and spatial input to basin-wide hydrologic and salinity modelling.

Accountability arrangements for the Murray-Darling Basin Commission's Basin Salinity Management Strategy 2001-2015 initiative (BSMS) rely on the definition and adoption of agreed baseline conditions across the Murray-Darling Basin. This project addresses the requirements for baseline land use conditions a 1 January 2000 (Year 2000 map); and underpins understanding of the variability in baseline conditions over the 25 year benchmark period 1975 – 2000 (Year, 1993, 1996, 1998 and 2000 maps plus concorded SLA statistics 1983-2000)

The aims of the current project were to implement the already established SPREAD procedure for mapping the MDB for the required period and also to include the newly developed SPREAD II probability mapping. R&D components to improve accuracies are partially established and these include modifications of primary input data, adjustment of methods and the use of constraining surfaces

The SPREAD method spatially allocates agricultural census data (AgStats 96/97, Australian Bureau of Statistics), using satellite imagery, to determine agricultural land use. The satellite data inputs were monthly "cloud-free" composites of NDVI (normalised difference vegetation index) derived from daily overpasses of the Advanced Very High-Resolution Radiometer (AVHRR). Known land–uses were assigned time-series NDVI profiles by the use of (1) a database of control sites that was earlier acquired from farm surveys and (2) Assignment of land-use from satellite data was based on three metrics derived from the monthly NDVI values for each image pixel. These were range, mean and time of maximum NDVI.

Probability surfaces for individual land use were then produced by comparison of metric results for the image data with the same metrics for the target land–use profiles. Subsequent to this, the number of pixels assigned a particular land use was constrained to the area reported in the agricultural census for the SLA being solved. The resulting agriculture commodity distributions were then combined with non-agricultural land use derived from other sources to create an encompassing land use map of the MDB. Output maps and probability surfaces have been derived for the years mentioned above.

Spatial accuracy of land use classification within SLA boundaries depends on the SPREAD analysis of AVHRR data. In this phase of the product, verification procedures have been developed that rely on comparison of the SPREAD outputs with Catchment Scale Land Use Mapping (CLUM). Although there are issues with the degree of detail, spatial resolution, timing and methods of the local mapping, the verification results indicate that reasonable accuracies are being achieved at primary levels of classification (i.e. cropping, pastures) and also for some commodities at crop-type level (e.g. cotton). Although preliminary, testing indicates that the use of CLUM derived target profiles and constraint surfaces, such as a map of irrigated areas, can improve the classification.

Output land-use maps generated by the project are statistically and spatially accurate to the Statistical Local Area (SLA) level due to the spatial constraint to census agricultural data. The data shows there has been a progressive increase in dryland cropping (38%), irrigated cropping (59%) and irrigated horticulture (73%) between 1993 and 2001.

Contents

Figures

Tables

1. Introduction

1.1 Objectives of the Project

The objective of this project is to produce land use data that meets the needs of the Murray-Darling Basin commission's (MDBC) baseline condition assessment for the Murray Darling Basin. The Bureau of Rural Sciences (BRS) has produced a time series of Broadscale Basin wide maps detailing Land use from 1994 to 2001. Mapping has involved the analysis of AVHRR Satellite imagery combined with ABS Agricultural Statistics data. Descriptive tables depicting the annual areas of commodities based on concorded SLA-based Agricultural statistics for the period 1983 through to 2001 have also been produced

The requirement for regional land use mapping (both historical and current)

Land use mapping has many applications relevant to national scale issues that include:

- Salinity and Hydrological modelling.
- Supporting information on diversity of environment and related patterns in economic use of land (e.g. risk assessment of land use with climate reliability projections; relating land use to social factors).
- Strategic industry-based planning.
- Integrating with other data sets to allow multi-objective assessment on land use and land use change.
- Background information products to inform decision-making and assist in the presentation of policy proposals.

Method for regional land use mapping

National agricultural statistics and time-series satellite NOAA-AVHRR data is being used as a primary input for a method of generating regional land-use maps across Australia. The method is designed to be a cost effective means of integrating available data sources on land use. The method can be briefly summarized in the following steps:

1. Acquisition of time series cloud-free AVHRR composites for continental Australia.
2. Compilation known field sites (approx. 1000) for commodity NDVI signatures.
3. Masking non-agricultural land.
4. Allocation of likely land-use classes based on known total areas (census data) for local areas using a statistically based method (SPREAD II)

Since there is a requirement for broad-scale data on land-use in both a current and historical context, the method is designed to provide data where no other data exists. The national land use mapping (NLUM) product supplements the often incomplete catchment-scale land-use mapping (CLUM) that is undertaken by State government surveys. This issue is expanded in the next section.

2. Land use mapping methodology

2.1 Overview

The mapping outputs comprised a set of digital maps of the Murray-Darling Basin showing land use for the years 1993, 1996, 1998 and 2000. The methodology used to construct the maps is a derivative of that used to construct the 1996/97 Land Use of Australia, Version 2 data set for the Audit (Stewart et al., 2001). The main difference between the current methodology and that used to construct the Audit map is that the algorithm used to allocate agricultural land uses to agricultural land pixels has been modified. In the original methodology, the agricultural land uses were determined using the SPREAD algorithm (Walker and Mallawaarachchi, 1998). The modified version of SPREAD used in this project will be referred to as SPREAD II. The remaining changes to the original methodology all involved varying the input data including the introduction of additional data to provide an irrigation constraint see Appendix 2.

SPREAD determines the best-case land use allocation for each unknown agricultural land pixel and provides what is, at best, a qualitative guide to reliability. SPREAD II, by contrast, determines, for each agricultural land pixel, the probability that it is each of the mapped commodity groups. Ie, a probability surface is produced for each of the mapped commodity groups, giving for each agricultural land pixel the probability that its land use is the commodity group. Derived probability surfaces can be produced for aggregations of the mapped commodity groups by adding the probability surfaces for the individual commodity groups. A summary map embodying an approximation to a maximum likelihood set of agricultural land use allocations that is consistent with the AgStats-based area constraints is derived from the probability surfaces.

The probability surfaces and the summary grids were constructed in ARC/INFO grid format with geographical coordinates referred to WGS84 and 0.01 degree cell size. The summary grids have a value attribute table with attributes defining the agricultural commodity group, irrigation status and land use according to the Australian Land Use and Management Classification (ALUMC), Version 4. The ALUMC Version 4 is the land use classification used in the 1996/97 Land Use of Australia, Version 2 data set The current version of ALUMC, Version 5 (Bureau of Rural Sciences, 2002), can readily be derived from Version 4. Visit http://www.affa.gov.au and search the site for 'ALUMC' for information about both Version 4 and Version 5 of the ALUMC.

2.2 Input data

NDVI imagery

A set of thirteen 28 day composite NDVI images covering the period 1st April to the following 31st March was used to disaggregate the AgStats data for a given year. In the original methodology used for the Audit map (Stewart et al., 2001), a set of twenty-six 14 day composites was used, covering the same period. Cloud correction of the NDVI images was undertaken by the Department of Environment and Heritage using a similar methodology to that used for the Audit map. It was expected that using longer composite periods would reduce the need for cloud correction and thereby reduce the chance of error due to spurious cloud correction, without significantly reducing the discriminating power of the NDVI imagery.

Each pixel in the map for a given year is characterized by the time sequence of 13 NDVI values from the NDVI images covering the period from 1st April in the year mapped to 31st March in the following year. Similarly, each agricultural control site is characterized by the time sequence of 13 NDVI values from the NDVI images for the control site location and covering the one-year period from 1st April to the following 31st March during which the commodity group represented by the control site is known to have been produced.

AgStats data

AgStats data are supplied by the Australian Bureau of Statistics (ABS) and were based on full censuses reported on Statistical Local Areas (SLAs) up to and including the 1997 AgStats database. Thereafter, except when the release coincides with the five yearly population census, they were based on surveys with much of the data reported only on Statistical Divisions. Current ABS policy is that AgStats data released in the five-yearly population census years are to be based on full censuses and reported on SLAs; in the intervening years the data will be based on surveys. The 2001 AgStats database is the first census-based release since the 1997 AgStats database. Special releases of agricultural survey data reported by SLA can be obtained for specific purposes but the standard errors are large for certain items.

AgStats censuses and surveys up to and including the 1999 survey report commodities produced in the period 1st April in the previous year to 31st March in the year of the census or survey – together with a few commodities harvested after 31st March but largely produced in the one-year period to 31st March. The 2000 AgStats survey and subsequent censuses and surveys report commodities produced in the period 1st July in the previous year to 30th June in the year of the census or survey. The change in the reporting period does not have any significant effect on the area data used for land use mapping.

The 1993 map was based on the 1994 AgStats database, which is a reasonably complete census. The original intention was to produce a map for the year 1994 based on the 1995 AgStats database. The 1995 AgStats database, however, lacks many important commodity items. Missing data would need to be estimated by interpolation and extrapolation from other years. Consequently, a 1993 map was made instead. The 1993 map was made using 1994 AgStats data concorded to ASGC 1996 SLA boundaries. The 1994 AgStats database has one significant omission: for NSW SLAs, though the total area of cotton is reported, the proportion that is irrigated is not. This omission also affects earlier AgStats collections available in digital form. However, the areas of irrigated and dryland cotton are available in the 1995 AgStats database. The irrigated proportions for cotton in NSW SLAs that were missing from the 1994 AgStats database were assumed to be the same as those reported in 1995 AgStats data concorded to ASGC 1996 SLA boundaries. Where a proportion was required for an SLA for which no cotton was reported in the concorded 1995 AgStats data, a default value of 0.84 was assumed – this was the average proportion of cotton irrigated for all of NSW according to the 1995 AgStats database. The 1994 AgStats database has a probable error in the area of berry fruit reported for the South Australian SLA entitled Lacepede (DC): the value was reported as 1916.0 ha though the corresponding values from the 93 and 97 AgStats databases were both 0.0 ha. Before running SPREAD II the area constraint for berry fruit for this SLA was reduced to less than 100 ha to ensure that no allocation would occur.

The 1996 map was based on the 1997 AgStats database, a very complete census. No estimation of missing commodity data was undertaken: the 1997 AgStats database is considered to be sufficiently complete in its own right for land use mapping purposes.

The 1998 map was based on the 1999 AgStats database, which is a reasonably complete survey. The 1999 AgStats database does not report the areas of irrigated pasture, irrigated cereals and irrigated 'other crops'. The missing data values were estimated by interpolation between values from 1997 AgStats data (concorded to ASGC 1998 SLA boundaries) and

2001 AgStats data (concorded to ASGC 1998 SLA boundaries). The 1999 AgStats database incorrectly reports the total area of holdings value for the South Australian SLA entitled Unincorp. Riverland as 18 264 ha. This value was changed to 450 203.7 ha, which is the average of the values for this SLA from the 1997 and 2001 agricultural censuses.

The 2000 map was based on the 2001 AgStats database, a very complete census. No estimation of missing commodity data was undertaken. Like the 1997 AgStats database, the 2001 AgStats database was considered to be sufficiently complete in its own right for land use mapping purposes.

For each map the AgStats data were processed in the same way as for the Audit map:

- orchard tree numbers were converted to areas;
- vegetable areas were adjusted for multiple cropping;
- various cropping areas were adjusted for double cropping using data from the Australian Bureau of Agricultural and Resource Economics (ABARE) 1996 – 97 Farm Survey (Australian Bureau of Agricultural and Resource Economics, 1997);
- AgStats irrigation data were used to disaggregate other AgStats commodity data into irrigated and dryland components; and
- the irrigated and dryland areas for each SLA were scaled to fit the mapped areas potentially available for agriculture.

Control sites

Two sources of control sites were available. Firstly, control sites can be constructed from the control site information collected for the Audit map. The control site information collected for the Audit map comprises locations with land use information for the year 1996 – 97 and for subsequent years. All locations have more than adequate land use information for the year 1996 – 97. Most of the locations have more than adequate information for one or both of the years 1997 – 98 and 1998 – 99 as well. A given control site location with adequate land use information for a given year was treated as a single control site. This means that there are many instances of different control sites sharing the same location and, in some cases, sharing the same land use. The control sites used are, in the main, based on control site information collected for the Audit map. The control site information collected for the Audit map comprises locations with land use information for the year 1996 – 97 and for subsequent years. All locations have more than adequate land use information for the year 1996 – 97. Most of the locations have more than adequate information for one or both of the years 1997 – 98 and 1998 – 99 as well. A given control site location with adequate land use information for a given year was treated as a single control site. This means that there are many instances of different control sites sharing the same location and, in some cases, sharing the same land use. The control sites collected for the Audit map were supplemented with a relatively small number of control sites for grapes that were constructed using catchment scale land use mapping (CLUM) data currently being compiled by BRS (Bureau of Rural Sciences, 2002). The total number of control sites is approximately 3000; these are spread over the whole of Australia.

Topographic data

The topographic data set used is an update of the Division of National Mapping's TOPO-250K (Series 1) acquired in February 1999. This is the same as the topographic data set used in the construction of the Audit map. At the scale of mapping, this data set only provides usable information about the spatial distribution of water bodies, watercourses, built-up areas and airports.

Tenure data

The tenure data set used is the same as that used in the construction of the Audit map. It is a modified version of the tenure data set compiled by the National Forest Inventory (NFI), an agency within the Bureau of Rural Sciences, and completed in 1997. The modifications entailed classifying aboriginal freehold and aboriginal leasehold land according to whether or not it is used for agriculture and removing sliver polygons.

Protected areas data

The protected areas data set used, the Collaborative Australian Protected Areas Database 2000 (CAPAD 2000), was published by the Department of Environment and Heritage in 2000. This is probably the most suitable protected areas data set for making land use maps for the year 2000 and earlier years. The more recent version of this data set, CAPAD 2002, would be more suitable for making land use maps for the year 2001 and subsequent years. It seems preferable to use CAPAD 2000 rather than the earlier versions of the data set, CAPAD 1999 and CAPAD 1997, for mapping years prior to 2000. This is because CAPAD 1999 has erroneous data for Tasmania and is otherwise the same as CAPAD 2000 and CAPAD 1997 reflects early interpretations of IUCN guidelines that have fallen out of favour. The Audit map was based on a hybrid data set that combined data for mainland Australia from CAPAD 1999 and data for Tasmania from CAPAD 1997.

Forest data

The forest data set used, entitled Forests of Australia 2003, was compiled by the NFI and completed in 2003. This is an updated version of the NFI forest data set completed in 1997, which was used to make the Audit map. In the Forests of Australia 2003 data set, forest is defined in the same way as it is in the 1997 NFI forest data set, ie as woody vegetation with height greater than 2 m and crown cover greater than 20%. Further, the same crown classes are used, ie 20 – 50% (woodland), 50 – 80% (open forest) and 80 – 100% (closed forest). The crown cover classes were used in conjunction with the scaling of the AgStats data in the same way as for the Audit map.

Irrigation constraint

An irrigation area boundaries data set supplied by the Murray-Darling Basin Commission showing designated irrigation areas in the southern Murray-Darling Basin and actual irrigation areas in the northern Murray-Darling Basin can be used in an attempt to refine the probability surfaces for the agricultural land uses by introducing an irrigation constraint. If the irrigation constraint is used, the prior probability that the irrigated agricultural land pixels will be located inside the irrigation areas must be set to a suitable value.

2.3 SPREAD II

Image-based mapping for agricultural commodities agrees with ABS statistics at the Statistical Local Area (SLA) level, with finer levels of accuracy dependant on the success of the statistically driven SPREAD II classification procedure using AVHRR imagery at 1 km resolution. Both SPREAD and SPREAD II rely on the fact that different crops and pastures having different growth characteristics through a one year period so that different agricultural land uses can be characterized by the changing NDVI values over that period. Both algorithms allocate land use in accordance with these profiles and subject to the area constraints provided by the agricultural statistics.

The SPREAD II algorithm uses a Bayesian technique - a Markov chain Monte Carlo algorithm - to find a solution that is consistent with the ABS statistics and the available imagery. A detailed description of the mathematics is provided in Appendix 1. The format of the algorithm means that it is simple to include additional constraints if they exist. It discriminates based on three metrics derived from the AVHRR time-series profiles; mean NDVI, annual range of NDVI and time of maximum NDVI.

The conditions for the runs of SPREAD II used to produce the final maps were:

- All control sites derivable from the control site information collected for the Audit map were used. No control sites derived from CLUM data were used.
- For each SLA, the control site pool used to characterize each land use comprised all available control sites for that land use; the pools were not restricted to control sites situated geographically close to the SLA.
- Kernel smoother bandwidths used to construct the probability density functions characterizing the control sites for each land use were set to 100 for the NDVI mean and range, and 1 for the time of maximum NDVI.
- The irrigation constraint was used with the prior probability that the irrigated agricultural land pixels will be located inside the irrigation areas set to 100%.

3. Results

3.1 Outputs (SPREAD II

The land use mapping outputs will be referred to collectively as the 1993, 1996, 1998 and 2000 Land Use of the Murray-Darling Basin, Version 2. This set of maps is the second and final version of a first attempt at construction of a time series of land use maps of the Murray-Darling Basin for the years 1993, 1996, 1998 and 2000. The maps are supplied as a set of ARC/INFO grids with geographical coordinates referred to WGS84 and 0.01 degree cell size. For each year there is a set of probability maps, one for each land use(see figure 3.1), and a single summary map made from the probability maps using simple rules (discussed in the methodology section) to make an approximation to a maximum likelihood land use map (see figure 3.2). As supplied, the probability maps are floating point grids with cell value between 0 and 1 and no value attribute table while the summary map is an integer grid with a value attribute table with attributes defining the agricultural commodity group, irrigation status and land use according to the Australian Land Use and Management Classification (ALUMC), Version 4 (the land use classification used in the Audit map). See the page 0 metadata (Appendix 4) and the user guide (Appendix 5) for more information about the mapping outputs.

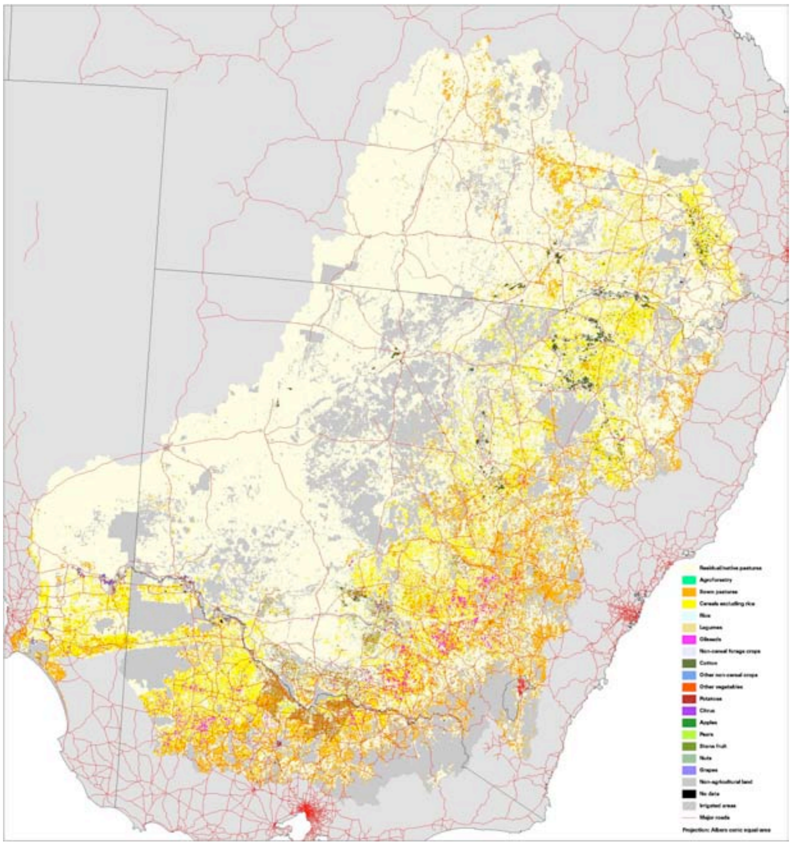


Figure 3.1. 2000 map showing the agricultural commodity groups mapped by SPREAD II. Major roads (courtesy Geoscience Australia) overlain for reference.

