DECLARATION

I certify that this thesis does not incorporate, without acknowledgement, any material previously submitted for a degree or diploma in any university and that, to the best of my knowledge, it does not contain any material previously published or written by another person except where due reference is made in the text. The work in this thesis is my own, except for the contributions made by others as described in the Acknowledgements.

Simon Knapp
ACKNOWLEDGMENTS

Much of the material in Sections 2.2 and 2.3 was based on work I was heavily involved with while working at the Bureau of Rural Sciences, and which was done in collaboration with colleagues. The specific sections are noted in the text. Of particular note, Simon Barry initially conceived of SPatial REallocation of Aggregated Data Version II (SPREAD II) and Robert Smart prepared and documented most of the data.

Professional copy editing was provided by Matthew Sidebotham of workwisewords. Finalisation of this thesis has been far easier since his preening.

Most of the data used in this thesis was generously provided by the Australian Department of Agriculture and Water Resources. My ex-colleagues there provided interesting insights on the work presented here and the peculiarities of land use and its analysis in Australia.

The support of my supervisors, and in particular Simon Barry, has been invaluable. His ability to focus on the facts and innovate has been inspirational.

Perhaps most importantly, the work undertaken in this thesis would not have been possible or, at least, would have been much harder, more time consuming, and of lower quality, had it not been for the assistance of Robert Smart. He prepared most of the data used herein and has always had time to chat about, and help resolve, both technical and practical issues.

Finally, I’d like to thank my family. Joanna’s patience and support through what has been a longer journey than anticipated, one that has taken place alongside starting a family and a small business, has been extraordinary. As for Alex and Dustin, a main motivator for knocking this thing over is to be able to spend my weekends with them and pay them my full mind!
We present a flexible, automated, Bayesian method designed for broad scale land use mapping. The method is based on a Monte Carlo Markov Chain and integrates a number of sources of ancillary data. It produces a probability density over a finite set of land use classes that can be used directly in further analyses or to classify individual pixels. The method assumes a multinomial prior over the possible land use types, and uses agricultural statistics to form stochastic constraints over the total area allocated to each land use within a region. A supervised learner is then used to allocate pixels within the region, while respecting the constraints. We then extend this method in three ways. First, supplementary mapping is used to form further constraints over subsets of the original land use classes. Second, two spatial models are considered; the first considers the use of partially labelled pixels, where the labels are based on the current state of the Markov Chain, and the second assumes a Markov Random Field. Third, the form of the prior is relaxed, and the method extended to allow the creation of a time series of maps using either cascade or compound classification techniques. The methods are benchmarked against the probabilistic classifier upon which they are built and simple Bayesian modifications to the raw classifier that incorporate the same data. The techniques are demonstrated and assessed using Australian data generated by a national Land Use (LU) mapping program and show promise in many of the test regions we consider.
CONTENTS

1. Introduction ................................................................. 8
   1.1 Background ............................................................ 9
       1.1.1 Land Cover and Land Use ...................................... 10
       1.1.2 The Use of Ancillary Data ..................................... 11
       1.1.3 Land Use Mapping in Australia .................................. 13
   1.2 Outline of This Thesis ............................................... 15

2. Spatial REallocation of Aggregated Data
   Version II ................................................................. 17
   2.1 Introduction ........................................................... 17
   2.2 Methods ............................................................... 18
       2.2.1 SPatial REallocation of Aggregated Data (SPREAD) ......... 18
       2.2.2 SPatial REallocation of Aggregated Data Version II (SPREAD II) ......... 19
       2.2.3 A Simple Bayesian Classifier ................................... 22
       Map of Australia (NLUM) ............................................ 22
       2.3.1 Data ............................................................... 23
       2.3.2 Running SPREAD II ................................................ 26
       2.3.3 Software ........................................................... 27
       2.3.4 Results ............................................................ 27
   2.4 SPREAD II Maps for 2005–06 and 2010–11 .......................... 35
       2.4.1 Data ............................................................... 35
       2.4.2 Software ........................................................... 38
       2.4.3 Results ............................................................ 40
   2.5 Simulation Study ..................................................... 46
   2.6 Discussion ............................................................ 50
3. Including Sub-regional Constraints in SPREAD II ........................................ 53
   3.1 Introduction .................................................................................. 53
      3.1.1 Possible Approaches ............................................................... 54
   3.2 Methods ....................................................................................... 55
      3.2.1 A Single Zonal Constraint ....................................................... 55
      3.2.2 Multiple Constraints ............................................................... 57
      3.2.3 A Simple Bayesian Classifier .................................................. 58
   3.3 Results ......................................................................................... 59
   3.4 Discussion .................................................................................... 61