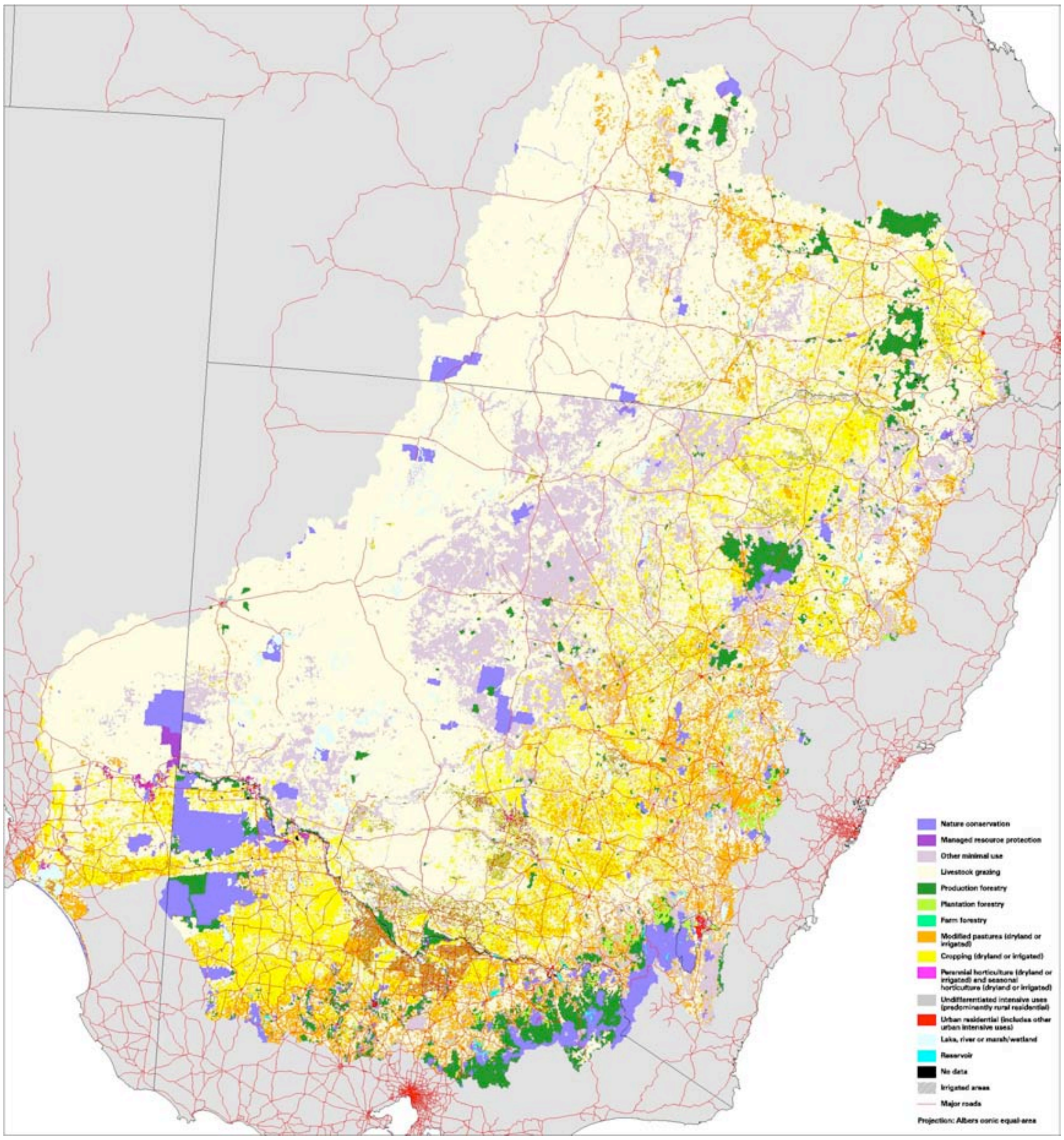


Figure 3.2. 2000 map summarised using land use categories based on the Australian Land Use and Management Classification, Version 4. Major roads (courtesy Geoscience Australia) overlain for reference.

Comparisons of land use maps, which reflect the agricultural statistics, are broadly informative about the changes in land use that have occurred in the MDB. In the 1993/94 period, dryland cropping occupied 7.6% of the MDB area and irrigated cropping occupied 0.5%. Changes in land use at the level 2 of classification are shown in Figure 3.3 compared with the 1993/94 baseline. Between 1993 and 2001 there is a general increase in cropping, horticulture and irrigation. This is generally associated with a reduction in modified and natural grazing areas, the latter occupying 60% of the MDB in 1993/94.

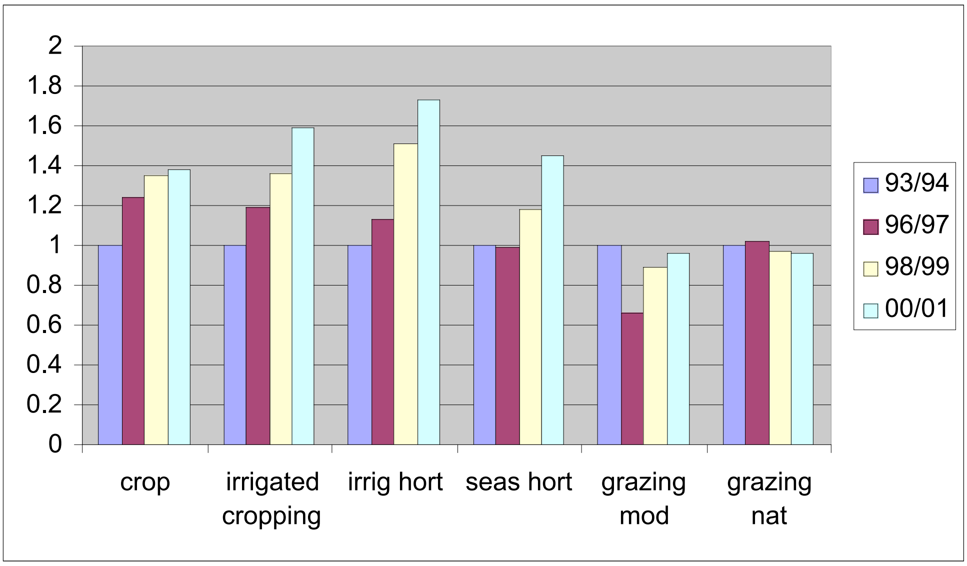


Figure 3.3 comparison of the areas of agricultural land use for years of outputs relative to 1993/94.

There is a general increase in crop (dryland), irrigated cropping, irrigated horticulture and seasonal horticulture balanced against the grazing areas of modified and native pasture.

4. Field control-site AVHRR NDVI signatures for Land-Use categories

4.1 Control site data -background

Control sites for the dominant agricultural land use (commodity based) were acquired in collaboration with state agencies (Stewart et al, 2001). The strategy was to sample the range of major agricultural enterprises occurring within similar biophysical regions. Details of the collection process are given in the NLUM report for 1996/97 (Stewart et al, 2001). As well as land use and NDVI, the database of control sites provides general information that includes sowing dates, harvest times and irrigation status. The sites were selected for largeness relative to the AVHRR pixel size (1.1 km) (see description below) and 4 successive years of information are present.

The aim of this document was to examine whether it is possible to derive characteristic NDVI signatures for these land uses and to determine what level the AVHRR- NDVI alone can discriminate land use.

4.2 Control site AVHRR-NDVI signature profiles

The problem of noise from mixed land-use

Due to the difficulty in finding truly homogeneous areas for control sites, there is noise present in the associated AVHRR NDVI time-series profiles. Also some control sites are small compared with AVHRR pixels (1.1 km). For the latter case slight spatial mis-registration errors can cause impurities or "noise" in control point data. The following is taken from the earlier NLUM report (Stewart, et al, 2001):

"The specification for control sites was that a commodity should dominate (> 50%) an area in the centre of a 4 x 4 km area (.ie. >200 ha within 400 ha). This should be possible in most cases for Pastures and Cereals excluding rice. For other commodities where an attempt has been made to locate a control site within a 4km^2 area has failed, a smaller representative area was allowed. For Rice, Legumes, Sugar Cane, and Cotton an area **2 km^2** dominated by the commodity was allowed (i.e. > 100 ha within 200 ha area). For Agroforestry, Oilseeds, Non-cereal forage crops, Other non-cereal crops, Other Vegetables, Potatoes, Citrus, Apples, Pears, Stonefruit, Nuts, Plantation fruit, and Grapes an area **1 km^2** dominated by the commodity is allowed (i.e. > 50 ha within 100 ha area)."

An example of the noise present in control site profiles is shown in Figure 4.1. AVHRR profiles for wheat harvested in October or November in a single 12 month period (1996/1997) are shown in Figure 4.1(a). The variation could be explained by the natural NDVI variability of a wheat crop but when compared with the average wheat signatures (Figure 4.1b), which are consistent between different years, it is apparent that the differences in Figure 4.1(a) are due to the presence of other non-wheat materials. Given that (from the above paragraph) wheat cereal control-sites are likely to be one of the most pure, there will be significant problems due to noise for many other commodities.

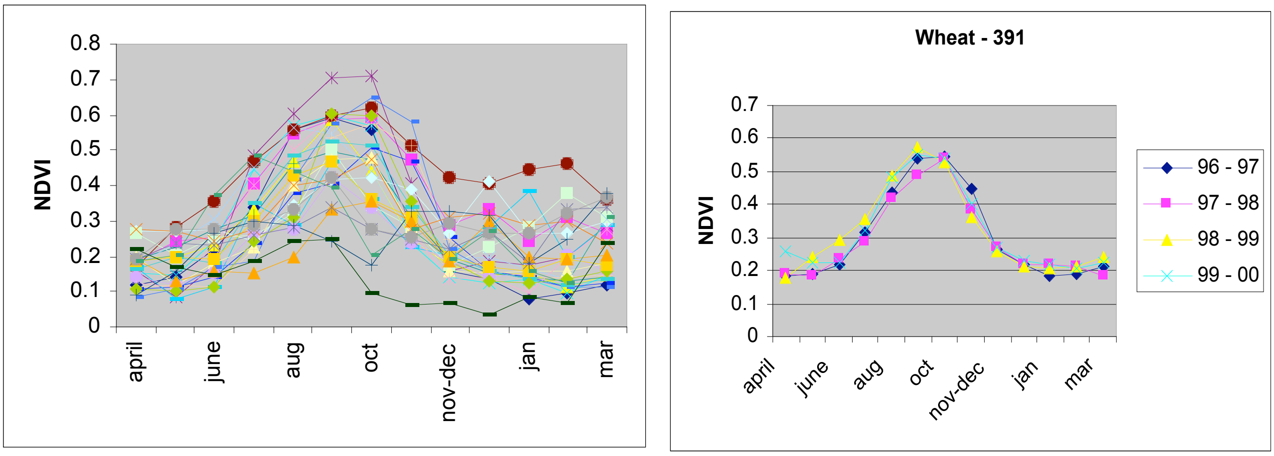


Figure 4.1: (a) AVHRR NDVI profiles for 27 wheat control-sites harvested in October or November for a single 12 month period (1996/1997) and (b) average wheat NDVI signatures for all harvest times in different years.

Variations in Commodity NDVI signature for a single land use will also arise due to climatic variability in factors such as rainfall and temperature. Farmer practices such as irrigation and the time of sowing and harvest will also have an influence. In all of this, though, there may be a typical signature with a characteristic range of deviation for particular crops and land-uses. Separation of this signature from the noise due to other materials is likely to be important for any mapping method using the satellite data. Ideally a mathematical technique such as Principal Components Analysis (PCA) might be employed to define the most common characteristics thereby isolating and removing the noise. This idea has not yet been explored fully and the use of data averages over a large number of control sites should at least smooth the noise and provide an approximation of the true NDVI signature. As in Figure 4.1(b), similarities in these profiles between years should identify whether there is a consistent signature.

4.3 Yearly AVHRR-NDVI averages

For commodities where there are a significant number of samples, the average NDVI signatures would be expected to show the features of the commodity itself rather than the noise. Figure 4.2 shows the average for each of the four years for a selection of commodity-types from a variety of audit classifications. The similarity between years and the smoothness of the profiles is evidence that we are seeing a close approximation of the true signature. However, given that many of the control sites have the same commodity from year to year, some of the “noise” or background materials could also be preserved.

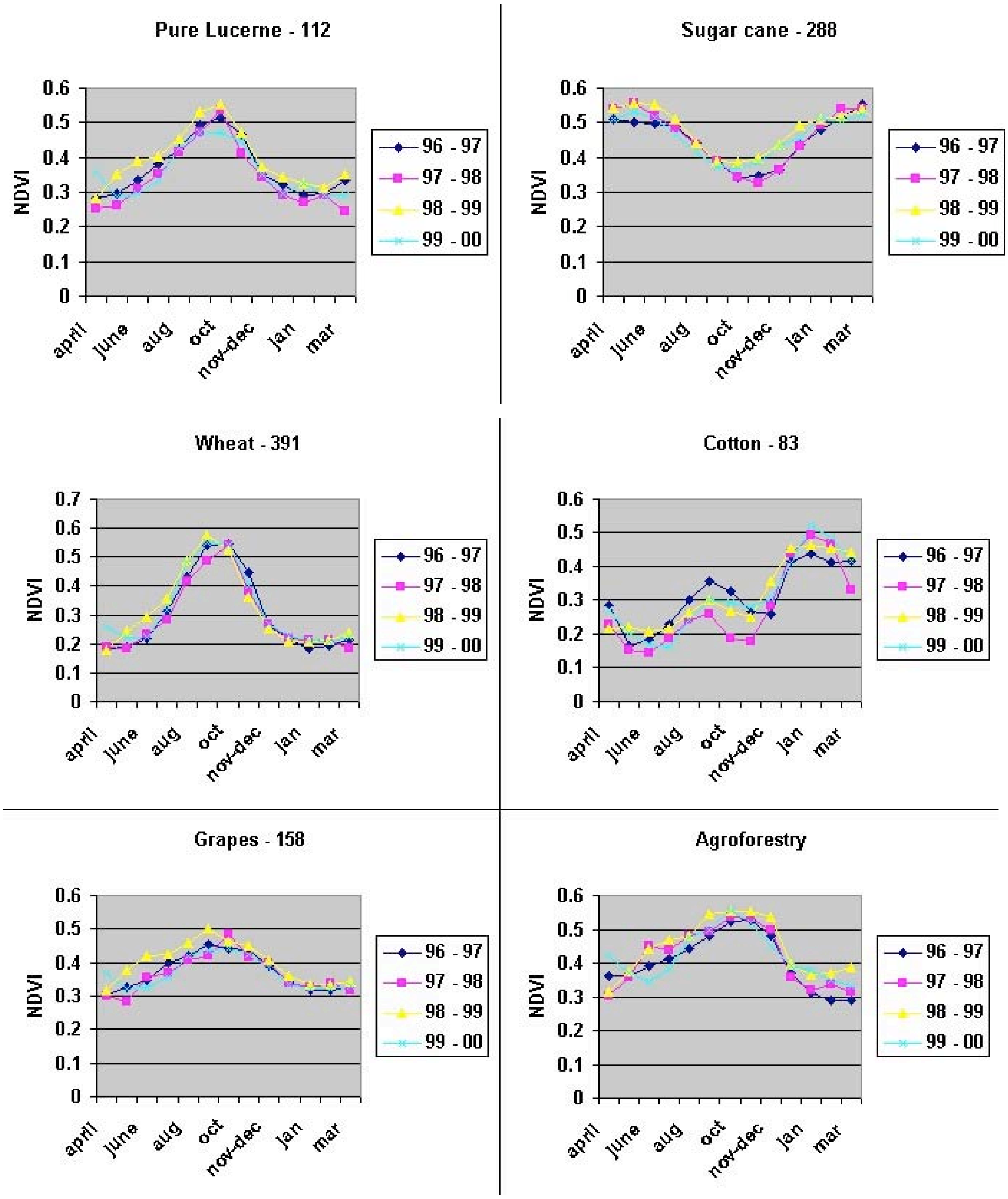


Figure 4.2. Yearly averages showing consistent NDVI time-series profiles for a number of different commodities.

Variation of average profiles within Audit commodity classes

Figure 4.3 shows the variation of NDVI signatures for ABS level 3 categories within 4 audit commodity classes – cereals excluding rice, legumes, oil seeds and stone fruit. Clearly there is variation between ABS level 3 classes and any attempt to classify AVHRR should be at this level prior to amalgamation to audit commodity class level.

Similarities between commodity profiles and potential mis-classification due to mixing

Total averages are shown in Figure 4.4 for commodities where there was considered to be enough data to provide a reasonable approximation for a commodity signature at ABS level 3 classification. While there are distinct differences between commodity signatures, there are also similarities between commodities of different audit classes. For example there is very little difference in the signature between wheat, lupins and canola – three ABS level 3 classes that correspond to the 3 audit classes cereals, legumes and oil seeds. As well, there are similarities between mangoes, avocadoes, sugar cane and bananas corresponding to 3 different audit classes. If for example, sugar cane was mixed with something with a flat signature, say native pasture, it would look like mangoes. Double cropping features are observed for sunflower, rice, cotton and maize raising the question of how to discriminate between them.