4. Including Spatial Context in SPREAD II ...................................................... 64
   4.1 Introduction .................................................................................. 64
   4.2 Methods ....................................................................................... 65
      4.2.1 Markov Random Field Approach ........................................... 66
      4.2.2 The Transductive Approach ................................................... 67
   4.3 Results ......................................................................................... 68
      4.3.1 Neighbourhood Size ............................................................... 69
   4.4 Discussion .................................................................................... 71

5. Producing a Time Series of Maps Using SPREAD II .................................... 73
   5.1 Introduction .................................................................................. 73
   5.2 Methods ....................................................................................... 74
      5.2.1 Cascade Classification ............................................................. 78
      5.2.2 Compound Classification ....................................................... 79
      5.2.3 Transition Matrix .................................................................. 85
      5.2.4 A Simple Bayesian Classifier ................................................. 87
   5.3 Results ......................................................................................... 87
   5.4 Simulation Study .......................................................................... 91
   5.5 Discussion .................................................................................... 93

6. Summary and Discussion ............................................................................. 96
   6.1 Summary of Results ...................................................................... 96
   6.2 Discussion ..................................................................................... 101
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.3 Conclusions</td>
<td>102</td>
</tr>
<tr>
<td>6.4 Further Research</td>
<td>105</td>
</tr>
<tr>
<td>6.4.1 Applying the Techniques in Concert</td>
<td>105</td>
</tr>
<tr>
<td>6.4.2 Producing Maps for Non-Census Years</td>
<td>105</td>
</tr>
<tr>
<td>6.4.3 Inclusion of Textures or Objects</td>
<td>105</td>
</tr>
<tr>
<td>6.4.4 Incorporation of Multi-resolution Data</td>
<td>106</td>
</tr>
</tbody>
</table>

Bibliography .............................................. 107

Appendices .................................................. 115

A. Test Regions ............................................ 115
1. INTRODUCTION

Continental scale Land Use (LU) information is an important input into many areas of natural resource management. Examples of areas of current interest include the measurement of land based Green House Gas (GHG) fluxes induced by LU change (Department of Climate Change, 2009; Australian Greenhouse Office, 2002), agricultural productivity (Bryan and Marvanek, 2004; Marinoni et al., 2012), environmental sustainability (Gardi et al., 2010), food security (Morrison et al., 2011), and as inputs into biophysical models (Hurtt et al., 2001), health studies (Maxwell et al., 2004) and salinity modelling (Kiiveri et al., 2003). This broad applicability and the dynamic nature of the studies to which LU mapping is applicable has rendered it one of the most widely studied problems employing satellite data (Cihlar and Jansen, 2001).

This thesis considers methods for broad (continental) scale LU mapping using ancillary data in the form of agricultural statistics and supplementary mapping. The techniques are demonstrated and assessed using Australian data generated by a national LU mapping program. We will review the literature on satellite based LU mapping, focusing on studies that have incorporated ancillary data. We then outline the available LU and Land Cover (LC) data in Australia and describe the design and structure of this thesis. The methods developed here are extensions to the method currently used for mapping the agricultural components regions within the National Land Use Maps of Australia (NLUMs).
1. Introduction

1.1 Background

The use of remotely sensed data in mapping large areas began in the early 1980s with the National Oceanic and Atmospheric Administration (NOAA) series of satellites. This was the first generation of satellites with radiometric sensors advanced enough to be of use for LC mapping (Tucker et al., 1985; DeFries et al., 1995; Franklin and Wulder, 2002). One of the first scientific studies reported in the literature that used remotely sensed data for continental scale mapping is Tucker et al. (1985). This study used Advanced Very High Resolution Radiometer (AVHRR) Global-Area Coverage (GAC) Normalised Differenced Vegetation Index (NDVI) data to classify land cover and monitor vegetation dynamics for Africa over a 19-month period. Those authors note that prior to that study Landsat was the main source of remotely sensed data, but had not been used to map large regions due to its fine spatial scale and relatively low temporal collection frequency, which rendered it expensive to process and difficult to use in automated classification procedures. The high temporal frequency of the AVHRR data allows for the creation of regular cloud free mosaics and hence the use of classifiers based on temporal vegetation dynamics. Prior to that work, broad scale maps were derived from preexisting maps and atlases (Friedl et al., 2002).

Since the original study of Tucker et al. (1985), mapping of large areas has been studied extensively and reviews of the status of the science are published regularly (Loveland et al., 1991; Franklin and Wulder, 2002; Aplin, 2004). This has been driven largely by the emergence of global environmental issues such as climate change and biodiversity, which require LU information at various scales, and by the broad utility of these data to a wide range of scientific disciplines (Cihlar, 2000). More recently, technical advances in computational infrastructure, the advent of high performance and cloud computing systems, and advances in statistical and machine learning have opened up new research opportunities and increased interest.

The phenomenal increase in computing power and storage space over past decades and the range and quality of remotely sensed data available have progressively facilitated increases in the resolution of LU maps. While early broad scale studies were conducted using data with an approximate 4 or 16km (NDVI GAC or Global Vegetation Index (GVI)) resolution (Loveland et al., 1991), it is now feasible to produce broad scale maps with resolutions of 1km and below. For example, the National Land Cover Database (NLCD) covers the conterminous United States at a nominal resolution of 30m and is produced approximately every five years (Vogelmann et al., 2001; Homer et al.; Fry et al., 2011; Homer et al., 2015).

Developments in classification techniques used in LU and LC mapping closely follow developments in machine learning, statistics and pattern recognition. Some of the automated techniques that were being applied at various spatial scales at the time of the work by Tucker et al. (1985) included maximum likelihood (either Linear Discriminant Analysis (LDA) or Quadratic Discriminant Analysis (QDA)) (e.g. Townshend et al., 1987; DeFries et al., 1995), principal components (e.g. Tucker et al., 1985), classification trees (e.g. Lloyd, 1990) and Neural Networks (NNs) (e.g. Benediktsson et al., 1990). Since then, techniques such as boosting (e.g. Friedl et al., 1999; Chi and Bruzzone, 2005), Support Vector Machines (SVMs) (e.g. Bruzzone and Marconcini, 2009) and variations on those previously considered, which either approximate the likelihood function numerically or calculate classification boundaries which are less sensitive to the curse of dimensionality (Oommen et al., 2008; Hughes, 1968), have been trialled. Many of
these techniques are themselves computationally intensive and are only now becoming feasible for large scale studies.

Many of the more recent techniques are based on iterative and/or transductive techniques (Vapnik, 1998) which exploit unlabelled or partially labelled data, and allow the use of data from neighbouring time periods (Bruzzone and Marconcini, 2009), and small training samples of relatively poor quality (Bruzzone et al., 2006; Chi and Bruzzone, 2005). These techniques exploit the fact that in many thematic mapping applications, the data, and in particular the satellite imagery, are known for all points of interest a priori and that one is only concerned with labelling a fixed set of points as opposed to any point in the domain of the spatial imagery or some metrics thereon. This allows transductive techniques to exploit information in the observations to be classified, as well as that in the training data. These are important developments, as one common problem faced by LU and LC studies is the scarcity of good quality training data that are representative of the classes being mapped and can capture correlations between observations that are close spatially and/or temporally (Bruzzone et al., 2006).