Hopefully many of these problems can be resolved by the use of ABS statistics, particularly where the commodity types with similar signatures don't occur together in a SLA. It is likely that there will be a considerable number of SLAs where these problems will not be resolved. The method can probably see further improvements in the use of surface temperature and other factors such as, slope, elevation and climatic suitability

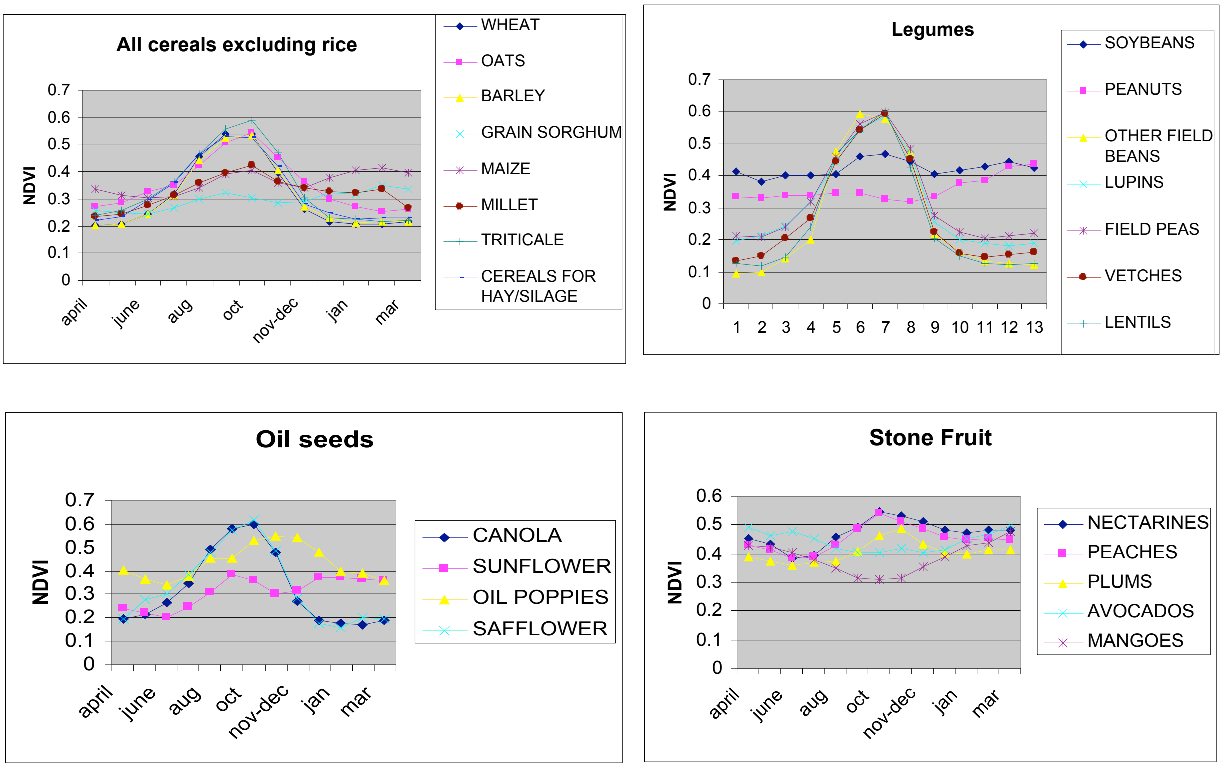


Figure 4.3. NDVI signatures for ABS level 3 categories within 4 audit commodity classes

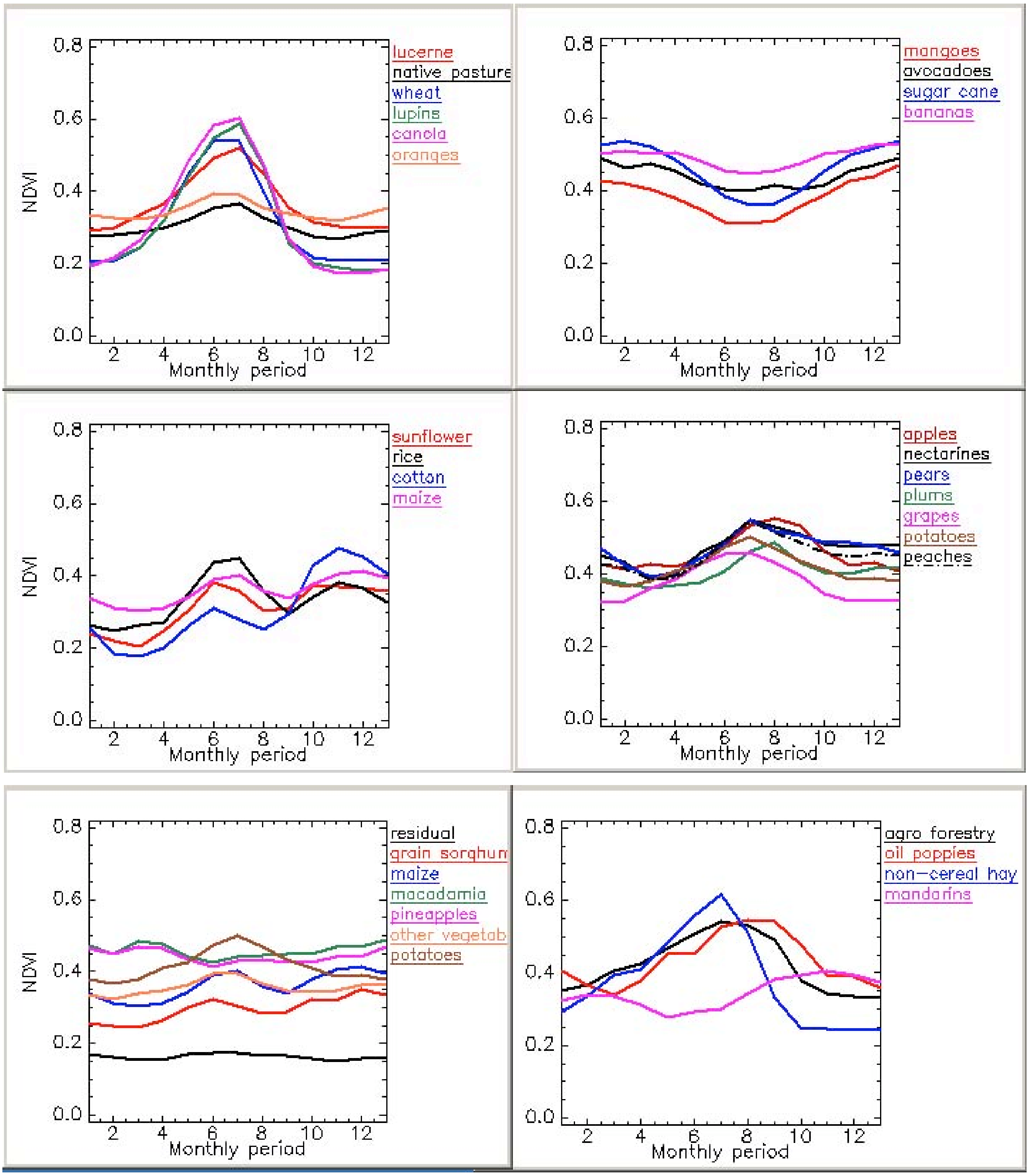


Figure 4.4. Total NDVI averages for commodities at ABS level 3 classification where there was considered to be enough data to provide a characteristic.

4.4 Climatic differences in NDVI signatures

Wheat profiles separated by harvest time show significant differences in their average NDVI profile (Figure 4.5). The NDVI profiles for native pasture are also highly variable on a regional basis (see Figure 4.6a), and these can be seen to be broadly influenced by rainfall

(Figure 4.6b). The native pasture signatures are so variable that it is likely that in some areas (particularly Victoria and South Australia) they will be confused with cropping.

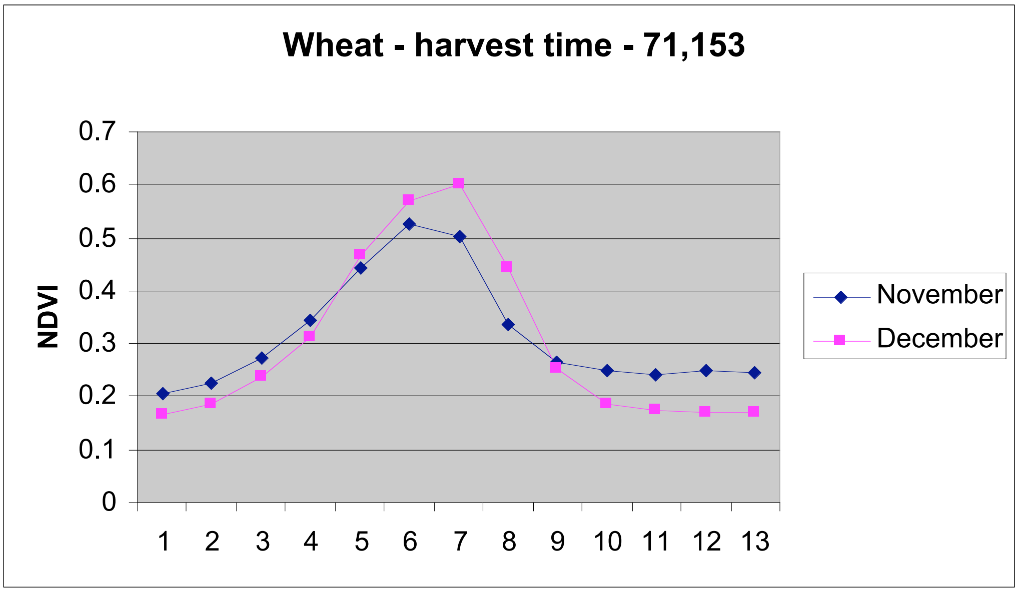


Figure 4.5. Comparison of average NDVI profile for wheat harvested at different times

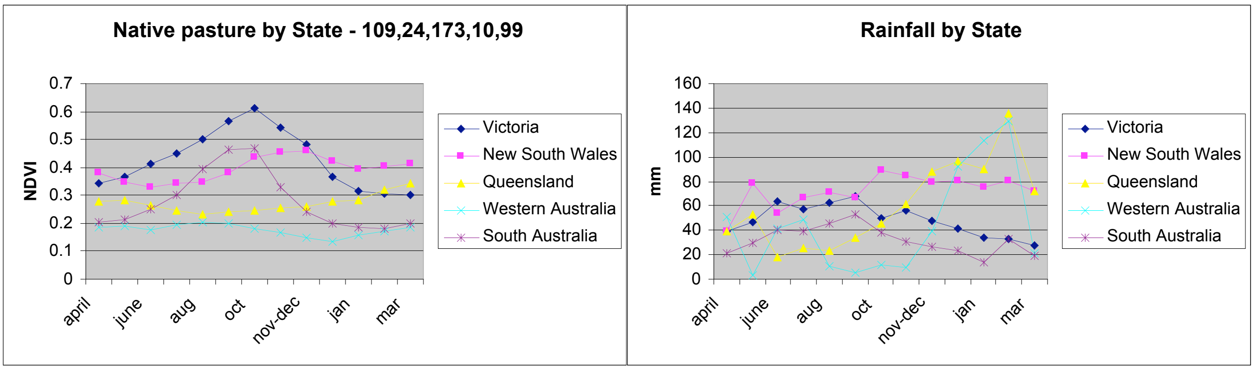


Figure 4.6. (a) Average NDVI profiles for native pasture by state and (b) the corresponding average monthly rainfall.

5. Testing and verification of results

5.1 Objective

A verification or accuracy-assessment needs to be conducted to test how well the SPREAD II method is allocating commodities and the spatial arrangement of the commodity allocations need to be verified against independent data (spatial accuracy). As discussed in the previous section, there will be limitations based on the between-class and within-class separability of the NDVI signatures for the various commodities. A number of variations of the method, with respect to NDVI signature inputs, are possible. With a procedure in place it is possible to test the effectiveness of the various approaches see Appendix 3 for more detailed information.

5.2 TAGALUM discussion on SPREAD initial outputs.

The Technical Advisory Group on Australian Land Use Mapping (TAGALUM) was established as a steering committee whose primary function is to provide a coordinated and strategic approach to developing and managing the production of consistent land use datasets for the continent at catchment scale. The group consists of representatives from the Commonwealth, State and Territories agencies and assists in synchronising statewide data capture and mapping programs. The group is also responsible for establishing and standardising the Australian Land Use Mapping (ALUM) categories and ensuring the regional scale Land Use data is appropriately categorised. Each state member of the TAGALUM possesses significant expert knowledge of land use activities occurring in their respective states. Accordingly the TAGALUM is suitably qualified to provide peer review of the SPREAD land use outputs of the Murray Darling Basin.

Members of the TAGALUM met to review, discuss and provide comments on the series of Murray Darling Basin outputs from SPREAD I and II. TAGALUM's main objectives were to facilitate the verification process in assessing the allocation of land uses from the SPREAD 1 and II by:

- Examining the SPREAD I and SPREAD II outputs;
- Providing comment on the known distribution of Land Uses throughout the MDB;
- Using expert knowledge, highlight discrepancies between the spatial allocation of land uses from the SPREAD outputs to actual locations of land uses and agricultural activities;
- Providing comment and recommendations on ways to improve the allocation of Land Uses.

Key issues

The spatial allocation of land uses from SPREAD I and II were examined by the TAGALUM through visually assessing these results against catchment scale land use data, Landsat TM images and expert knowledge of the location of agricultural activities. Following this examination the following key inaccuracies pertaining to the SPREAD allocated distribution of land uses throughout the MDB were highlighted:

1. The far north western area of the of the Cobar SLA adjacent to the Darling River flood plain was designated as cropping. TAGALUM advice assured that no cropping activities occurred within that particular area of the SLA and were generally restricted to areas within the southern portions of the SLA. A visual inspection of this region using Landsat imagery showed no visible cropping zones along the floodplain but clear evidence of cropping in the southern zone;
2. Large areas of central western New South Wales in the vicinity of the Pilliga district were designated as non-agricultural land due to these forested areas containing crown cover greater than the 50% threshold set for agricultural and non-agricultural activity. It was indicated by the NSW TAGALUM representative that some grazing activity does occur within these forested areas with crown covers of up to 80%;
3. Areas flanking the South Australian western section of the Murray River at distances greater than 15km from the river were designated as horticulture. It was put forward that irrigated horticultural activities generally occur within 15km of the river. Visual verification along this portion of the Murray River using Landsat TM images confirmed this.
4. It was indicated that areas designated under the category 'Lake, river or marsh/wetland' may be used for grazing in periods of low or absent water conditions.

Improvements to land use allocation methods

Several changes to the SPREAD procedure and or the SPREAD outputs resulted from discussions with TAGALUM on the key inaccuracies of the initial SPREAD allocated land use outputs. Improvements to the 4 key inaccuracies listed above were achieved by the employing the following approaches:

1. Cropping pixels incorrectly allocated to the far north western area of the Cobar SLA were correctly allocated to the southern portion of the SLA. This was achieved by running SPREAD using NDVI profiles from all of the available control sites. A visual comparison using Landsat images of the area verified that this subsequent reallocation had greatly improved the positioning of the cropping pixels

2. Changing the 50% crown cover threshold for grazing activities to 80% will not be used regardless of grazing activities possibly occurring within forests and woodlands with up to 80% crown cover. All outputs however will carry a caveat stating that: "The definitions of potentially agricultural land and potentially agricultural holdings, assume that the compatibility of native vegetation with agricultural activities is determined by its crown cover. Specifically, it has been assumed that agricultural activities occur where the native vegetation crown cover is less than 50% and do not occur where the crown cover is greater than 50%. Thus it is generally assumed that non-forest and woodland are compatible with agricultural activities but that open forest and closed forest are not (*forest_type* = 3 or 4). There may be instances however where livestock grazing activities may occur where the crown cover exceeds 50% cover but generally less than 80% crown cover."

3. An irrigation constraint layer was constructed using best available data of actual and designated irrigation areas of the Murray Darling Basin. The incorporation of this

constraint layer as part of the SPREAD runs resulted in horticultural pixels being correctly allocated within the irrigation zones.

4. Areas defined under the land use category ‘Lake, river or marsh/wetland’ will still retain this category definition despite possible intermittent grazing occurring within these areas. The caveat “ Some areas designated under the Land Use category ‘Lake, river or marsh/wetland’ may be used for grazing activity during periods of low or absent water conditions.” will be placed on all outputs.

5.3 Testing Issues – land-use AVHRR-NDVI profiles input to SPREAD

SPREAD II results rely on the input of AVHRR NDVI signatures for particular commodities or land use types. Various approaches relating to the use of input signatures need to be tested. For example, the use of local, regional or noisy signatures will have a big impact on the results of the method. The issues related to these tests are discussed below and the conducted tests are summarized in Table 5.1. Once these issues had been tested and the results analysed, T7 (use of an irrigation constraint, and decreased “time of max. NDVI” bandwidth) was used for the project.

Table 5.1. Tests conducted by varying the input signature profiles. T1 – T7 are all SPREAD 2

Test Name	Description
T0	Original – Spread 1 using a limited number of local control site profiles
T1	Original – Spread 2
T2	all profiles
T4	increasing the kernel smoother bandwidth
T5	average signatures profiles
T6	Inclusion of profiles selected from CLUM bin 10 data
T7	Use of an irrigation constraint, and decreased “time of max. NDVI” bandwidth

Test T0 - SPREAD 1 using a limited number of local control site profiles

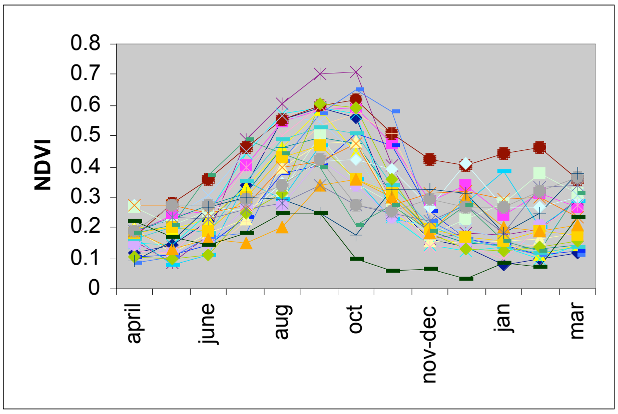
This test is the original SPREAD procedure that uses control-site profiles that are proximal to the target pixel, at similar latitude and with a constraint on the minimum number of control sites. This minimum number was 2 and this could translate to 6 profiles if the land use was the same for the three years of control-site information collection.

Test T1 - SPREAD 2 using a limited number of local control site profiles

Here, SPREAD 2 is used as an improvement on the methodology of SPREAD 1, but using the same signature profile inputs as test T0. The details of these two approaches are given in Section 2.

Test T2 – all profiles

Clearly if the closest sites have noisy NDVI profiles (as discussed in section 4.2), the probability assessment will be incorrect. The question is, with the current implementation, how many control sites are needed to reduce this noise effect and improve accuracy. AVHRR profiles for wheat harvested in October or November in a single 12 month period (1996/1997) are shown in Figure 5.1(a). The variation could be explained by the natural NDVI variability of a wheat crop but when compared with the average wheat signatures (Figure 5.1b), which are consistent between different years, it is apparent that the differences in Figure 1a are due to the presence in the control-sites of other non-wheat materials. Given that wheat cereal control-sites are likely to be one of the most pure (based on the preferred size of control site selection – see section 4.2), there will be significant problems due to noise for many other commodities.



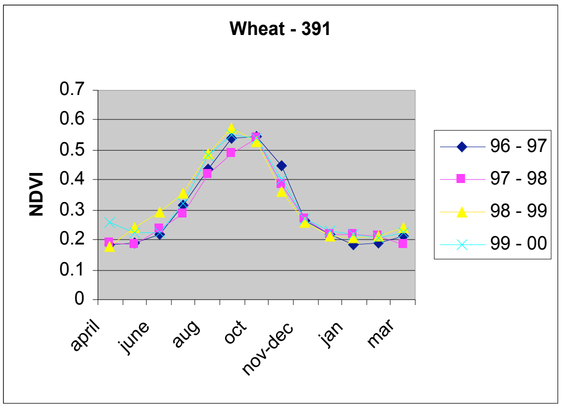


Figure 5.1. (a) AVHRR profiles for wheat harvested in October or November in a single 12 month period (1996/1997) (b) average wheat signatures for different years

Test T4 – increasing the kernel smoother bandwidth

One variation that needs to be tested is the statistical bandwidth (see method section 2). Test T4 was conducted using limited numbers of control sites as in the original version (Test T1) but the probability density functions, that characterize the control site profiles for each commodity, were made smoother, more spread out and less spike. This was done by increasing the kernel smoother bandwidth from 100 to 500.

Test T5 – Using average signatures profiles

In this test, average signatures for commodities were used as input to the process. This was done at the Audit commodity level (21 classes) (general crop-types, e.g. legumes) as opposed to the ABS level 3 classification (specific crop-types, e.g, soybeans, peanuts). Figure 5.2 shows average signatures for a number of commodities. The concept behind this test was that the averaging of profiles should reduce noise associated with non target materials in the control-sites.

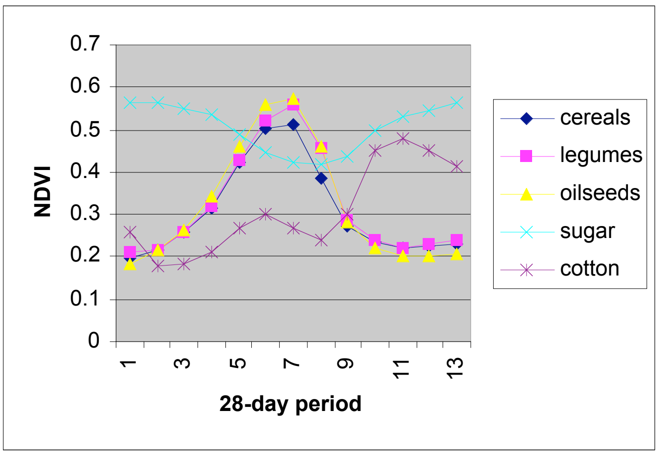


Figure 5.2. Average AVHRR-NDVI profiles for various crop-types

Test T6 - using profiles selected from CLUM bin 10 data

Another way of generating signature profiles was to use catchment scale land use mapping (CLUM) data to identify 1 km AVHRR pixels that were covered by a particular land-use or crop type. The CLUM polygon data was first converted to 25m pixels and compared with the 1 km image data. AVHRR pixels were selected where the 25m pixels for a unique land-use covered greater that 90% of the 1km pixel. Unfortunately only a limited number of Audit commodity classes (the NLUM - SPREAD target commodities) can be equated to the CLUM data which is classified according to ALUM version 5. Signature NDVI profiles were extracted for the following commodities:

1. Cereals – dryland and irrigated
2. Oil seeds - dryland
3. Cotton - dryland and irrigated
4. Legumes - dryland and irrigated
5. Irrigated vine fruits (grapes)

A comparison between the control site data and the CLUM derived profiles is shown in Figure 5.3. Figure 5.3 (b) shows a group of 20 randomly selected profiles from the CLUM derived data for cereals compared with the control-site derived cereals (Figure 5.3a). The CLUM profiles are much closer to the mean (see Figure 5.1b) and are clearly less affected by variations likely due to noise from other materials. The means and standard deviations for the two types of selection methods are show in Figure 5.4 for wheat and cotton. Standard deviations are lower for the CLUM method this also indicates less noise in the profiles. The cotton CLUM-derived mean lacks a second peak which appears to corrospond with the greenness period of other crops that are likely to be contributing to the CS mean. So generally it appears that the CLUM-derived profiles are more pure.