1.1.1 Land Cover and Land Use

While LC can theoretically be defined precisely at a given point in space and time, the appropriate LU label to apply to a given LC will vary with the purpose for which a map is being used (Cihlar and Jansen, 2001) and appropriate taxonomies for both LC and LU must be defined with respect to a specific ontology or purpose of a map. For example, Matsuoka et al. (2007) produced a map primarily for use in hydrological modelling of the Yellow River in China; consequently the LC classification is focused on separating classes of primary importance to water flows. In contrast, and due largely to their expense, most broad scale mapping projects aim to produce classifications designed to be useful for a wide range of analyses, either directly or through reclassification appropriate to the goals of a specific project. Reclassification usually involves grouping of classes or disaggregation using ancillary data. These large scale studies also aim to create products enabling the comparison of LC and/or LU over broad regions or between regions, or consistent inputs to global climatological or biophysical models (Bartholomé and Belward, 2005; Cihlar and Jansen, 2001; Loveland et al., 1991).

The methods used in mapping LU and LC can differ. Since LU is more an anthropological issue (Cihlar and Jansen, 2001; Wästfelt, 2009), it is possible to map some LU classes in some classification schemes based on administrative boundaries, cadastres or other spatial datasets. This is the case for the production of the NLUMs, in which all non-agricultural LUs are derived directly from administrative boundaries. LC, on the other hand, can change in response to a wide range of climatological and environmental factors beyond the control of man and hence is generally determined through data analysis or surveys. LC is also much more sensitive to scale. For example, at a broad scale a LC might be commercial forest, but at a finer scale subregions within a commercial forest may be (perhaps in a different classification scheme) bare ground or pine forest.

Where LU cannot be mapped from administrative boundaries or the like, it is generally inferred from LC. Consequently, most broad scale mapping projects involve some element of LC mapping (Wästfelt, 2009) or make use of existing LC maps (Cihlar and Jansen, 2001; Hurtt et al., 2001) during the preprocessing or post-classification procedures, or as an integral part of the LU
classification method.

While the techniques explored in this thesis were developed for the classification of LU, they are equally applicable to LC. Further, some of the LU classes contained in the target classification schemes essentially map to a specific LC when considered over a single year (e.g. agroforestry, citrus, apples, grapes and nuts), while many others may encompass multiple LC classes (e.g. seasonal cropping and pastures in crop/pasture rotations). In the latter case, we would hope that the variety of LC combinations that occur within a given LU class are represented in the training data. In any case, due to their increased heterogeneity, it is likely that mapping such classes will be more difficult without relatively large amounts of training data that is geographically close to the region being mapped unless the LC sequences vary little spatially.

1.1.2 The Use of Ancillary Data

Over large regions, limitations in the availability of observational and training data become more pronounced, and climatological and/or other environmental conditions may increase heterogeneity of observed data within an individual LU or LC, hence reducing the accuracy of classifiers. Further, large scale mapping projects generally use relatively coarse scale remotely sensed imagery due to the expense, availability, low collection frequency and computational difficulties associated with obtaining and processing the large volumes of fine scale imagery that would otherwise be required. Individual pixels within coarse scale imagery often contain multiple LUs or LCs and the accuracy of classifiers is likely to be decreased relative to finer scale imagery.

Classification methods that only use remotely sensed imagery are generally not sufficient for discriminating between LU or LC classes due to both similarities in the reflectance spectra from various LUs or LCs (Bryan and Marvanek, 2004; Kiiveri et al., 2003; DeFries et al., 1998; Jewell, 1989). Technical issues with the collection of the data such as intertemporal and spatial climate variability, disturbance events (e.g. fires and land management), sensor conditions (e.g. view angles and sensor calibration drift) and signal contamination (e.g. atmospheric contamination and soil color) (Lu et al., 2003) also reduce the discrimination power of remotely sensed imagery. These issues are particularly problematic when there are a large number of classes in a classification scheme, as is case in the work undertaken in this thesis.