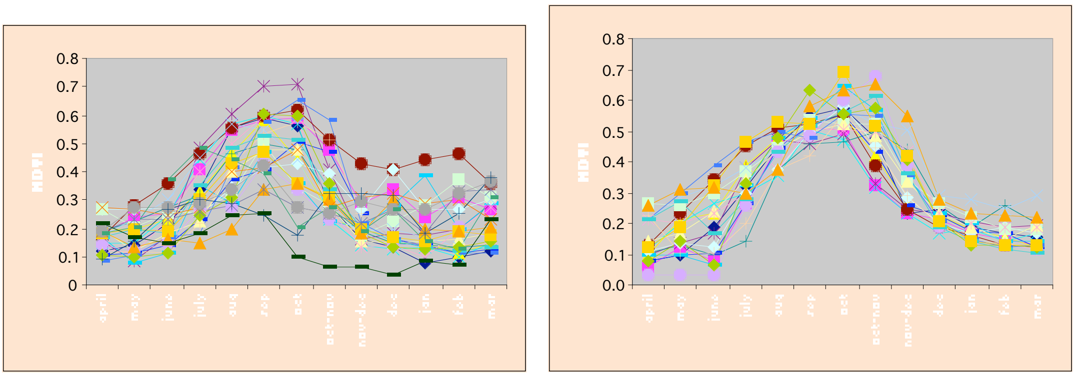


Figure 5.3. Comparison of (a) control site derived wheat profiles (same as Figure 1) and (b) CLUM derived profiles for cereals

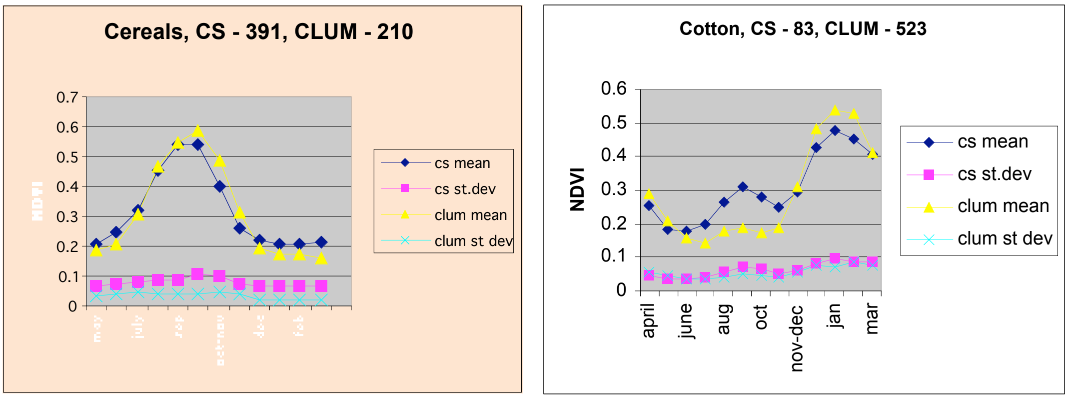


Figure 5.4 Comparison of the mean and standard deviations of commodity AVHRR NDVI profiles derived from control sites (cs) and from catchment scale land use mapping (CLUM) for (a) cereals and (b) cotton. The numbers of available profiles are indicated at the top of the charts.

Test T7 - use of an irrigation mask constraint

This test was conducted largely in response to the results of earlier tests. An irrigation mask for the MDB was available to the project and this was used to constrain the spatial distribution of irrigated crops (see section 2 for method description). For this particular test there was no allowance made in the probability assessment to allow for inaccuracies in the irrigation mask. Another change was that the bandwidth for the “time of maximum NDVI” was made considerably smaller.

5.4 Accuracy of the results

Independent data for accuracy testing

The data sets that can be use for verification include catchment-scale land use (CLUM) and Landsat TM satellite imagery for the years 1992, 1995, 1998, 2000 and 2002. The verification method is based on the NLUM 1996/97 methodology reported in Stewart et al (2001).

Catchment scale land use mapping (CLUM) appropriate for verifying the NLUM data are available for a significant proportion of the Murray-Darling Basin (MDB) and are predominantly located in the southern and eastern pasts of the basin (figure 5.5). These data have been translated into the ALUM classification version 5. The ALUM classification is a hierarchical one with:

- Primary levels indicating level of intervention in the landscape eg 3.0.0 Production from dryland agriculture and plantations
- Secondary levels indicating broad land use classes and are the minimum level of mapping required eg 3.3.0 Cropping
- Tertiary levels have been aligned to ABS agricultural census and usually require field verification or another source of data for attribution eg 3.3.1 Cereals
- Commodities have in practice been restricted to the requirements of the mapping agency, for example identification of pests in apples and have been attributed at a quaternary level eg 3.3.1.1 Wheat

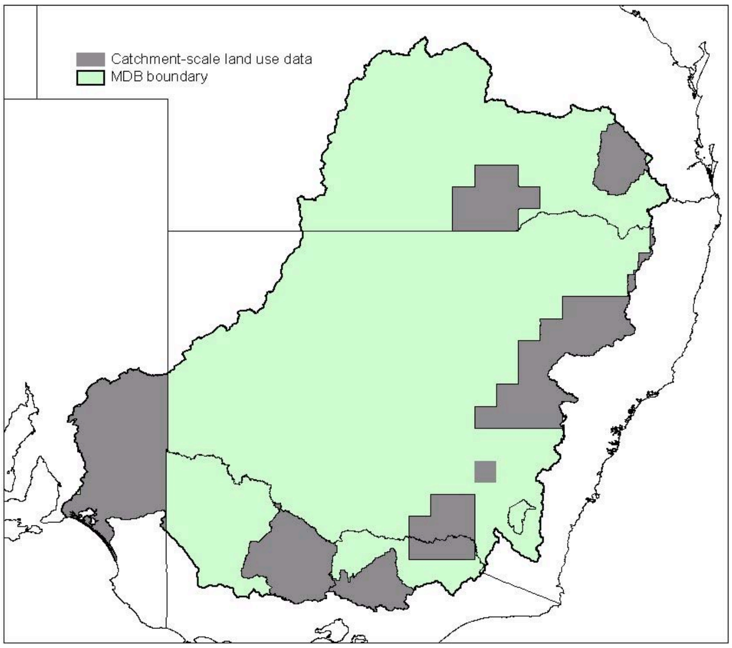
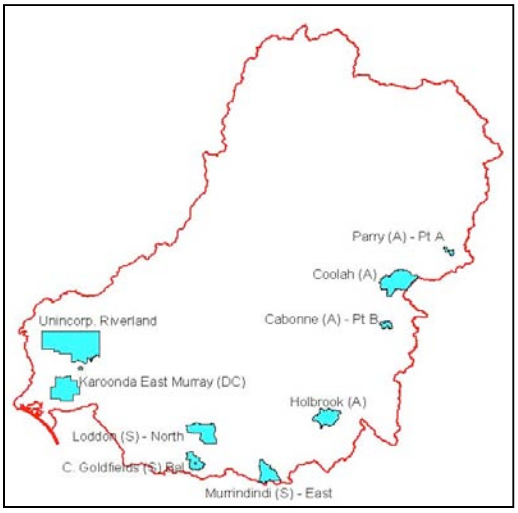


Figure 5.5. Catchment scale land use mapping (CLUM) available for verifying the NLUM data

A large number (375) of statistical local areas (SLA's) intersect with the MDB and nearly half of these intersect with the catchment-scale land use data.

Matching SPREAD outputs to CLUM data

Representative SLA's were selected that were wholly contained within the CLUM and covering the full extent of the data. Ten SLAS were selected and are shown in Figure 5.6



Outputs from spread were coded via a look-up table so that mapped commodities were translated to the ALUM code represented in the CLUM. In most cases, the NLUM data was allocated codes at the tertiary level. Since most of the catchment-scale data in the MDB indicate the land use at a secondary level, this required also aggregating the SPREAD data to similar levels. The CLUM data was further re-sampled from approximately 25m pixels to 1 km pixels based on the most common land use.

Figure 5.6. SLA's used for comparison of CLUM mapping and NLUM SPREAD data.

The CLUM and NLUM datasets were overlain and pixel-by-pixel comparisons carried out. For the classified NLUM data and probability maps, areas were cross-tabulated to analyse for correct and incorrect pixels, the omissions and commissions and accuracy of classification.

Accuracy of SPREAD allocation maps

SPREAD allocation maps for the various tests were compared with the CLUM 1 km data at the primary, secondary and tertiary levels of land-use classification. For the allocation maps, the proportion of NLUM pixels that were allocated correctly (according to the CLUM data) was determined. These values or "accuracies" are shown in Table 5.2 along with the number of pixels analysed.

Table 5.2. Accuracy assessment of NLUM allocations based on resampled CLUM data. Descriptions of tests t0 – t6 are given in Table 5.1

Level 1

	Proportion of NLUM pixels confirmed by CLUM			Number of NLUM pixels analysed		
Test	non-ag	pastures	crops	non-ag	pastures	crops
t0	0.733	0.806	0.295	10842	46800	7624
t1	0.737	0.814	0.350	10851	46512	7903
t2	0.736	0.821	0.412	10850	46607	7809
t4	0.736	0.819	0.394	10851	46493	7922
t5	0.736	0.820	0.404	10849	46539	7878
t6	0.736	0.822	0.404	10848	46514	7904
t7	0.737	0.823	0.408	10851	46547	7868

Level 2

	Proportion of NLUM pixels confirmed by CLUM				Number of NLUM pixels analysed			
Test	native	modified	crops	perennial hort	native	modified	crops	perennial hort
t0	0.617	0.617	0.289	0.191	38539	8261	7498	110
t1	0.624	0.573	0.348	0.114	37676	8836	7774	114
t2	0.641	0.602	0.412	0.070	37645	8962	7679	115
t4	0.641	0.605	0.395	0.061	37690	8803	7791	115
t5	0.644	0.615	0.402	0.223	37487	9052	7753	112
t6	0.638	0.605	0.400	0.254	37812	8702	7771	118
t7	0.642	0.578	0.403	0.304	37280	9267	7738	115

Level 3

	Proportion of NLUM pixels confirmed by CLUM		Number of NLUM pixels analysed	
Test	Cereals	Cotton	Cereals	Cotton
t0	0.127	0.361	986	402
t1	0.188	0.612	835	412
t2	0.187	0.633	845	417
t4	0.183	0.588	814	434
t5	0.184	0.162	839	377
t6	0.166	0.195	856	338
t7	0.182	0.659	829	416

Level 1 accuracies

At level 1, crops include horticulture. Table 5.2 shows that the SPREAD 2 results (T1) are better than SPREAD 1 (T0). The T2 test (all profiles) provides the best results at level 1 with an improvement for cropping of 6% on the original results. The accuracy (0.41) is relatively low in a statistical sense. However there are issues with respect to comparing the dominant CLUM land-use at 1 km that might reduce the classification accuracy. A further test was conducted on only those pixels where cropping was greater that 90%. These pixels were found by using the CLUM data gridded at 25m, and the accuracy of T2 cropping improved to 0.45. The CLUM data is also sometimes mapped on infrastructure alone and some of these data may not relate precisely enough in time to the collection of the satellite data. Unfortunately the amount of CLUM data available for verification is limited although it serves as an indication and a means of comparing results. At level 1 pasture results are good while surprisingly the accuracy of non-agricultural layers (derived from various map sources) is not as good as expected. Clearly there are some problems due to noise present in control-site profiles and this is amplified in some SLA's where only limited numbers of noisy profiles were used. An example of this is shown in a visual comparison of t1 and t2 outputs for the Cobar (A) SLA (SLA code 135151750) with 2001 Landsat imagery (Figure 5.7). The Landsat image (Figure 5.7a) strongly suggests that cultivated fields are confined to the southern part of the SLA while the north-western part of the SLA is occupied by the Darling River flood plain. The t1 summary map (Figure 5.7 b), based on a run of SPREAD II using limited numbers of control sites, incorrectly allocates the land uses involving cultivation in the Darling River flood plain in the north-western part of the SLA. Figure 5.7(c) shows that the t2 summary map, based on a run of SPREAD II using all control sites, more correctly allocates the land uses involving cultivation in the southern part of the SLA.

(a) (b) (c)

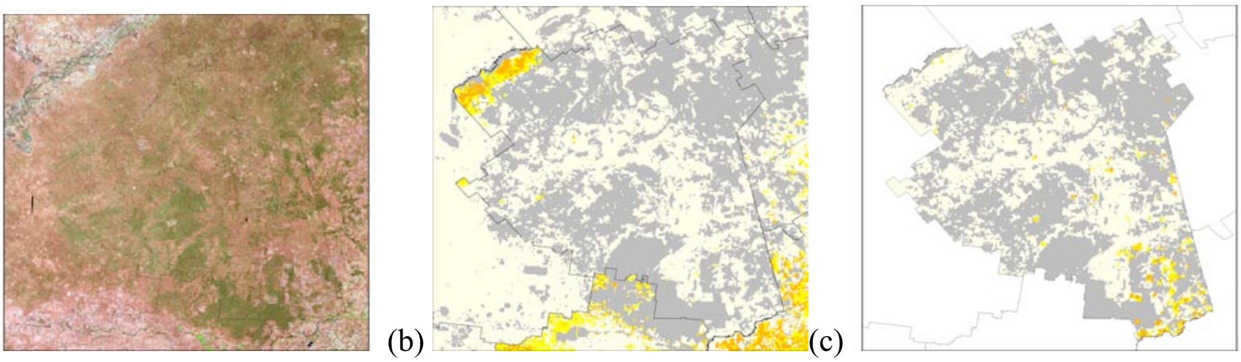


Figure 5.7. (a) Landsat image showing the Cobar - black lines are SLA boundaries (A) SLA in 2001. (b) The t1 summary map for the year 2000 - pale yellow, residual/native pasture; orange, sown pastures; deep yellow, cereals and oilseeds; grey, non-agricultural land (mainly open forest); white, no data. (c) The t2 summary map of the Cobar SLA for the year 2000.

At level 2 pastures have been separated into native and modified while cropping is divided into horticulture and other cropping. Unfortunately there was not enough data on seasonal horticulture for the analysis. Again for cropping, the T2 test is the more accurate. However, for perennial horticulture, there is a significant improvement in accuracy of identification for tests T5 and T6. This is likely because these tests use more realistic AVHRR profiles for perennial horticulture subclasses than those provided by the control sites. In some areas there are clear advantages in using CLUM derived profiles. An example of this is seen along the Murray River in S.A. near the border (Figure 5.8) in the Loxton Waikerie SLA. The SPREAD T6 test allocation for grapes (shown in red in Figure 5.8b) closely corresponds with the CLUM mapping (Figure 5.8a) while the T2 results show grapes as dispersed and unrealistically present in mallee areas.

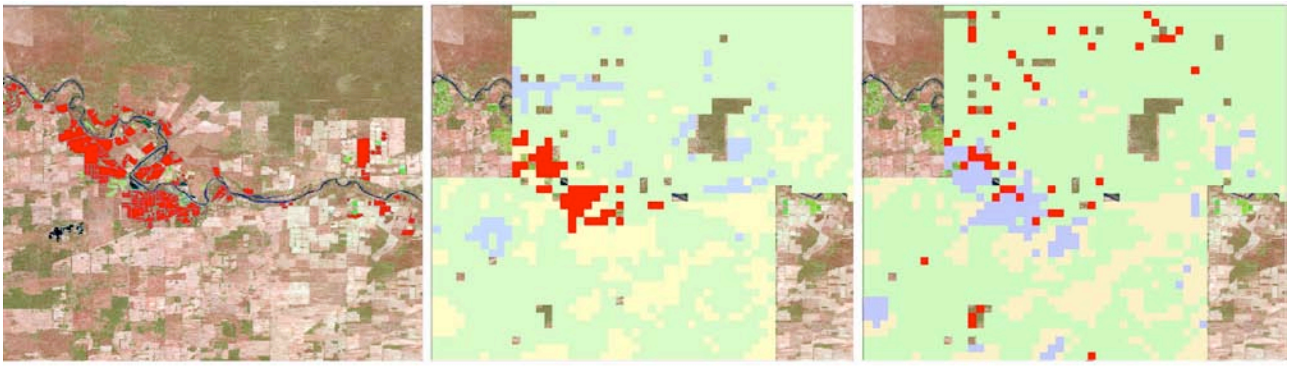


Figure 5.8. NLUM results for mapping grapes along the Murray River in S.A near the Victorian border. (a) CLUM mapping of perennial horticulture (mostly grapes – Russell Flavel pers com.) shown in red. Grapes are also shown in red for SPREAD results from (b) test T6 and (c) test T2. Cereals are shown as yellow, sown pastures – blue and native pastures – green.

Level 3 CLUM data is scarce within the basin and the verification procedure is complicated by the fact that crops are variably mapped to both level 2 and 3, sometimes within a survey area. Table 5.2 shows accuracy results for the NLUM allocated pixels for cereals (excluding rice) and cotton. The cereal data was taken from two CLUM survey areas, North Central and Goulburn-Broken, where the mapping was done mostly at tertiary level. Best results (15% accuracy) were again achieved for test T2. These values are generally poor probably due to the overlap of cropping signatures using the NDVI data alone, although some of the cereals here may have been mapped as "cropping" at the secondary level, reducing the accuracies. Figure 5.9 shows the distribution of "ground truth" classes for cereal and cotton pixels from the T2 test. From this it appears that modified pastures were commonly misidentified by SPREAD as cereals. For cotton the accuracies are high at 63% (Table 5.2) although some grazing areas are mistaken for cotton.

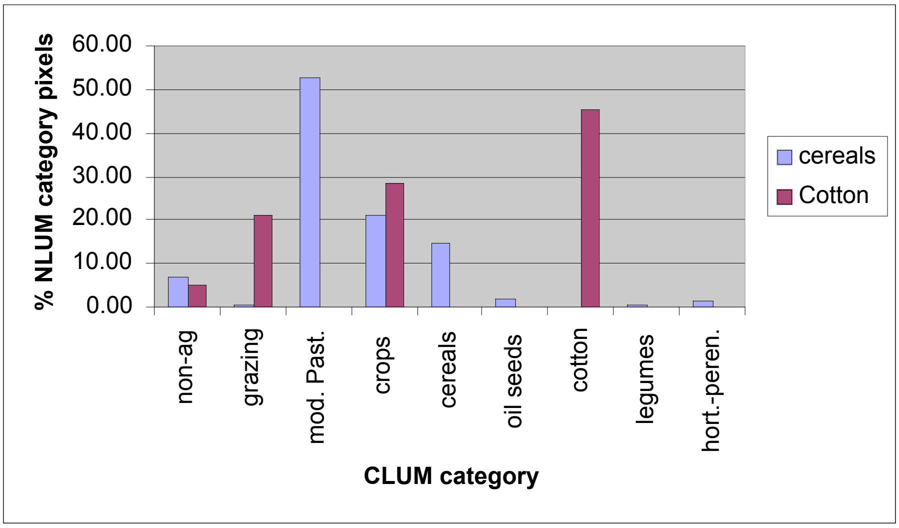
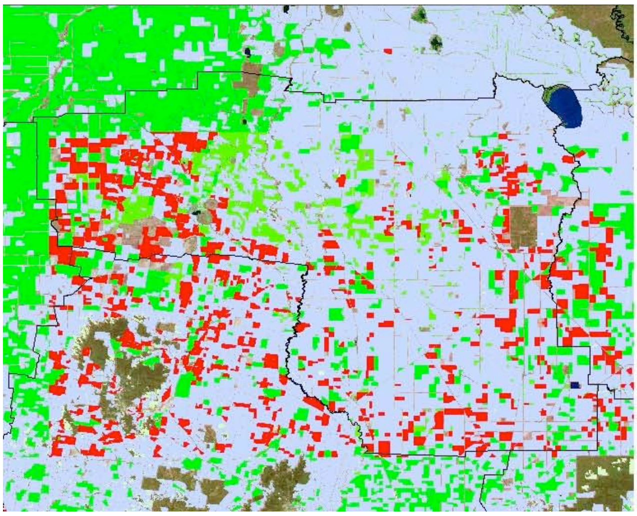


Figure 5.9 CLUM classes for NLUM T2 allocated cereals for the North Central and Goulburn Broken areas and cotton pixels.

An example of NLUM cereal allocation versus CLUM ground mapped data is shown in Figure 5.10 In general, the cereal allocations from the SPREAD-AVHRR method (Figure 5.10b) are similar to the mapped areas of cereals. However some of the cereals have been mapped in CLUM only as cropping and this illustrated the difficulty in assessing the accuracies of NLUM at tertiary class level. Also as illustrated in Figure 5.9 above, significant areas of modified pasture have been allocated to cereals by the SPREAD NLUM method.

Although spatial accuracy results are low, it appears that, on a broad scale, the NLUM mapping is reasonable.



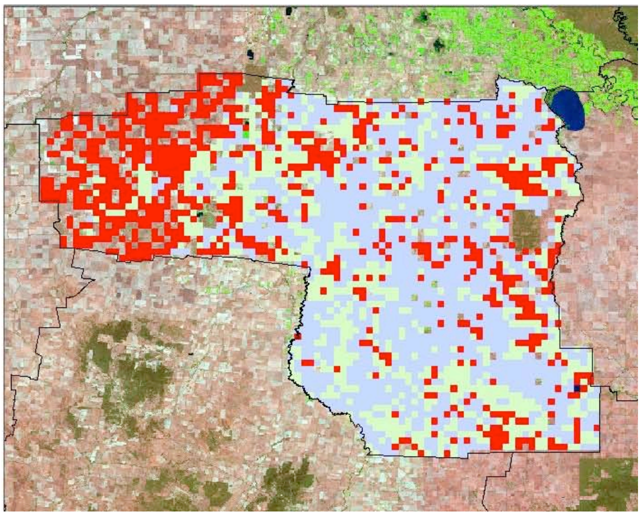
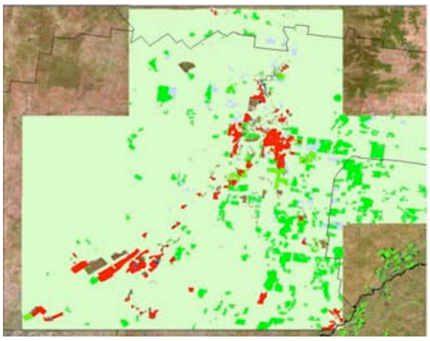
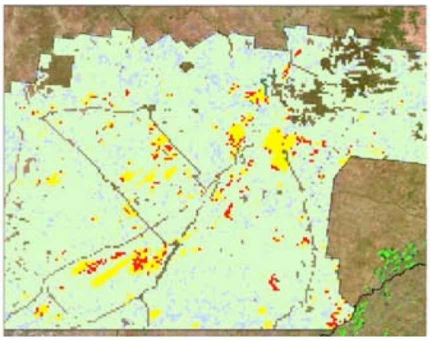


Figure 5.10 (a) CLUM mapping of cereals in North-Central, Victoria survey area and (b) NLUM results for the T2 test. Secondary level mapping in (a), i.e. cropping, is shown as green. In both (a) and (b) native pastures are shown as light yellow and modified pastures as grey-blue. A Landsat TM composite is shown as the background for areas not represented

The cotton results for the T6 test using CLUM-derived profiles are much less accurate at only 19.5% and the reasons for this are unclear particularly since areas that were used to derive cotton signatures are not being mapped by the SPREAD method. Only two CLUM surveys (Balone and Condamine) contained Cotton mapped at the tertiary level. A comparison of results are shown for the Balone area in Figure 5.11. In this SLA, the T2 test provides accurate mapping of the cotton growing areas whereas the T6 test shows poor results. It this test, the cotton areas are being mapped as cereals (yellow) indicating that there is confusion between the cereals and cotton signatures. A comparison of wheat and cotton NDVI phenology profiles (see Figure 4.2) shows that there are distinct differences and that SPREAD should be able to distinguish between these crop-types. From this it was apparent that the bandwidth for “time of maximum NDVI” was too large and that many different crops were not being effectively discriminated. It was decided that further testing should include a modified bandwidth for this parameter.





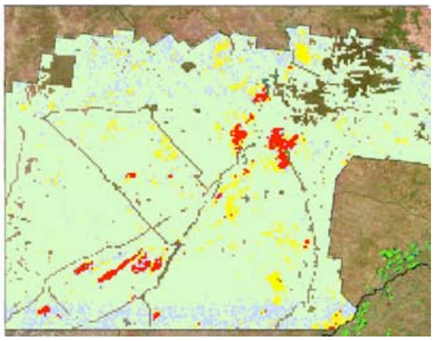


Figure 5.11 Comparison of cotton data in the Balone area, N.S.W for (a) CLUM mapping and (b) NLUM results for the T2 test. . Cotton is also shown in red for SPREAD results from (b) test T6 and (c) test T2. Secondary level mapping in (a), i.e. cropping, is shown as bright green. In all cereals are shown as yellow, native pastures as light green and modified pastures as grey-blue. A Landsat TM composite is shown as the background for areas not represented

At this stage of the testing, time was running out to finalise this stage of the NLUM product. It was concluded that the best strategy was to produce a last composite test based on previous evidence. This was to effectively reproduce the T2 test, but with an inclusion of two variations, i.e. the use of (1) an irrigation constraint and (2) a significantly smaller bandwidth for “time of maximum NDVI”. Table 5.2 indicates that accuracies are similar at level 1, while there has been improvement in the results for perennial horticulture and cotton. This improvement for these generally irrigated land-uses indicates that constraining by irrigation areas is a useful procedure. Visual results of this last test (T7) are shown in Figure 5.12 for the two areas displayed in Figures 5.8 and 5.11. While cotton allocation in Figure 12(b) is consistent with CLUM mapping, the grape allocations are not totally accurate. The grapes were well allocated in test T6 and these should have been used for T7 but were left out by mistake. Nevertheless the use of the irrigation constraint improves the allocation compared with no constraint (T2) (see Figure 5.8c). It is likely that the new bandwidth for “time of maximum NDVI” has improved the allocation, although with the current stage of testing there is no real evidence for this. Further testing using improved signature profiles will be conducted (see recommendations).

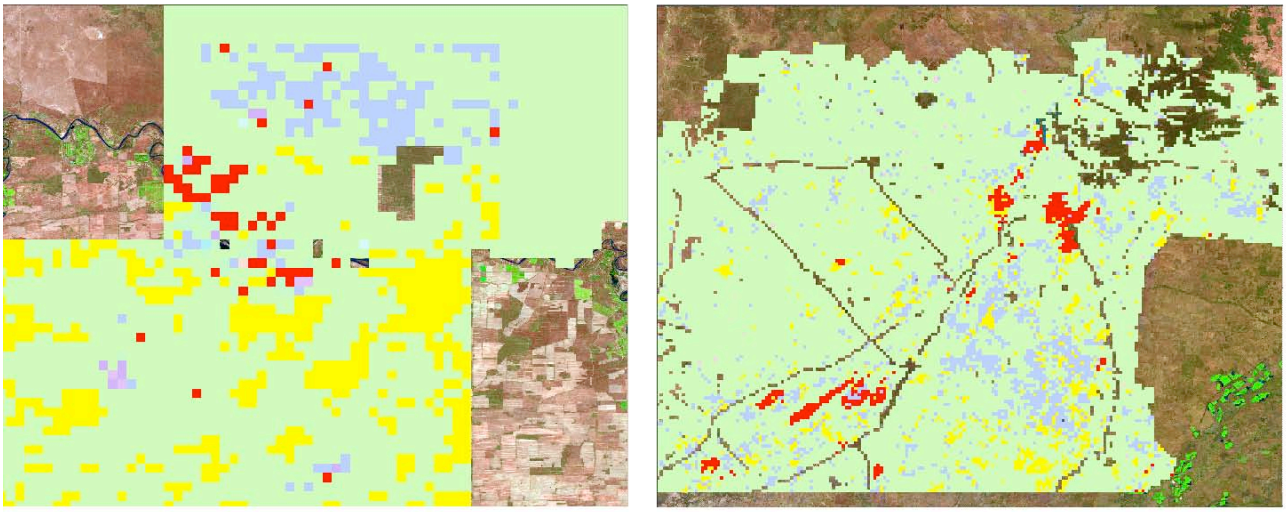
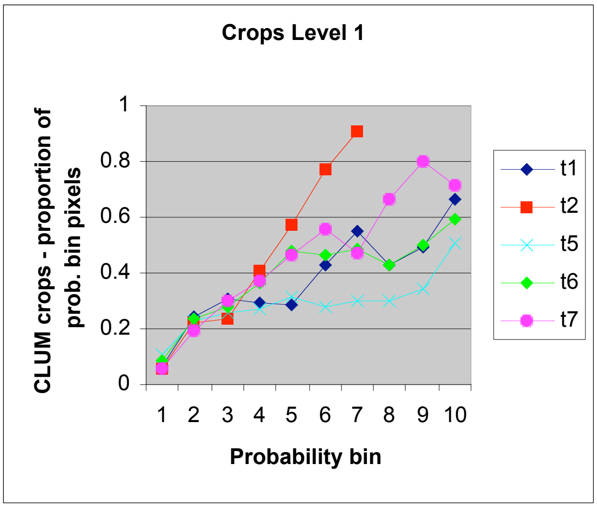


Figure 5.12. SPREAD T7 test results for (a) Loxton Waikerie (grapes = red) and (b) Balonne (cotton = red). In both images, cereals are yellow, native pastures are light green and modified pastures are pale blue. Compare with Figures 5.8 and 5.11 respectively.

Although, as above, it is important to verify the final allocated class mapping, it is possibly more important to assess the probability maps that are used to make these allocations.

Accuracy of SPREAD probability surfaces

Probability surfaces derived from SPREAD were also aggregated into ALUM codes at three levels for comparison with the CLUM data. To test the accuracy of the probabilities, these values were aggregated into bins where bin 1 = 0 - 0.1, bin 2 = 0.1 - 0.2 etc. For each bin, the proportions of total number of pixels, that were mapped by CLUM as the same land-use, were determined. In this way the probabilities and the CLUM were compared and results are shown in Figure 5.12 for both cropping and pastures for the various tests at level 1.



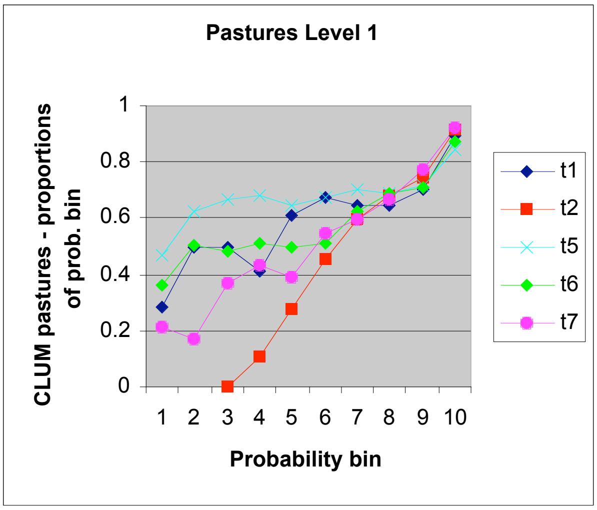


Figure 5.12 The proportion of binned probability data mapped in the CLUM data as (a) crops and (b) pastures.

The numbers of pixels present in each probability bin are given in Table 5.3 Generally for all tests, results show that for the higher the probability of crops or pastures, the more likely the CLUM mapping will be crops or pastures respectively. In Figure 5.12 (a) the T2 test probabilities are lower than the other tests but the linear result and the 90% accuracy of the highest probabilities for both crops and pastures indicate that this test has produced the best results and that the range of probability values are realistic. Other tests, such as the use of average profiles (T5) and CLUM derived profiles (T6) have produced poorer results. It was though that the reason for this was that the bandwidth used for "time of maximum NDVI" was too large and that crop-types with different growing seasons, such as cereals and cotton, were not being effectively discriminated. This would explain the fact that average or CLUM-derived profiles produced worse results, both visually (see Figure 5.11b) and statistically (see cotton results - Table 5.2). As mentioned earlier, time was running out to deliver this stage of the NLUM and the T7 test was designed to include an irrigation constraint and a modified bandwidth for "time of maximum NDVI". The analysis of probability data for T7 shown in Figure 5.12 indicates that the results are as good as T2, but generally better than other tests. One possible explanation in the slight degradation of the results is that the data are too tightly constrained to the irrigation mask that contains inaccuracies. It would be preferable to test each of the new components of the T7 and this is recommended for future development. Also given that the earlier tests T5 and T6 were not optimum, these should also be repeated.

Table 5.3 Numbers of AVHRR pixels falling into probability bins for level 1 crops and pastures. Bin 10 represents probabilities of 0.9 – 1.0

Probability	**Crops**					**Pastures**				
NLUM Bin	**t1**	**t2**	**t5**	**t6**	**t7**	**t1**	**t2**	**t5**	**t6**	**t7**
1	33255	33103	41220	36731	30031	516	0	4088	1606	541
2	6501	6062	2355	3765	7081	452	0	1245	481	228
3	5330	6943	1439	3153	5822	512	2	984	438	186
4	3369	6522	1102	2502	4770	933	29	928	529	280
5	3063	2053	939	1084	2288	1770	402	924	747	983
6	1234	140	867	614	621	3379	3839	979	1636	3200
7	815	11	949	486	220	3792	6719	1188	2866	5496
8	466	0	1028	438	214	6118	6995	1624	3372	5909
9	499	0	1430	557	237	6726	8760	3502	4557	7930
10	302	0	3505	1379	436	30636	28088	39372	34477	26967

5.5 Conclusions

Verification and accuracy testing of the NLUM SPREAD data has been achieved using catchment scale land-use data (CLUM). The latter data has been used to assess both the final SPREAD allocations and also the probability values for the various agricultural land-uses. Unfortunately the amount of CLUM data available for verification is limited although it serves as an indication and a means of comparing results.

In general, the SPREAD 2 results show a significant improvement in accuracy over SPREAD 1. Accuracy results are improved by using a large suite of NDVI signatures for a particular commodity. At the primary level of classification, best spatial accuracies at 1km resolution are suggested at around 50% for cropping and over 80% for pastures.

At the tertiary or crop-type level, the CLUM data is scarce within the basin and the verification procedure is complicated by the fact that crops are variably mapped to both secondary and tertiary levels. The use of CLUM-derived NDVI profiles significantly improved the accuracy for grapes, since the original method had the difficulty of finding control-site areas large enough to derive 1km pixel AVHRR NDVI profiles. However possibly inappropriate bandwidths for "time of maximum NDVI" led to worse results for other crop-types. A final approach (T7) was adopted that included CLUM grapes profiles, a modified bandwidth and an irrigation constraint.

The constraint of concordance with ABS statistics means that the accuracy at a coarser resolution is much higher. Accordingly, this data is appropriate only for use in applications where high accuracy is only required at low spatial resolution (see caveats section).

6. Conclusions and recommendations

The main objectives of the project were to apply the already established SPREAD methodology to produce broad-scale land use maps for the Murray Darling Basin in time slices for the period 1994-2001. At this stage, it was planned to include improvements to the method (SPREAD II) and the development of procedures for parameter/input testing and validation. These improvements have been developed although the restricted time frame of the project has meant that only limited research and development could be achieved.

Output land-use maps generated by the project are statistically and spatially accurate to the Statistical Local Area (SLA) level due to the spatial constraint to census agricultural data. Verification and accuracy testing of the finer-scale NLUM SPREAD allocation data suggests that, at the primary level of classification, best spatial accuracies at 1km resolution are around 50% for cropping and over 80% for pastures. Accuracy results are complicated by the limited amount catchment scale land-use data (CLUM) for the basin and the lack of detail of this data at a tertiary level. Inconsistencies in the exact time of mapping for the two methods are also likely to be a source of reduced accuracy results. Analyses of SPREAD II probability surfaces show that the amount of correct allocation linearly increases with the probability values. One test achieved 90% accuracy for cropping for the highest binned probability value. Accuracy results are improved by using a large suite of NDVI signatures for a particular commodity.

A valuable component of the project is the control-site database that allows for the analysis and characterisation of crop-types in terms of NDVI time-series signatures. The database has ancillary information that can be used to determine the effects, of location and farm practice etc, on crop signatures. As a result of the analysis of control-site data, a number of problems with the current SPREAD approach have been observed and these can be translated into potential improvements. The complimentary use of CLUM data has also been identified as an important technique for deriving less-noisy NDVI profile signatures.

Current NLUM mapping is targeted at the level of Audit commodity classes. There are a number of problems with this in relation to attempts to characterise the NDVI phenology of the Audit classes. One problem is that these classes may often contain a number of different surface types with variable phenology. For example, cereals include wheat, barley, grain sorghum and maize; legumes include soybeans, peanuts and lupins; oilseeds include canola and sunflowers, and these specific crops actually have differing NDVI characteristics. Another problem is that crop-types falling within different classes may have similar profiles. An example of this is that wheat, lupins and canola have average time-series NDVI signatures that are indistinguishable. Attempting to identify materials at the current Audit commodity class level is likely to be contributing to lower accuracies.

Improvements to the primary satellite data input should also have a major effect on map accuracies. The current method does not incorporate available AVHRR surface temperature data that can provide important information about evapotranspiration and soil moisture. Research is currently underway to incorporate these data. Also, the accuracy of the NLUM mapping product will be significantly enhanced using data from the new MODIS sensor since these data have improved spatial and radiometric characteristics. Currently there are problems with the pre-processing of the AVHRR data as the ERIN procedures often identify unrealistically large amounts of invalid data and subsequent splining has produced no-data artefacts throughout the image sequence. There is also a dependence on ERIN for processing of primary data and an advantage with MODIS is that there should no problems with the supply of calibrated and cloud-free NDVI data.

Recommendations:

There are a number of recommendations that are likely to contribute to improvements to the NLUM product:

1. The current land-use classification scheme is not considered optimal and is likely to be contributing to lower accuracies. More research is required to fully evaluate what level of detail of land use is achievable using satellite data in combination with agricultural census statistics. This should include a study of the degree of separation that certain metrics can produce and accordingly a modified classification scheme that is both practical and accurate. It is recommended that a simpler scheme, perhaps more aligned with secondary land-use classes, be determined.

2. Further investigation should be conducted into the use of AVHRR surface temperature data (Ts) for historical NLUM products.

3. More testing of SPREAD outputs is required in relation to input constraints. These constraints include topography, spatial rainfall patterns, an irrigation mask, the use of particular NDVI or Ts metrics and varying the statistical bandwidths of these metrics. Another potential constraint is land cover that can be derived by initial stage processing of the satellite data.

4. The most recent CLUM data needs to be incorporated for verification and accuracy testing and this is an ongoing process.

5. Migration of the method to the new MODIS sensor, for future NLUM outputs, is recommended since these data have improved spatial and radiometric characteristics.

7. References

Australian Bureau of Agricultural and Resource Economics, 1997. *1996 – 97 Farm Survey*.

Bureau of Rural Sciences, 2002. *Land use mapping at catchment scale: principles, procedures and definitions*, Edition 2, Bureau of Rural Sciences, Canberra.