To deal with these issues, broad scale mapping projects tend to depend on highly sophisticated, automated algorithms that incorporate ancillary data in order to improve discrimination and reduce labour costs (Richards et al., 1982; Rogan et al., 2003). Techniques for including ancillary data range from expert opinion being applied post hoc (e.g. Loveland et al., 1991) through to fully automated techniques based on training data (e.g. Melgani and Serpico, 2002). Aplin (2004) and Franklin and Wulder (2002) review this research and provide references to particular studies. Sources of ancillary data that have been used in LU and LC mapping include other satellite imagery collected over the region of interest, information on landforms (e.g. elevation, slope and soil type information), expert knowledge, administrative data, official statistics, biophysical models, classifications at other time periods and other classifications at the same time period.
One important source of ancillary data used in LU and LC mapping generally, and central to the methods explored in this thesis, is Agricultural Statistics (AgStats), which give estimates of the total area within a region being used for each LU. These data provide direct estimates of quantities of interest, albeit at a generally broad spatial scale, which complements the relatively fine-scale contextual data provided by remotely sensed imagery (Bryan and Marvanek, 2004). Most methods reviewed in the course of this thesis incorporate these data through post hoc calibration procedures which lack theoretical foundations and/or make strong assumptions about the extent to which relationships between LUs hold between regions. Several methods that rely on AgStats and demonstrate the types of techniques used to incorporate them are described below, and Kuenmerle et al. (2013) provide a review of more recent studies in the context of LU intensity mapping.

Ramankutty and Foley (1998) reclassify the DISCover LC dataset (Loveland et al., 1999) into six classes that are each assumed to have a globally homogeneous fractional crop cover. They then systematically assign a fractional crop cover to each class; use linear regression to compare the resulting weighted sum (an area) to the area of cropping reported in various AgStats data within each political unit; and choose the set of fractional crop cover that yields the highest correlation coefficient such that the estimated slope is between 0.9 and 1.1.

Khan et al. (2010) use the ISODATA algorithm (Ball and Hall, 1965) to cluster pixels based on their temporal NDVI profiles, then use linear regression to relate AgStats, (the dependent variable) to the area of each cluster type, (the explanatory variables) within a region, one crop type at a time. For each crop type, the estimated coefficients for cluster type are interpreted as the fraction of a cluster of type that is being used for that crop type. Biggs et al. (2006) use the same approach for LU types irrigated using surface water, irrigated using groundwater and non-irrigated.

Hurtt et al. (2001) also use the DISCover dataset and estimate the composition of the LU classes defined in the AgStats in terms of the LC types of the DISCover. They use a multiple response regression of the form \( \mathbf{c} = \mathbf{Ar} \), where \( \mathbf{c} \) is a vector (of length 4) of areas of various LU types based on AgStats data, \( \mathbf{r} \) is a vector (of length 16) of areas of various LC types based on the LU classes from the DISCover data, and \( \mathbf{A} \) is a matrix (of dimension 4 × 16) of fractional covers.

Maxwell et al. (2004) use areal estimates of corn harvest to choose training pixels that are likely to be corn. After measuring the spectral distances between the training pixels and all other pixels, they label each pixel as highly likely to be corn, likely to be corn or unlikely to be corn, such that the combined area of the pixels labelled highly likely to be corn is 75 percent of the area of corn according the areal estimates, the area of likely to be corn is the remaining 25 percent, and the remaining pixels are labelled unlikely to be corn. These percentages were based on a sensitivity analysis performed on the three test counties through a trial and error process.

Frolking et al. (2002) use sown area of each crop and total crop area (the combined sown area of all crop types) estimated from AgStats data along with a set of Multicrop Rotation Priorities in an iterative procedure to determine the proportion of area within subregions that is under one of several crop rotation regimes. They combine this data with a fine scale LC database derived from Landsat TM data that includes categories of rice paddies and nonflooded cropland to produce a map of where rice farming occurs and under what rotation regime.
Other studies use agstat! (agstat!) only. Morrison et al. (2011) prorate from provinces to sub-provinces using an auxiliary variable that is reported at both scales. Aalders and Aitkenhead (2006) use neural networks, Bayesian networks and regression trees to demonstrate that a range of social variables reported in an agricultural census can be used to assist in LU mapping.

Wästfelt (2009) presents a method for post-classifying a LC map developed from an unsupervised classification algorithm into a LU map. The method encodes the spatial configuration of the LC within a study area of known LU through maximum nearest-neighbour distances. Other regions can then be re-classified based on the similarity of their LC configuration based on the same unsupervised learning algorithm to that of the study region. A similar method is presented in Chen (2002), who uses homogeneity metrics to separate urban density based on a method using the spatial context of a pixel to help determine the LU of that pixel. By introducing the spatial context, they demonstrate that it becomes possible to distinguish different LUs, even when the spectral profiles of these pixels are very similar.

Two methods that incorporate the AgStats data into the primary classification procedure are You and Wood (2005) and Walker and Mallawaarachchi (1998). You and Wood (2005) seek to estimate the proportion of each pixel that is used for each of a set of crop types within a set of pre-defined “production systems” by using linear programming to minimise the cross-entropy between these proportions and ‘prior’ estimates thereof, subject to a range of constraints. Production systems partition the space of possible production techniques with respect to equipment, inputs (e.g. chemical and mechanical) and plant varietals (e.g. high/low yield). The priors are developed from any available data, including: production statistics, land use data, satellite-based land cover map, biophysical crop ‘suitability’ assessments, population density, distance to urban centers and any prior knowledge of crop distribution (You et al., 2009). The data are combined into the prior using practical methods that suit the specific forms of data. The method also requires estimates of the total proportion of the total physical area of each crop type within each production system. The authors adapt and extend this approach in You and Wood (2006) and You et al. (2009) by incorporating various other data sources.

The second method is that of Walker and Mallawaarachchi (1998), which presents the algorithm SPatial REallocation of Aggregated Data (SPREAD). This is a linear programming algorithm that allocates a LU to each pixel within a region while ensuring that the total area of each LU is equal to that reported by the Australian Bureau of Statistics (ABS) AgStats for the corresponding time period. This method is discussed further below.

1.1.3 Land Use Mapping in Australia

Here we review the continental scale LU and LC products available in Australia and describe the context for how some of the methods considered in this thesis could contribute to the Australian national mapping system through the NLUM.

Australia has two existing continental LU products and one continental scale LC product, each of which serve different purposes. The LU products are the Catchment Scale Land Use Map of Australia (CLUM) (Department of Agriculture, Fisheries and Forestry, 2014) and the NLUMs (Stewart et al., 2001; Smart et al., 2006). The LC product is the Dynamic Land Cover Dataset (DLCD) (Lymburner et al., 2010), which provides continental scale information and has significant categorical overlap with the CLUM and NLUM in terms of spatial area in Australia.
The CLUM is effectively a high resolution ongoing census of LU across Australia. This data is used extensively for Natural Resource Management (NRM) evaluation and activities such as water quality, soil erosion and acidification (Australian Bureau of Agricultural and Resource Economics, 2011). The data is collected by various state government agencies across Australia, with regions being updated as new data becomes available. While detailed, the currency of the CLUM varies spatially and at any one time varies by around a decade. Consequently, it has limited utility for analyses that require regular ongoing measurement of a phenomena, such as GHG emissions or modelling with a temporal aspect. Data collection is largely manual and the methods used differ between agencies and hence spatially. While the data is presented nationally using a consistent classification scheme (the Australian Land Use and Management (ALUM)), since LU is somewhat subjective and the data is collected and prepared by a diverse group of individuals working for different agencies, it is likely that there will be coding discrepancies between jurisdictions. The resources allocated to the CLUM also vary between agencies and consequently, there are regional differences in the level of accuracy. The required attribute accuracy of these data is at least 80 percent (Australian Bureau of Agricultural and Resource Economics, 2011). We use the CLUM data in the current work for validation of our results in selected regions where the currency aligns.

The NLUMs provide a time series of LU maps of Australia with each map corresponding to a twelve month period. These are at a coarser resolution and are less accurate than the CLUM. They are much cheaper to produce, as described below, and the same (repeatable) methodology is applied across the entire continent and (more or less) to each time period. These data are used for “synoptic-level land use assessments, and for strategic planning and evaluation (such as setting regional investment priorities and developing programs for natural resource management). They are also used in modelling applications, such as national carbon accounting and salinity assessments at the river basin level.” (Australian Bureau of Agricultural and Resource Economics, 2011). To date, national maps have been produced for the financial years 1992–93, 1993–94, 1996–97, 2000–01, 2005–06 and 2010–11. Presently a new map is prepared for every agricultural census year, which, since 2000–01, is every five years. The NLUMs are published using the ALUM classification and all non-agricultural LUs determined deterministically from administrative datasets (i.e. only agricultural LUs are estimated statistically).