Stewart, J B, Smart, R V, Barry, S C and Veitch S M, 2001. *1996/97 Land Use of Australia: final report for project BRR5*, National Land and Water Resources Audit, Canberra, available online http://audit.ea.gov.au/ANRA/atlas_home.cfm

Walker, P A and Mallawaarachchi, T, 1998. 'Disaggregating agricultural statistics using NOAA-AVHRR NDVI', *Remote Sens Environ*, 63:112 - 125.

Appendix 1 Caveats for the 1993, 1996, 1998 and 2000 Land Use of the Murray Darling Basin

1. The purpose of these maps is to provide basin-wide representations of major commodity types for mapping and display, and spatial input to basin-wide hydrologic and salinity modelling.

1. Finer resolution land use data are available for many areas of the Murray-Darling Basin and, when appropriate, should be used in preference to the1993, 1996, 1998 and 2000 Land Use of the Murray-Darling Basin, Version 2.

2. The land use maps should be used at an appropriate scale (nominally 1:1,500,000). For the agricultural land uses, the summary maps cannot be expected to have high attribute accuracy on a pixel-by-pixel basis (each pixel is ~ 1.1km).

3. Attribute accuracy is likely to be particularly low for pixels in the summary maps representing agricultural land used for more than one commodity group. This can occur where different commodity groups are close in space (strip cropping in particular and small scale planting in general) or in time (multiple cropping). Attribute accuracy is generally dependant on how distinct the commodity appears in the satellite image. The most distinct commodities include primary level classifications and some homogeneous irrigated agricultural types.

4. Agricultural census data and, in the case of the 1998 map, agricultural survey data supplied by the Australian Bureau of Statistics (ABS) provide the areas of each commodity group in each Statistical Local Area (SLA) that were built into the maps. It should be noted that the ABS data were processed on the basis of various assumptions during construction of the maps, as discussed in the project report. The maps should therefore be used with appropriate caution. In relation to the 1998 map, it should, further, be noted that the agricultural survey data include many records with large relative standard error (exceeding 50%). This is because the ABS designs its agricultural surveys to be reported on larger areas than SLAs and only releases agricultural survey data at SLA level by special request. The 1998 map should therefore be used with additional caution.

5. The native forest data set used in the construction of the maps includes a crown cover attribute with three crown cover classes: 20 – 50% (woodland), 50 – 80% (open forest) and 80 – 100% (closed forest). For agricultural land with native forest cover it has been assumed that grazing will be the dominant land use if the crown cover is between 20% and 50% but that grazing will be subordinate to conservation ('other minimal use' in terms of the Australian Land Use and Management Classification, Version 4) where the crown cover exceeds 50%. It might, arguably, be more appropriate to assume that grazing will be dominant when the crown cover is as high as 80%. Users of the maps should be aware that grazing might have been the dominant land use, from time to time, in some areas classified as 'other minimal use'. Some cross state border differences in crown cover will be evident; these are artefacts of separate interpretations.

6. Non-perennial and perennial hydrographic features have not been distinguished. Users of the maps should be aware that grazing might have been the dominant land use, from time to time, in some areas classified as 'lake', 'river' or 'marsh/wetland'

Appendix 2 BRS Technical Bulletin Regional Land Use Mapping Using Satellite AVHRR Data

Appendix 3 BRS Technical Bulletin. National Land Use Mapping Project: SPREAD 2 Methodology – Testing and Validation

Appendix 4 Metadata

Title 1993, 1996, 1998 and 2000 Land Use of the Murray-Darling Basin, Version 2
Custodian
Bureau of Rural Sciences

Jurisdiction
Australia

Abstract
The 1993, 1996, 1998 and 2000 Land Use of the Murray-Darling Basin, Version 2, is a time series of land use maps of the Murray-Darling Basin for the years 1993, 1996, 1998 and 2000. Temporal variation is shown for agricultural land uses only. The non-agricultural land uses are based on existing digital maps. The agricultural land uses shown in each map are based on the Australian Bureau of Statistics' agricultural census or survey collection with appropriate reference period. The spatial distribution of agricultural land uses is interpretive and has been determined using AVHRR satellite imagery with ground control data. The maps are supplied as a set of ARC/INFO grids with geographical coordinates referred to WGS84 and 0.01 degree cell size. For each year there is a set of probability maps, one for each agricultural land use, and a single summary map showing the non-agricultural land uses and a likely arrangement of the agricultural land uses. The arrangement of agricultural land uses in the summary map was determined from the probability maps using some simple rules to make an approximation to a maximum likelihood land use map. As supplied the probability maps are floating point grids with cell value between 0 and 1 and no value attribute table while the summary map is an integer grid with a value attribute table with attributes defining the agricultural commodity group, irrigation status and land use according to the Australian Land Use and Management Classification (ALUMC), Version 4 (http://www.affa.gov.au).

Search Words
AGRICULTURE
AGRICULTURE Crops
AGRICULTURE Horticulture
AGRICULTURE Irrigation
BOUNDARIES
BOUNDARIES Administrative
BOUNDARIES Biophysical
BOUNDARIES Cultural
FLORA
FLORA Exotic
FLORA Native
FORESTS
FORESTS Agroforestry
FORESTS Natural
FORESTS Plantation
HUMAN ENVIRONMENT
LAND
LAND Conservation
LAND Conservation Reserve
LAND Cover
LAND Ownership
LAND Use
VEGETATION
VEGETATION Structural

WATER
WATER Lakes
WATER Surface
WATER Wetlands

North Bounding Latitude
-24.485

South Bounding Latitude
-37.785

East Bounding Longitude
152.585

West Bounding Longitude
138.475

Beginning Date
1993-04

Ending Date
2003-08

Progress
Complete

Maintenance and Update Frequency
As required

Stored Data Set Format
DIGITAL ARC/INFO 8.2 under SunOS

Available Format Type
DIGITAL - ARC/INFO raster

Access Constraint
Access is unrestricted. There are two conditions of use:

1. Users of the data should acknowledge the following in any visual or published material: the data set was derived and compiled by the Bureau of Rural Sciences and land uses were derived using the Collaborative Australian Protected Areas Database 2000 (Dept of Environment and Heritage), TOPO-250K Version 1 (Geoscience Australia, Division of National Mapping), Australian Tenure and Forests of Australia 2003 (Bureau of Rural Sciences, National Forest Inventory), Normalised Difference Vegetation Index data (Dept of Environment and Heritage), agricultural census and survey data collected in 1994, 1997, 1999 and 2001 (Australian Bureau of Statistics) and control site data (compiled for the National Land and Water Resources Audit by NSW Agriculture, Victorian Dept of Natural Resources and Environment, Queensland Dept of Natural Resources and Mines, Primary Industries and Resources SA, Agriculture Western Australia, Tasmanian Dept of Primary Industries, Water and Environment and Northern Territory Dept of Lands, Planning and Environment).

2. Any errors, omissions or suggestions for improvement should be made known directly to BRS (by e-mail to dataman@brs.gov.au or by mail to the Data Manager, Bureau of Rural Sciences).

Lineage

I. The following existing digital maps were overlaid to determine the non-agricultural land uses and the distribution of agricultural land: 1. TOPO-250K (Version 1), 1:250,000 scale vector topographic data, published by Geoscience Australia, Division of National Mapping, February 1999 update. Line and point features were buffered prior to conversion to raster format.

2. Collaborative Australian Protected Areas Database (CAPAD 2000), 1:250,000 scale vector protected areas data, published by the Department of Environment and Heritage.

3. Australian Tenure, 250m raster tenure data, compiled by the Bureau of Rural Sciences, National Forest Inventory, in 1997. Information compiled by state and territory agencies in 1997 was used to classify aboriginal freehold and aboriginal leasehold land as agricultural or non-agricultural.

4. Forests of Australia 2003, 250m raster native and plantation forest data, compiled by the Bureau of Rural Sciences, National Forest Inventory, in 2003.

II. The spatial distribution of specific agricultural land uses for each of the

four years was determined using SPREAD II, a modified version of the SPREAD (SPatial REallocation of Aggregated Data) algorithm of Walker and Mallawaarachchi (1998). The method requires 3 inputs relating to a particular time period. These are a time sequence of NDVI images, a set of control sites (known location and agricultural land use) and agricultural census or survey data (reported on small regions and giving the area devoted to each agricultural land use). A computer program embodying an adaptation of SPREAD II was implemented by the Bureau of Rural Sciences. NDVI images were obtained from Advanced Very High Resolution Radiometer (AVHRR) data processed to correct for cloud cover by ERIN, Department of Environment and Heritage. Control site data were collected by State and Territory agencies. The irrigation status of most control sites is known and the method was used to determine the distribution, not only of commodity groups, but also of their irrigation status. Agricultural census and survey data reported on Statistical Local Areas (SLAs) were obtained from the Australian Bureau of Statistics. Modifications made to the agricultural census and survey data are documented in the Final Project Report. The SPREAD II methodology is statistically based, using a Bayesian technique - a Markov chain Monte Carlo (MCMC) algorithm. An irrigation area boundaries data set supplied by the Murray-Darling Basin Commission showing designated irrigation areas in the southern Murray-Darling Basin and actual irrigation areas in the northern Murray-Darling Basin was used to refine the prior probabilities used in the MCMC algorithm. The irrigation constraint was set so that 100% of irrigated land uses would fall within the designated irrigation areas (to the extent that the area inside the irrigation areas was sufficient to accommodate them). For each of the four years, SPREAD II generated outputs comprising the 42 probability maps described in the abstract and a summary agricultural land use map (which constitutes the agricultural component of the summary map described in the abstract). III. Land uses were assigned to pixels in the summary grids by using a macro to construct lookup tables which were then permanently joined to the value attribute tables of the summary grids. Non-agricultural land uses were assigned according to the attributes of the four layers overlaid in step I. Agricultural land uses were assigned according to the attributes of the summary agricultural land use map produced in step II. The land use classification used is the Australian Land Use Management Classification V4 (http://www.affa.gov.au).

Positional Accuracy

The data type and stated positional accuracy of the major existing data sets used to determine the non-agricultural land uses and the distribution of agricultural land (as discussed in the lineage section) are as follows: . CAPAD 2000 - vector data, spatial errors are in the range 1m to 500m . TOPO-250K (Version 1) - vector data, error less than 160m for at least 90% of well-defined points . Australian Tenure - 250m raster data, spatial errors, in the main, do not

exceed 125m . Forests of Australia 2003 - 250m raster data; source data has variable pixel

size ranging up to 500m The input NDVI imagery and the output probability and summary grids have 0.01 degree pixel size. Therefore, the positional accuracy of the outputs is

approximately 1 - 2 km.

Attribute Accuracy

Non-agricultural land uses were assigned, initially, on the basis of existing data sets showing protected areas, tenure, forest type and topographic features. Specific agricultural land uses were then assigned by automated interpretation of NDVI images. Accuracy of assignments based on existing data sets depends mainly on the attribute accuracy of the underlying data sets but also on the validity of the rules used for land use assignment. The attribute accuracy of the underlying data sets has not been tested except for the topographic features data set (TOPO-250K, Version 1 of Geoscience Australia) for which the range of allowable attribute errors is from 0.5% to 5% at a 99% confidence level. However, the attribute accuracy of the other three underlying data sets is expected to be high, with consequent high accuracy in non-agricultural land use assignments. The accuracy of the specific agricultural land use allocations based on automated interpretation of NDVI images is variable. The probability grids give an indication of the accuracy of the agricultural land use allocations. The final report shows how attribute accuracy varies as a function of probability for certain agricultural land uses.

Logical Consistency

The attribute combination corresponding to each land use assignment in the summary grid was tested by inspection to verify that these automated assignments were as intended and were logically consistent.

Completeness

Coverage and classification are complete. Verification of spatial and attribute data is discussed in the final report. In brief, the grids constituting the 2000 map were compared against catchment scale land use mapping (CLUM) data currently being compiled by the Bureau of Rural Sciences (Bureau of Rural Sciences, 2002). This analysis was confined to 17 SLAs. The relationship between attribute accuracy and probability was established for various agricultural land uses and the extent of agreement on a pixel to pixel basis was also assessed. Difficulties with this approach include the fact that the currency of the CLUM data is variable, the fact that the the CLUM land use categories do not align well with the commodity groups mapped by SPREAD II and the fact that the level of attribute detail in the CLUM data is variable.

Analysis of spatial and attribute accuracy was supplemented using Landsat imagery for the year 2000.

Contact Organisation

Bureau of Rural Sciences

Contact Position

Data Manager

Mail Address

GPO Box 858

Locality

CANBERRA

State

ACT

Country

Australia

Postcode

2601

Telephone

612 6272 4000

Facsimile

612 6272 4687

Electronic Mail Address

dataman@brs.gov.au

Metadata Date
2004-06-11

Additional Metadata

Bureau of Rural Sciences, 2002, Land use mapping at catchment scale: principles, procedures and definitions, Edition 2, Bureau of Rural Sciences, Canberra.

Bureau of Rural Sciences, 2004, Caveats: 1993, 1996, 1998 and 2000 Land Use of the Murray-Darling Basin, Version 2.

Bureau of Rural Sciences, 2004, Final Report: 1993, 1996, 1998 and 2000 Land Use of the Murray-Darling Basin, Version 2.

Bureau of Rural Sciences, 2004, User Guide: 1993, 1996, 1998 and 2000 Land Use of the Murray-Darling Basin, Version 2.

Walker, P.A. & Mallawaarachchi, T. 1998, 'Disaggregating agricultural statistics using NOAA-AVHRR NDVI', Remote Sens. Environ., vol. 63, pp. 112-125.

Appendix 5 User Guide for the 1993, 1996, 1998 and 2000 Land Use of the Murray-Darling Basin, Version 2

Introduction

The 1993, 1996, 1998 and 2000 Land Use of the Murray-Darling Basin, Version 2, is the second and final version of a first attempt at construction of a time series of land use maps of the Murray-Darling Basin for the years 1993, 1996, 1998 and 2000 using the new SPREAD II algorithm, a derivative of the SPREAD algorithm of Walker and Mallawaarachchi (1998). The maps are supplied as a set of ARC/INFO grids with geographical coordinates referred to WGS84 and 0.01 degree cell size. For each year there is a set of probability maps, one for each agricultural land use, and a single summary map made from the probability maps using some simple rules to make an approximation to a maximum likelihood land use map. As supplied the probability maps are floating point grids with cell value between 0 and 1 and no value attribute table while the summary map is an integer grid with a value attribute table with attributes defining the agricultural commodity group, irrigation status and land use according to the Australian Land Use and Management Classification (ALUMC), Version 4. The ALUMC Version 4 is the land use classification used in the 1996/97 Land Use of Australia, Version 2 data set of the National Land and Water Resources Audit (the Audit). The ALUMC Version 4 can readily be converted to the current ALUMC Version 5.

Methodology

The 1993, 1996, 1998 and 2000 Land Use of the Murray-Darling Basin, Version 2 was constructed using a derivative of the methodology described by Stewart et al, (2001) used to construct the 1996/97 Land Use of Australia, Version 2, a data set of the National Land and Water Resources Audit (the Audit).

The major difference between the current methodology and that used to construct the Audit map was the use of the new SPREAD II algorithm in place of the original SPREAD algorithm. Like SPREAD, SPREAD II is an automated method that can be used to disaggregate the Australian Bureau of Statistics' agricultural census or agricultural survey data to determine the spatial distribution of the agricultural land uses reported for each statistical local area (SLA). Both methods use a time series of normalized difference vegetation index (NDVI) images derived from AVHRR images to characterize both agricultural land pixels with unknown land use and control site pixels with known land use. They then relate the unknown pixels to the known pixels based on similarities in the NDVI time series. The SPREAD methodology uses a nearest neighbour approach and generates a single interpretive map and some qualitative reliability data. The SPREAD II methodology, however, is statistically based, using a Bayesian technique – a Markov chain Monte Carlo (MCMC) algorithm – and generates outputs of the kind described above. There are 42 probability maps comprising a dryland and an irrigated probability map for each of the 21 land uses mapped in the Audit map. For a given cell, the sum of the cell values for the 42 probability maps is 1. The summary grid was made from the probability grids using the following algorithm applied to each SLA in turn:

1. Allocate land use of rarest commodity to the cells with highest probability for the commodity until the AgStats constraint is satisfied.
2. Allocate land use of next rarest commodity to the remaining cells with highest probability for the commodity until the AgStats constraint is satisfied.
3. Continue until all land uses allocated.

The area allocated is never greater than the constraint, but note that the AgStats data are manipulated during processing to fit the area using the methodology described in the report on the Audit map (Stewart et al, 2001). As in the original SPREAD algorithm, a land use with less than 100 ha in a given SLA is treated as though the area were zero – the probability surface for that land use is set to zero for all agricultural cells in the SLA and, hence, there is no allocation in the SLA to that land use in the summary grid.

Another major difference between the current methodology and that used to construct the Audit map is that an irrigation constraint was used. An irrigation area boundaries data set supplied by the Murray-Darling Basin Commission showing designated irrigation areas in the southern Murray-Darling Basin and actual irrigation areas in the northern Murray-Darling Basin was used to refine the prior probabilities used in the MCMC algorithm. For the Version 2 maps, the irrigation constraint was set so that 100% of irrigated land uses would fall within the designated irrigation areas (to the extent that the area inside the irrigation areas was sufficient to accommodate them).

Minor differences between the current methodology and that used to construct the Audit map concerned the selection of input data and are detailed in the project report.

Grid naming conventions and data dictionary

The grids for the four maps are supplied on a CD, the grids for each map being stored in a separate directory named *gridsYY* where *YY* indicates the map year and is one of 93, 96, 98 or 00.

The probability grids are floating point grids. They have been named *prYY_NA* where:

- *YY* indicates the map year and is one of 93, 96, 98 or 00;
- *N* is a one or two digit integer code for the modelled land use with values ranging from 1 to 21; and
- *A* is a one letter code for the irrigation status of the mapped land use with values being either d (dryland) or i (irrigated).

The meanings of the land use codes can be read from the *spread_desc* column of Table 2, reading from the row of the table that has the appropriate land use code in the *spread* column of the table. The probability grids have cell values that are either equal to 0, for residual potentially agricultural land pixels in SLAs where the total of the area constraints is less than the area of potentially agricultural land, or to a number between 0 and 1, which is the probability that the land use for the pixel concerned was the mapped land use.

The summary grids are integer grids called *mdluYYn2* where *YY* indicates the map year and is one of 93, 96, 98 or 00. (The final digit in the name of the summary grids is the version number.) The summary grids all have the same structure. Each summary grid has a value attribute table (VAT) and comprises three layers, each layer being defined by a group of attributes in the VAT. The summary grids all have their VAT attributes defined and named the same way, as set out in the following tables 1, 2, 3 and 4. Table 1 lists the summary grid VAT attributes and shows how they define the layers.

Table 1. VAT attributes of the summary grid showing their meanings and how they define the three layers.

Attribute	Meaning	Layer
value	Cell value	Not applicable
count	Number of cells with given value	"
spread	Agricultural commodity code: SPREAD output	Agricultural commodities layer
spread_desc	Agricultural commodity description: SPREAD output	"
irrigation	Irrigation status code: SPREAD output	Irrigation layer
irrigation_desc	Irrigation status description: SPREAD output	"
lu_code	Land use code: ALUMC Version 4	Land use layer
lu_desc	Land use description: ALUMC Version 4, primary land use	"
lu_desc2	Land use description: ALUMC Version 4, secondary land use	"
lu_desc3	Land use description: ALUMC Version 4, tertiary land use	"
t-code	Land use code: ALUMC Version 4, tertiary code	"

The agricultural commodities layer is defined by the attributes *spread* (a numerical code) and *spread_desc* (a brief description). The values of these attributes and their meanings are listed in Table 2.