The Dynamic Land Cover Dataset (DLCD) (Lymburner et al., 2010) is a LC dataset produced by clustering pixels based on summary metrics extracted from Moderate Resolution Imaging Spectroradiometer (MODIS) Enhanced Vegetation Index (EVI) 250m data for an eight-year period, followed by a post-classification step that combines a range of other datasets and expert opinion. While this is an LC dataset, some of the classes correspond directly to LU classes mapped in the CLUM and NLUM datasets: specifically, irrigated cropping and rain-fed (non-irrigated) cropping, pasture and sugar. Since these, in particular cropping and pasture, are extensive in most agricultural regions, there should be a reasonably strong link between this dataset and those mentioned above. The accuracy of the DLCD dataset is reported using a fuzzy logic system (Zhang and Foody, 1998), which reports matches at different levels of accuracy. The reported match rates against field validation sites are exact: 30 percent, very similar: 35 percent, moderately similar: 10 percent, somewhat similar: 18 percent and completely mismatched: 7 percent.

Since they are produced for a much narrower time period that is consistent across the entire
continent, as a time series, the NLUMs are suitable for a range of purposes that the CLUM or DLCD are not. One of the benefits of the methodologies used in compiling the NLUM, is that they preserve the regional total reported in the agricultural censuses and hence provide protection from serious error at the regional scale. This is important for applications such as GHG reporting, as the regional and national totals are of prime importance for international reporting requirements. Further, being derived from censuses, the regional totals are likely to be relatively accurate. It is also important for many modelling applications which aim to produce a time series of information related to LU. Stewart et al. (2001) describe many actual and potential applications to policy, regulation, NRM etc. of the NLUM data.

Analysis of the relative costs of producing the NLUM and CLUM was done in Stewart et al. (2001). They estimated that the cost of production of an NLUM was between 3.3 and 12.5 cents per km$^2$, while the cost of detailed mapping, based on similar approaches used in collecting CLUM data, was around 50 cents per km$^2$. It should be noted, however, that the NLUMs incorporate information from the CLUM data and their accuracy would be reduced without those data. Hence, the cost of producing an NLUM of similar quality would be higher without the CLUM data.

1.2 Outline of This Thesis

The work undertaken in this thesis is motivated by spatially consistent, repeatable, continental scale LU mapping, using agricultural classification schemes that contain in the order of 20 to 40 LU classes, in a manner in which the outputs accurately reflect uncertainties. The techniques employed must be highly automated in order to minimise labour costs, avoid spatial inconsistencies that could be introduced by analyst intervention, and allow for regular updates. Of the methods reviewed above, the methodology underpinning SPREAD, which is a highly automated technique that was designed for use with the same data we have available, is the most appropriate starting point. The methodologies underpinning the other methods are not suitable for one or more of the following reasons: they do not propagate uncertainties through to the outputs, require manual intervention or interpretation on the part of the analyst, are designed for simpler classification schemes, cannot readily be extended to include additional data sources, are targeted and much less complex classification schemes, pertain to spatial partitions other than rasters, or use input data which is unavailable (at acceptable cost) at a national scale.

As discussed previously, at geographic scale and complexity of classification scheme considered herein, remotely sensed imagery alone are not sufficient for mapping the LU. Hence, classifiers used for this type of mapping need to incorporate other sources of information in order to discriminate effectively between LUs. We would like the methods we develop here to be relatively easy to modify to incorporate new data sources as they identified or become available.

We present a Bayesian method that incorporates a probabilistic classifier with regional estimates of the area of land used for various LUs, which we refer to as “areal constraints”. The prior used by this method takes the form of a categorical distribution over the possible LU classes at each pixel and the posterior is categorical over the same set of classes. We then present a series of extensions to this algorithm that incorporate additional sources of ancillary data or exploit
spatial/temporal relationships between neighbouring pixels or time points.