Table 2. Attributes of the agricultural commodities layer showing values and meanings.

spread	spread_desc	Meaning
-1	Non-agricultural land or no data	Non-agricultural land or no data.
0	Unallocated potentially ag. land	Potentially agricultural land for which no agricultural land use was allocated by SPREAD. The total area submitted to SPREAD exceeds the total commodity area available, for the SLA concerned. The land is non-forested and non-public. It is probably mainly non-agricultural. Intensive uses may be prominent, especially rural residential ('hobby farms') in periurban areas.
1	Residual/Native pastures	Native pasture of variable quality.
2	Agroforestry	Agroforestry
3	Sown pastures	Sown pastures
4	Cereals excluding rice	Cereals excluding rice (eg wheat, oats, barley, grain sorghum, maize, millet)
5	Rice	Rice
6	Legumes	Legumes (eg soybeans, peanuts, lupins)
7	Oilseeds	Oilseeds (eg canola, sunflower)
8	Sugar cane	Sugar cane
9	Non-cereal forage crops	Non-cereal forage crops
10	Cotton	Cotton
11	Other non-cereal crops	Other non-cereal crops (eg tea, coffee, turf, herbs)
12	Other vegetables	Other vegetables
13	Potatoes	Potatoes
14	Citrus fruit	Citrus fruit (eg oranges, lemons)
15	Apples	Apples

16	Pears	Pears (includes quinces and nashi)
17	Stone fruit	Stone fruit (eg apricots, figs, olives, peaches, avocados)
18	Nuts	Nuts (eg macadamia, almonds)
20	Plantation fruit	Plantation fruit (eg bananas, kiwifruit, pineapples)
21	Grapes	Grapes

The irrigation layer is defined by the attributes *irrigation* (a numerical code) and *irrigation_desc* (a brief description). The values of these attributes and their meanings are listed in Table 3.

Table 3. Attributes of the irrigation layer showing values and meanings.

irrigation	irrigation_desc	Meaning
-999	Non-ag or unalloc pot ag land or no data	Non-agricultural land or no data; unallocated potentially agricultural land.
0	Dryland agriculture	Dryland agriculture
1	Irrigated agriculture	Irrigated agriculture

The land use layer is defined by the attributes *lu_code* (a numerical code), *lu_desc* (the primary classification), *lu_desc2* (the secondary classification), *lu_desc3* (the tertiary classification) and *t-code* (the tertiary code). The tertiary code is a string of three numbers separated by periods indicating, respectively, the primary, secondary and tertiary classifications. These attributes, and their values, use the ALUMC, Version 4, described in the next section. The values of *lu_code* are three digit integers. The three digits indicate the primary, secondary and tertiary classes in the ALUMC. The three digits are the same as the three numbers forming the t-code, and are in the same order. For example, *lu_code* = 500 indicates primary class 5 Intensive uses (t-code 5.0.0), *lu_code* = 540 indicates secondary class 5.4 Residential (t-code 5.4.0) and *lu_code* = 542 indicates tertiary class 5.4.2 Rural residential (t-code 5.4.2). The values of *lu_desc* (the primary classification in words) and their meanings and corresponding ranges of values for *lu_code* are listed in Table 4.

Table 4. Values and meanings for the attributes, *lu_code* and *lu_desc,* of the land use layer.

lu_code	lu_desc	Meaning
0	NO DATA	No data.
100 to less than 200	CONSERVATION AND NATURAL ENVIRONMENTS	Land used primarily for conservation purposes, based on the maintenance of the essentially natural ecosystems present.
200 to less than 300	PRODUCTION FROM RELATIVELY NATURAL ENVIRONMENTS	Land used primarily for primary production based on limited change to the native vegetation.
300 to less than 400	PRODUCTION FROM DRYLAND AGRICULTURE AND PLANTATIONS	Land used mainly for primary production, based on dryland farming systems.
400 to less than 500	PRODUCTION FROM IRRIGATED AGRICULTURE AND PLANTATIONS	Land used mostly for primary production, based on irrigated farming.
500 to less than 600	INTENSIVE USES	Land subject to extensive modification, generally in association with closer residential settlement, commercial or industrial uses.
600 to less than 700	WATER	Water features. Water is regarded as an essential aspect of the classification, but it is primarily a cover type.

The values of *lu_desc2, lu_desc3* and *t-code* follow the ALUMC Version 4. See the following section for more information.

Australian Land Use and Management Classification (Version 4)

Minimum expected level of attribution

s-code	Secondary class	t-code	Tertiary class
I Conservation and Natural Environments			
1.1	Nature conservation[1]		
		1.1.1	Strict nature reserves
		1.1.2	Wilderness area
		1.1.3	National park
		1.1.4	Natural feature protection
		1.1.5	Habitat/species management area
		1.1.6	Protected landscape
		1.1.7	Other conserved area
1.2	Managed resource protection[1]		
		1.2.1	Biodiversity
		1.2.2	Surface water supply
		1.2.3	Groundwater
		1.2.4	Landscape
		1.2.5	Traditional indigenous uses
1.3	Other minimal use		
		1.3.1	Defence
		1.3.2	Stock route
		1.3.3	Remnant native cover
		1.3.4	Rehabilitation[2]
2 Production from Relatively Natural Environments			
2.1	Livestock grazing[3]		
2.2	Production forestry[4]		
3 Production from Dryland Agriculture and Plantations			
3.1	Plantation forestry		
		3.1.1	Hardwood plantation
		3.1.2	Softwood plantation
		3.1.3	Plantation forest nurseries
3.2	Farm forestry		
		3.2.1	Woodlots
		3.2.2	Windbreaks
		3.2.3	Tree and crop production
3.3	Grazing modified pastures[5]		
		3.3.1	Native/exotic pasture mosaic
		3.3.2	Woody fodder plants
		3.3.3	Legumes
		3.3.4	Legume/grass mixtures
		3.3.5	Sown grasses
3.4	Cropping[6]		
		3.4.1	Cereals
		3.4.2	Beverage & spice crops
		3.4.3	Hay & silage
		3.4.4	Oil seeds & oleaginous fruit
		3.4.5	Sugar
		3.4.6	Cotton
		3.4.7	Tobacco
3.5	Perennial horticulture[6]		
		3.5.1	Tree fruits
		3.5.2	Oleaginous fruits
		3.5.3	Tree nuts
		3.5.4	Vine fruits
		3.5.5	Shrub nuts fruits & berries
		3.5.6	Flowers & bulbs

		3.5.7	Vegetables & herbs
3.6	Seasonal horticulture[6]		
		3.6.1	Fruits
		3.6.2	Nuts
		3.6.3	Flowers & bulbs
		3.6.4	Vegetables & herbs
4 Production from Irrigated Agriculture and Plantations			
4.1	Irrigated plantation forestry		
		4.1.1	Irrigated hardwood plantation
		4.1.2	Irrigated softwood plantation
		4.1.3	Irrigated plantation nurseries
4.2	Irrigated farm forestry		
		4.2.1	Irrigated woodlots
		4.2.2	Irrigated windbreaks
		4.2.3	Irrigated tree and crop production
4.3	Irrigated modified pastures		
		4.3.1	Irrigated woody fodder plants
		4.3.2	Irrigated pasture legumes
		4.3.3	Irrigated legume/grass mixtures
		4.3.4	Irrigated sown grasses
4.4	Irrigated cropping[6]		
		4.4.1	Irrigated cereals
		4.4.2	Irrigated beverage & spice crops
		4.4.3	Irrigated hay & silage
		4.4.4	Irrigated oil seeds & oleaginous fruit
		4.4.5	Irrigated sugar
		4.4.6	Irrigated cotton
		4.4.7	Irrigated tobacco
4.5	Irrigated perennial horticulture[6]		
		4.5.1	Irrigated tree fruits
		4.5.2	Irrigated oleaginous fruits
		4.5.3	Irrigated tree nuts
		4.5.4	Irrigated vine fruits
		4.5.5	Irrigated shrub nuts fruits & berries
		4.5.6	Irrigated flowers & bulbs
		4.5.7	Irrigated vegetables & herbs
4.6	Irrigated seasonal horticulture[6]		
		4.6.1	Irrigated fruits
		4.6.2	Irrigated nuts
		4.6.3	Irrigated flowers & bulbs
		4.6.4	Irrigated vegetables & herbs
5 Intensive Uses			
5.1	Intensive horticulture		
		5.1.1	Shadehouses
		5.1.2	Glasshouses
		5.1.3	Glasshouses (hydroponic)
5.2	Intensive animal production		
		5.2.1	Dairy
		5.2.2	Cattle
		5.2.3	Sheep
		5.2.4	Poultry
		5.2.5	Pigs
5.3	Manufacturing and industrial[7]		
5.4	Residential		
		5.4.1	Urban residential
		5.4.2	Rural residential
5.5	Services		
		5.5.1	Commercial services
		5.5.2	Public services
		5.5.3	Recreation and culture
		5.5.4	Defence facilities

			5.5.5	Research facilities
	5.6	Utilities		
			5.6.1	Electricity generation/transmission
			5.6.2	Gas treatment, storage and transmission
	5.7	Transport and communication		
			5.7.1	Airports/aerodromes
			5.7.2	Roads
			5.7.3	Railways
			5.7.4	Ports and water transport
			5.7.5	Navigation and communication
	5.8	Mining		
			5.8.1	Mines
			5.8.2	Quarries
			5.8.3	Tailings
	5.9	Waste treatment and disposal		
			5.9.1	Stormwater
			5.9.2	Landfill
			5.9.3	Solid garbage
			5.9.4	Incinerators
			5.9.5	Sewage
6 Water[8]				
	6.1	Lake		
			6.1.1	Lake - conservation
			6.1.2	Lake - production
			6.1.3	Lake - intensive use
	6.2	Reservoir		
			6.2.1	Water storage and treatment
			6.2.2	Reservoir - intensive use
			6.2.3	Evaporation basin
			6.2.4	Effluent pond
	6.3	River		
			6.3.1	River - conservation
			6.3.2	River - production
			6.3.3	River - intensive use
	6.4	Channel/aqueduct		
			6.4.1	Supply channel/aqueduct
			6.4.2	Drainage channel/aqueduct
	6.5	Marsh/wetland		
			6.5.1	Marsh/wetland - conservation
			6.5.2	Marsh/wetland - production
			6.5.3	Marsh/wetland - intensive use
	6.6	Estuary/coastal waters		
			6.6.1	Estuary/coastal waters - conservation
			6.6.2	Estuary/coastal waters - production
			6.6.3	Estuary/coastal waters - intensive use

Notes –

The definitions for each class are being updated from those contained in the draft land use classification (6 May 1999) used in mapping projects for the National Land and Water Resources Audit (Gippsland, Victoria and the Fitzroy Basin, Qld), and the Murray Darling Basin Commission's Landmark Project (St George, Qld, Cootamundra, NSW and Swan Reach, SA).

1 **Nature conservation** - Tertiary classes are based on the Collaborative Australian Protected Areas Database (CAPAD) classification except 1.1.7 'Other conserved area' and 1.2.5. 'Traditional indigenous uses'. Class 1.1.7 includes forms of nature conservation outside the IUCN definition such as heritage agreements, voluntary conservation, registered property agreements etc.

2 **Rehabilitation** – This tertiary class includes areas that are under active rehabilitation or are unused because of weed infestation, salinisation, scalding and similar hazards.

3 **Livestock grazing (natural environments) -** ***Optional*** Tertiary classes provide an opportunity to link the land use classification with NVIS native vegetation classes. The classification at this level could be attributed by a simplified structural formation classification of the dominant native vegetation (NVIS level II), as shown below, if these data are available.

- Forest
- Woodland
- Open woodland
- Shrubland
- Grassland
- Other

4 **Production forestry (relatively natural environments) -** ***Optional*** Tertiary classes provide an opportunity to link the land use classification with NFI forest types. The classification at this level could be attributed by NFI forest type, as shown below, if these data are available.

- Rainforest
- Tall eucalypt forest
- Medium eucalypt forest
- Callitris forest
- Acacia forest

5 **Grazing improved pastures** **-** ***Optional*** Tertiary classes for legume and grass pasture types can be fitted to the pasture attributes collected through the ABS Agricultural Census. Attribution to the classification at this level is likely to be facilitated when census geocoding is completed.

6 **Cropping/ Perennial Horticulture / Seasonal Horticulture** **–** ***Optional*** attribution of agricultural commodities and commodity classes at the Tertiary level of the classification is likely to be facilitated when the geocoding of the Agricultural Census is completed. States may wish to map the distribution of important crops such as sugar, cotton and rice.

7 **Manufacturing and Industrial–** ***Optional*** attribution at the tertiary level includes abattoirs and other agricultural processing activities. Agricultural production facilities such as feedlots, piggeries etc are included as Tertiary classes under 5.2 Intensive animal production.

8 **Water–** Water is regarded as an essential aspect of the classification, but its inclusion is complicated as it is normally classified as a land cover type.

At the secondary level the classification identifies water features, both natural and artificial.

At the tertiary level natural water features are classed according to levels of use. *'Conservation'* accounts for features associated with land uses included in 1. Conservation and Natural Environments. *'Production'* accounts for features associated with land uses included in 2. Production from Relatively Natural Environments, 3. Production from Dryland Agriculture and Plantations, and 4. Production from Irrigated Agriculture and Plantations. *'Intensive use'* accounts for features associated with land uses included in 5. Intensive Uses.

The classification of water features will be reviewed after the next round of land use mapping.

References

Stewart, J.B., Smart, R.V., Barry, S.C. and Veitch S.M. 2001, '1996/97 Land Use of Australia: final report for project BRR5', National Land and Water Resources Audit, Canberra, available online http://audit.ea.gov.au/ANRA/atlas_home.cfm

Walker, P.A. and Mallawaarachchi, T. 1998, 'Disaggregating agricultural statistics using NOAA-AVHRR NDVI', *Remote Sens. Environ.,* vol. 63, pp. 112-125.

Bureau of Rural Sciences

June 2004

Appendix 6 AgStats Agregation

This appendix shows how 1996 – 97 AgStats items were aggregated to construct the mapped commodity groups. AgStats items for other years were aggregated similarly. Table 1 shows how the AgStats items were aggregated to the level of the ABS level 3 classification. Table 2 shows how the ABS level 3 items were aggregated to the level of the Audit commodity classification. The Audit commodity classification defines the mapped commodity groups. Table 2 also shows how the Audit commodity items aggregate to the level of the ABS level 1 classification though this latter classification was not used.

Table 1. 1996 – 97 AgStats items extracted to construct commodity groups.

AgStats 96/97		**ABS level 3 classification**	
Commodity item	**Code**	**Commodity group**	**Code**
Trees and shrubs planted			
Seedlings - area sown	704011	Agroforestry	2
Seed sown – area	704311	Agroforestry	2
Pastures			
Lucerne and other species - area at 31 March 1997	1000401	Other sown pastures	4
Lucerne (pure) - area at 31 March 1997	1000501	Pure lucerne	3
Pasture legumes (excl lucerne) - area at 31 March 1997	1001201	Other sown pastures	4
Sown grasses - area at 31 March 1997	1001301	Other sown pastures	4
Native or naturalised pasture at 31 March 1997	1001701	Native pasture	1
Mix of perennial grasses & legumes - area at 31 March 1997	1002101	Other sown pastures	4
Mix of annual grasses & legumes - area at 31 March 1997	1002201	Other sown pastures	4
Cereals			
Wheat for grain – area	1500101	Wheat	6
Oats for grain – area	1500801	Oats	7
Barley for grain – area	1501701	Barley	8
Cereal Rye for grain - area	1502301	Cereal rye	9
Buckwheat for grain - area	1503001	Buckwheat	10
Grain sorghum (for grain) - area	1504101	Grain sorghum	11
Maize for grain – area	1505301	Maize	12
Millet & panicum (inc. canary seed) - area	1507601	Millet	13
Rice for grain – area	1508501	Rice	17
Triticale for grain - area	1508801	Triticale	14

Other cereals for grain (NEC) - area	1510601	Other cereals for grain	15
Cereals (incl. forage sorghum) cut for hay – area	1510801	Cereals for hay/silage	16
Cereals fed off or for silage - area	1511001	Cereals for hay/silage	16

AgStats 96/97		ABS level 3 classification	
Commodity item	**Code**	**Commodity group**	**Code**
Non-cereal crops			
Aloe vera - area	1800301	Aloe vera	43
Mung beans - area	1800601	Mung beans	20
Other field beans - area	1800901	Other field beans	21
Soybeans - area	1801701	Soybeans	18
Broom millet (fibre) - area	1802201	Broom millet	44
Cotton - irrigated - area	1803001	Cotton	38
Cotton - non irrigated - area	1803101	Cotton	38
Hops - total area	1805101	Hops	39
Lab Lab Purpureus (seed) - area	1806001	Lab lab purpureus	45
Fennel (bitter) - (Foeniculum Vulgae) - area	1806401	Fennel	46
Lavender - area	1806501	Lavender	47
Linseed / Linola (clean seed) - area	1806701	Linseed	31
Lupins for grain - area	1807001	Lupins	22
Oil poppies - area	1807701	Oil poppies	32
Peanuts - area	1808101	Peanuts	19
Field peas for grain - area	1809101	Field peas	23
Pigeon peas - area	1900201	Pigeon peas	24
Chick peas - area	1900301	Chick peas	25
Coffee - area	1900401	Coffee	40
Popcorn - area	1900501	Popcorn	48
Canola - area	1900901	Canola	29
Safflower - area	1901401	Safflower	33
Sesame - area	1901501	Sesame	34
Crops fed off or cut for green feed or silage	1901911	Non-cereal crops for silage/green feed	36
Sugar cane cut for crushing - area	1902101	Sugar cane	35
Sugar cane cut for plants - area	1902301	Sugar cane	35
Sugar cane (standover) - area	1902501	Sugar cane	35
Sugar cane (newly planted) - area	1902601	Sugar cane	35
Sunflower - area	1903901	Sunflower	30
Tea - area	1904101	Tea	41
Tobacco - area	1904401	Tobacco	49
Vetches for seed - area	1904801	Vetches	26
Lentils - area	1905201	Lentils	27

Coriander - area	1907201	Coriander	50
Ginger - area	1907401	Ginger	51
Faba beans (incl tick and horse) - area	1907601	Faba beans	28
Mustard seed - area	1908101	Mustard	52
Cultivated turf - area	1908601	Turf	53
Pyrethrum - area	1908801	Pyrethrum	54
Peppermint - area	1908901	Peppermint	55
Crops (excl cereals) cut for hay - area	1909301	Non-cereal crops for hay	37
Non cereal crops (Not Elsewhere Classified)	1910301	Other non-cereal crops	56

AgStats 96/97		**ABS level 3 classification**	
Commodity item	**Code**	**Commodity group**	**Code**
Non-cereal crops (cont'd)			
Nurseries - area	1918201	Nurseries/flowers	42
Cut flowers - area	1918401	Nurseries/flowers	42
Vegetables			
Beans french & runner - for seed - area	3400501	Other vegetables	57
Carrot seed - area	3401001	Other vegetables	57
Cabbage (chinese) for seed - area	3401501	Other vegetables	57
Cauliflower seed - area	3401801	Other vegetables	57
Onion seed - area	3402001	Other vegetables	57
Peas green - for seed - area	3402501	Other vegetables	57
Potatoes - for seed - area	3403101	Potatoes	58
Pumpkins - for seed - area	3403501	Other vegetables	57
Radish seed - area	3404001	Other vegetables	57
Vegetable seed - other (not elsewhere classified) - area	3409901	Other vegetables	57
Potatoes - early/spring - harvest before 31 March - area	3503101	Potatoes	58
Potatoes (main/autumn) harvested after 31 March - area	3503601	Potatoes	58
Artichokes - area	3600201	Other vegetables	57
Asparagus - total area	3600601	Other vegetables	57
Broad beans - area	3601001	Other vegetables	57
French and runner beans (processing) - area	3601101	Other vegetables	57
French and runner beans (fresh market) - area	3601201	Other vegetables	57
Beetroot - area	3601501	Other vegetables	57
Broccoli - area	3601701	Other vegetables	57
Brussel sprouts - area	3601801	Other vegetables	57

Cabbages - area	3601901	Other vegetables	57
Chinese cabbage (Buckchoi and Wombak) - area	3602001	Other vegetables	57
Capsicum chillies and peppers - area	3602101	Other vegetables	57
Carrots - area	3602401	Other vegetables	57
Cauliflower - area	3602701	Other vegetables	57
Celery - area	3602901	Other vegetables	57
Chokos - area	3603001	Other vegetables	57
Cucumbers - area	3603201	Other vegetables	57
Eggplant - area	3603401	Other vegetables	57
Witloof chicory (french endive) - area	3603601	Other vegetables	57
Fennel (sweet) - area	3603701	Other vegetables	57
Garlic - area	3603901	Other vegetables	57
Gherkins - area	3604101	Other vegetables	57
Herbs - lemon grass etc - area	3604201	Other vegetables	57
Horse radish - area	3604301	Other vegetables	57
Leeks - area	3604401	Other vegetables	57
Lettuce - area	3604501	Other vegetables	57

AgStats 96/97		ABS level 3 classification	
Commodity item	**Code**	**Commodity group**	**Code**
Vegetables (cont'd)			
Marrows and squashes - area	3604701	Other vegetables	57
Zucchini - area	3604801	Other vegetables	57
Melons rock (incl cantaloupe) - area	3605101	Other vegetables	57
Melons water - area	3605201	Other vegetables	57
Melons (not elsewhere classified) - area	3605301	Other vegetables	57
Melons bitter (gourd) - area	3605401	Other vegetables	57
Mushrooms - area	3605801	Other vegetables	57
Okra - area	3605901	Other vegetables	57
Onions spring (incl shallots) - area	3606001	Other vegetables	57
Onions white and brown - area	3606101	Other vegetables	57
Parsley - area	3606301	Other vegetables	57
Parsnips - area	3606401	Other vegetables	57
Peas green (for processing) - area	3606601	Other vegetables	57
Peas green (for fresh market) - area	3606701	Other vegetables	57
Peas snow - area	3606901	Other vegetables	57
Pumpkins triambles trombones etc. - area	3607101	Other vegetables	57

Radish - area	3607201	Other vegetables	57
Rhubarb - area	3607401	Other vegetables	57
Silver beet and spinach - area	3607801	Other vegetables	57
Sprouts (alfalfa mung bean etc.) - area	3607901	Other vegetables	57
Sweet corn - area	3608001	Other vegetables	57
Sweet potatoes - area	3608101	Other vegetables	57
Tomatoes (processing) - area	3608401	Other vegetables	57
Tomatoes (fresh market) - area	3608501	Other vegetables	57
Swedes - area	3608901	Other vegetables	57
Turnips (white) - area	3609101	Other vegetables	57
Vegetables other (not elsewhere classified) - area	3609701	Other vegetables	57
Orchards			
Oranges trees under 6 years	4200711	Oranges	59
Oranges trees 6 years and over	4200712	Oranges	59
Grapefruit trees under 6 years	4201001	Grapefruit	60
Grapefruit trees 6 years and over	4201002	Grapefruit	60
Lemon and lime trees under 6 years	4201201	Lemon/lime	61
Lemons and limes trees 6 years and over	4201202	Lemon/lime	61
Mandarins trees under 6 years	4201401	Mandarins	62
Mandarins trees 6 years and over	4201402	Mandarins	62
Tangelos trees under 6 years	4201501	Tangelos	63
Tangelos trees 6 years and over	4201502	Tangelos	63
Citrus fruit other (NEC) - trees under 6 years	4201701	Other citrus	64
Citrus fruit other (NEC) -trees 6 years and over	4201702	Other citrus	64
Apples trees under 6 years	4202211	Apples	65
Apples trees 6 years and over	4202212	Apples	65

AgStats 96/97		**ABS level 3 classification**	
Commodity item	**Code**	**Commodity group**	**Code**
Orchards (cont'd)			
Pears (excluding Nashi) trees under 6 years	4202811	Pears	66
Pears (excluding Nashi) trees 6 years and over	4202812	Pears	66
Quinces - trees under 6 years	4203001	Quinces	67
Quinces - trees 6 years and over	4203002	Quinces	67
Nashi trees under 6 years	4203111	Nashi	68
Nashi trees 6 years and over	4203112	Nashi	68
Pome fruit NEC) trees under 6 years	4203201	Other pome	69

Pome fruit (NEC) trees 6 years and over	4203202	Other pome	69
Apricots trees under 6 years	4203901	Apricots	70
Apricots trees 6 years and over	4203902	Apricots	70
Avocados trees under 6 years	4204101	Avocados	82
Avocados trees 6 years and over	4204102	Avocados	82
Carambola trees under 6 years	4204201	Carambola	83
Carambola trees 6 years and over	4204202	Carambola	83
Cherries trees under 6 years	4204301	Cherries	71
Cherries trees 6 years and over	4204302	Cherries	71
Custard apples trees under 6 years	4204501	Custard apples	84
Custard apples trees 6 years and over	4204502	Custard apples	84
Dates trees under 6 years	4204601	Dates	85
Dates trees 6 years and over	4204602	Dates	85
Jackfruit trees under 6 years	4204701	Jackfruit	86
Jackfruit trees 6 years and over	4204702	Jackfruit	86
Figs trees under 6 years	4204801	Figs	72
Figs trees 6 years and over	4204802	Figs	72
Guava trees under 6 years	4204901	Guava	87
Guava trees 6 years and over	4204902	Guava	87
Loquats - trees under 6 years	4205101	Loquats	88
Loquats - trees 6 years and over	4205102	Loquats	88
Lychees - trees under 6 years	4205201	Lychees	89
Lychees - trees 6 years and over	4205202	Lychees	89
Mangoes trees under 6 years	4205301	Mangoes	90
Mangoes trees 6 years and over	4205302	Mangoes	90
Nectarines trees under 6 years	4205501	Nectarines	73
Nectarines trees 6 years and over	4205502	Nectarines	73
Olives - trees under 6 years	4205701	Olives	74
Olives - trees 6 years and over	4205702	Olives	74
Longans - trees under 6 years	4205801	Longans	92
Longans - trees 6 years and over	4205802	Longans	92
Peaches - trees under 6 years	4206211	Peaches	75
Peaches - 6 years and over	4206212	Peaches	75
Peacharines - trees under 6 years	4206301	Peacharines	76
Peacharines - trees 6 years and over	4206302	Peacharines	76

AgStats 96/97		ABS level 3 classification	
Commodity item	Code	Commodity group	Code
Orchards (cont'd)			
Persimmons - trees under 6 years	4206401	Persimmons	77
Persimmons - trees 6 years and over	4206402	Persimmons	77
Plums - trees under 6 years	4206601	Plums	78
Plums - trees 6 years and over	4206602	Plums	78
Prunes - trees under 6 years	4206701	Prunes	79
Prunes - trees 6 years and over	4206702	Prunes	79
Rambutan trees under 6 years	4206901	Rambutan	91
Rambutan trees 6 years and over	4206902	Rambutan	91
Stone fruit (NEC) trees under 6 years	4207701	Other stone fruit	80
Stone fruit (NEC) trees 6 years and over	4207702	Other stone fruit	80
Orchard fruit NEC - trees under 6 years	4207801	Other orchard fruit	81
Orchard fruit NEC - trees 6 years and over	4207802	Other orchard fruit	81
Almonds trees under 6 years	4208101	Almonds	94
Almonds trees 6 years and over	4208102	Almonds	94
Cashews trees under 6 years	4208301	Cashews	95
Cashews trees 6 years and over	4208302	Cashews	95
Chestnuts trees under 6 years	4208501	Chestnuts	96
Chestnuts trees 6 years and over	4208502	Chestnuts	96
Filberts hazelnuts & cobnuts trees under 6 years	4208701	Filberts	97
Filberts hazelnuts & cobnuts trees 6 years and over	4208702	Filberts	97
Macadamia trees under 6 years	4209001	Macadamia	93
Macadamia trees 6 years and over	4209002	Macadamia	93
Pecans trees under 6 years	4209201	Pecans	98
Pecans trees 6 years and over	4209202	Pecans	98
Pistachios trees under 6 years	4209301	Pistachios	99
Pistachios trees 6 years and over	4209302	Pistachios	99
Walnuts trees under 6 years	4209501	Walnuts	100
Walnuts trees 6 years and over	4209502	Walnuts	100
Nuts (NEC) trees under 6 years	4209701	Other nuts	101
Nuts (NEC) trees 6 years and over	4209702	Other nuts	101
Berry/Tropical fruit			
Black currants – area not yet bearing	4300401	Blackcurrants	102
Black currants - bearing area	4300402	Blackcurrants	102

Blueberries not yet bearing area	4300901	Blueberries	103
Blueberries bearing area	4300902	Blueberries	103
Gooseberries area not yet bearing	4301101	Gooseberries	104
Gooseberries bearing area	4301102	Gooseberries	104
Loganberries not yet bearing area	4301401	Loganberries	105
Loganberries bearing area	4301402	Loganberries	105
Raspberries - not yet bearing area	4301901	Raspberries	106
Raspberries - bearing area	4301902	Raspberries	106

AgStats 96/97		ABS level 3 classification	
Commodity item	Code	Commodity group	Code
Berry/Tropical fruit (cont'd)			
Strawberries - not yet bearing area	4302201	Strawberries	107
Strawberries - Bearing area	4302202	Strawberries	107
Berry and other small fruit (NEC) not yet bearing area	4302901	Other berry fruit	108
Berry and other small fruit (NEC) - bearing area	4302902	Other berry fruit	108
Babacos not yet bearing area	4304001	Babacos	110
Babacos bearing area	4304002	Babacos	110
Bananas - not yet bearing area	4304601	Bananas	109
Bananas - bearing area	4304602	Bananas	109
Kiwi fruit /zespri - not yet bearing area	4305001	Kiwifruit	111
Kiwi fruit / zespri - bearing area	4305002	Kiwifruit	111
Papaws / Papaya - not yet bearing area	4305401	Papaws	112
Papaws / Papaya - bearing area	4305402	Papaws	112
Passionfruit not yet bearing area	4305701	Passionfruit	113
Passionfruit bearing area	4305702	Passionfruit	113
Pepinos not yet bearing area	4305801	Pepinos	114
Pepinos bearing area	4305802	Pepinos	114
Pineapples - not yet bearing area	4306001	Pineapples	115
Pineapples - bearing area	4306002	Pineapples	115
Rosella not yet bearing area	4306301	Rosella	116
Rosella bearing area	4306302	Rosella	116
Tropical fruit (NEC) -not yet bearing area	4306601	Other tropical fruit	117
Tropical fruit (NEC) - bearing area	4306602	Other tropical fruit	117
Grapes			
Grapes - total area	4803011	Grapes	118

Table 2. Hierarchical classification of agricultural land uses used to classify SPREAD II inputs.

ABS level 1 code	ABS level 1 classification	Audit com-modity code	Audit commodity classification	ABS level 3 code	ABS level 3 classification
1	Residual/Native pastures	1	Residual/Native pastures	1	Residual
2	Agroforestry	2	Agroforestry	2	Agroforestry
3	Sown pastures	3	Sown pastures	3	Pure lucerne
				4	Other sown pastures
				5	Native pastures
4	Cereals	4	Cereals excluding rice	6	Wheat
				7	Oats
				8	Barley
				9	Cereal rye
				10	Buckwheat
				11	Grain sorghum
				12	Maize
				13	Millet
				14	Triticale
				15	Other cereals for grain
				16	Cereals for hay/silage
		5	Rice	17	Rice
5	Non-cereal crops	6	Legumes	18	Soybeans
				19	Peanuts
				20	Mung beans
				21	Other field beans
				22	Lupins
				23	Field peas
				24	Pigeon peas
				25	Chick peas
				26	Vetches
				27	Lentils
				28	Faba beans
		7	Oilseeds	29	Canola
				30	Sunflower
				31	Linseed

				32	Oil poppies
				33	Safflower
				34	Sesame
		8	Sugar cane	35	Sugar cane
		9	Non-cereal forage crops	36	Non-cereal crops for silage/green feed
				37	Non-cereal crops for hay
		10	Cotton	38	Cotton
		11	Other non-cereal crops	39	Hops
				40	Coffee
				41	Tea
				42	Nurseries/ Flowers
				43	Aloe vera
				44	Broom millet
				45	Lab lab purpureus
				46	Fennel
				47	Lavender
				48	Popcorn
				49	Tobacco
				50	Coriander
				51	Ginger
				52	Mustard
				53	Turf
				54	Pyrethrum
				55	Peppermint
				56	Other non-cereal crops
6	Vegetables	12	Other vegetables	57	Other vegetables
		13	Potatoes	58	Potatoes
7	Orchards	14	Citrus fruit	59	Oranges
				60	Grapefruit
				61	Lemon/lime
				62	Mandarins
				63	Tangelos
				64	Other citrus
		15	Apples	65	Apples
		16	Pears	66	Pears

				67	Quinces
				68	Nashi
				69	Other pome
		17	Stone fruit	70	Apricots
				71	Cherries
				72	Figs
				73	Nectarines
				74	Olives
				75	Peaches
				76	Peacharines
				77	Persimmons
				78	Plums
				79	Prunes
				80	Other stone fruit
				81	Other orchard fruit
				82	Avocados
				83	Carambola
				84	Custard apples
				85	Dates
				86	Jackfruit
				87	Guava
				88	Loquats
				89	Lychees
				90	Mangoes
				91	Rambutan
				92	Longans
		18	Nuts	93	Macadamia
				94	Almonds
				95	Cashews
				96	Chestnuts
				97	Filberts
				98	Pecans
				99	Pistachios
				100	Walnuts
				101	Other nuts
8	Berry/	19	Berry fruit	102	Blackcurrants

	Plantation fruit				
				103	Blueberries
				104	Gooseberries
				105	Loganberries
				106	Raspberries
				107	Strawberries
				108	Other berry fruit
		20	Plantation fruit	109	Bananas
				110	Babacos
				111	Kiwifruit
				112	Papaws
				113	Passionfruit
				114	Pepinos
				115	Pineapples
				116	Rosella
				117	Other tropical fruit
9	Grapes	21	Grapes	118	Grapes

Appendix 7 SPREAD II Statistics

1 Theory

In the following discussion we assume that the map is a lattice of pixels, and that all pixels have area one. In practice there is slight variation in the areas of the pixel due to the nature of the satellite data and we will discuss this in the next section.

We make the following definitions and associated notation. Let there be K landuses that we wish to assign. Assume that the SLA we are allocating has N pixels, and that the nth pixel has an actual landuse C_n that we wish to estimate. C_n can take the values $1, \ldots, K$ depending on the actual land use. The total area, ignoring variation in pixel size, for landuse k is defined as

$$T_k = \sum_{n=1}^{N} I(C_n = k),$$

where $I()$ is the indicator function. We define bold notation to represent the vector version of arguments over their natural dimension. For example $\mathbf{T} = (T_1, \ldots, T_K)$, ie over the index k.

We assume a prior structure for the map as follows. First we assume that the total area in the map for each commodity is drawn from some joint distribution defined such that $\sum_k T_k = N$. This function is

$$[\mathbf{T}|\theta],$$

where θ is a vector of parameters.

Given $\mathbf{T} = \mathbf{t}$, we assume that each pixel has a prior probability $\pi_n = (\pi_{1n}, \ldots, \pi_{Kn})$, where π_{kn} is the prior probability that the nth pixel is landuse k. The prior distribution of landuses over the map is assumed to be

$$[C|\mathbf{T}] \propto \prod_n \prod_{k=1}^{K} \pi_{kn}^{I(C_n=k)}, \tag{1}$$

which is defined over all C such that the allocated landuses equal the totals.

The normalising constant for this is therefore

$$\sum_{\text{all C permutations that satisfy the constraint}} \prod_n \prod_{k=1}^{K} \pi_{kn}^{I(C_n=k)}.$$

Combining these equations we have that the joint distribution for the map is

$$[C, \mathbf{T}, \theta] \propto [C|\mathbf{T}][\mathbf{T}|\theta][\theta],$$

where $[\theta]$ is the prior distribution supplied by the user.

The aim of the inference is to infer the C_n (and hence the T_k). We assume for each pixel we have M measurements $\mathbf{Y}_n = (Y_{1n}, \ldots, Y_{Mn})$, and that over the pixels we have conditional joint densities $[\mathbf{Y}_n|k] = [Y_{1n}, \ldots, Y_{Mn}|k]$, which define how the measured attributes vary with differing landuse k.

1

The posterior is therefore

$$[C, \mathbf{T}|\mathbf{Y}, \theta] \propto [C, \mathbf{T}, \theta] \prod_n [\mathbf{Y}_n|C_n].$$

The normalising constant in this case is intractable due to the number of permutations involved. Instead we explore this posterior via a Gibbs sampling algorithm.

2 Gibbs Sampler

Sampling from the posterior distribution of C can be performed if the full conditionals are easy to calculate. In this case, where the prior probabilities are constant, the calculations are simple. In this case, we have that the product term in the normalising constant is constant for all permutations. In this case

$$[C|\mathbf{T}] = \frac{N!}{T_1!T_2!\cdots T_k!}\frac{1}{K^n}.$$

To use the Gibbs sampler we need to calculate the conditionals for the C_n. We can ignore the T_k as they are degenerate if the C_n are known. The conditional is

$$[C_n|C_{-n}, \mathbf{T}|\mathbf{Y}, \theta],$$

where C_{-n} denotes all pixels except the nth. This distribution is a multinomial with

$$[C_n = i|C_{-n}, \mathbf{T}, \mathbf{Y}, \theta] \propto \frac{T_1!T_2!\cdots T_k!}{N!} K^n \frac{1}{K^n} [\mathbf{T}|\theta][\theta][\mathbf{Y}_n|C_n = i].$$

We note that $\sum_i [C_n = i|C_{-n}, \mathbf{T}, \mathbf{Y}, \theta] = 1$. To gain insight into the conditional and to ease calculation, we note that if we divide through by the current value of the $C_n = h$ we have

$$[C_n = i|C_{-n}, \mathbf{T}, \mathbf{Y}, \theta] \propto \frac{T_i + 1}{T_h} \frac{[\mathbf{T}|\theta]}{[\widetilde{\mathbf{T}}|\theta]} \frac{[\mathbf{Y}_n|C_n = i]}{[\mathbf{Y}_n|C_n = h]},$$

where $[\widetilde{\mathbf{T}}|\theta]$ is evaluated at the previous value of C_n.

In the maps we are producing, $[\mathbf{T}|\theta]$ has a Gaussian distribution with mean equal to the adjusted agricultural statistics and variance equivalent to a standard deviation of 200 hectares, so that the constraint is quite hard. The densities $[Y_n|C_n]$ are constructed from the control site data by using trivariate Gaussian kernels to produce kernel density estimates. The three statistics calculated are mean greeness, greeness range and time of maximum greeness. The Gibbs sampling algorithm was run for 10000 iterations with an additional burn in of 1000 iterations. For each pixel the proportion of times it was allocated to each landuse was calculated to give the probability surfaces. The final allocation map was produced by the following process:

- Sort pixels based on probability value for the rarest commodity.
- Allocate the most likely pixels to satisfy the agricultural statistics constraint for the rarest commodity.
- Sort the remaining pixels based on probability value for the next rarest commodity.
- Allocate the most likely pixels to satisfy the agricultural statistics constraint for the next rarest commodity.
- Continue till complete.

3 Imposing constraints

The simplest way of imposing a constraint is when it is absolute. In this case we simply set the density for the commodities to zero in the pixels that cannot have the particular landuse. When constraints are of the form "70% of activity should occur in this area" the approach is currently to define a pseudo density such that if there are two areas A and B, and a ratio p to be applied to some landuse or group of landuses we calculate a density:

	LU1	LU2
A	$p * A_{lu1}/A_A$	$1 - p * A_{lu1}/A_A$
B	$(A_{lu1} - p * A_{lu1})/A_B$	$1 - (A_{lu1} - p * A_{lu1})/A_B$

where A_i is the area of i. This density is then applied independently to produce the posterior. It is motivated by noting that if the imagery is uninformative this construction will return the correct expected allocation. The obvious adjustments are made if there is insufficient area to apply the constraint.

4 Improvements

A number of improvements are possible. First, a swapping sampler (bivariate Gibbs) may allow more efficient sampling of the posterior. If we are willing to fix the constraint there are a number of more powerful approaches that could be used in imposing the constraints. The efficiency occurs because the ratio of the normalising constant is tractable under this assumption.