In the second chapter we present SPatial REallocation of Aggregated Data Version II (SPREAD II), the algorithm developed for the production of version 3 of the NLUM. This is contrasted to SPREAD, which was used to produce versions 1 and 2 of the NLUM. We explore the performance of SPREAD II on a study region, by comparing its results to the CLUM and to SPREAD. We conduct a simulation study to ascertain its robustness to biases in the areal constraint data with respect to noise in the satellite imagery. In both these analyses we compare it to ‘simpler’ classifiers which incorporate the same inputs using different methodologies.

In the third chapter we introduce another source of ancillary data, in the form of a digital map that specifies regions where certain groups (which we refer to as “super types”) of LUs occur, and contrast the results with those obtained in the second chapter. Such data are often cheaply available from other industry-specific or regional studies and provide valuable information that we would like to exploit. Indeed, for some LUs such data are taken as deterministic in the production of the NLUMs when the LU is unambiguous; for example military areas and national parks. The challenge addressed in Chapter 3 is to leverage such data when it is not deterministic. Our experience in producing version 3 of the NLUM suggested that this will improve classification performance regardless of the quality of the satellite imagery.

In the fourth chapter we evaluate two methods for including spatial context. The first of these methods is based on Markov Random Fields (MRFs) and should improve our results if pixels of the same LU are clustered spatially. It is unlikely that this will improve classification accuracy for LUs that occur with spatial extents smaller than that of the satellite imagery being used, but may benefit broad scale LUs. The second method is a transductive technique designed to exploit local similarities in climatic conditions, seasonality, phenology, management and other factors that affect the spectral signature of specific crops spatially. Both of these methods are only likely to improve classification performance if the satellite imagery is informative in its own right.

In the fifth chapter we explore making a sequence of maps. Both compound and cascade classifiers are introduced and trialled with and without transition probabilities (a matrix of probabilities specifying the probability that each LU will ‘transition’ to itself or some other between two time points). In the work here we develop the transition matrix from regional statistics published by the ABS and discuss the limitations of this approach. Once again, these methods are only likely to improve classification accuracy if the satellite imagery is informative in its own right.

In the sixth and final chapter, we summarise the results from a number of study regions, discuss our findings, draw conclusions and describe further research.
2. SPATIAL REALLOCATION OF AGGREGATED DATA
VERSION II

2.1 Introduction

This chapter considers a new approach to the incorporation of regional statistics in LU mapping, SPREAD II. This method combines information on crop types from AVHRR NDVI satellite data with regional statistics providing estimates of the area of land within a region is used for each of the LUs being mapped. We refer to these data as “areal constraints”. The areal constraints used here are derived from AgStats data published by the ABS. The background theory underpinning SPREAD II is briefly described in Bureau of Rural Sciences (2004, Appendix 7), but is restated below for convenience and because this method provides a starting point for the developments considered in this thesis.

A wide range of remotely sensed data is available for pixel classification and a great deal of work has been done on the development and selection of appropriate data and classifiers. The methods explored in this thesis assume that the remotely sensed data and the underlying probabilistic classifier used to estimate \( Y_i | C_i = k \), where \( k \) is a specific LU label, \( Y_i \) is the remotely sensed data available at pixel \( i \) and \( C_i \) is the LU of pixel \( i \), have been chosen. We also assume that all data have been registered and cleaned; that is, we do not address issues related to the preparation of the remotely sensed data, the choice of ‘primary classifier’, or any other data used in the classification process beyond describing the specific data we have used and how they have been prepared.

The SPREAD II approach addresses shortcomings of SPREAD that were identified when producing version 1 of the NLUMs. It allows for the integration of both AgStats data and other forms of ancillary data with the outputs of an existing probabilistic classifier and the posterior distribution provides a meaningful measure of uncertainty.

We consider three sets of analyses. The first considers regions that were used in earlier versions of the NLUM. The results presented for these regions are based on version 3 of the NLUMs. It is presented here to allow comparison with SPREAD. The second set of analyses consider more recent time periods. The third set of analyses are based on synthetic data. These analyses are important in understanding when SPREAD II does and does not perform well and how it compares to ‘simpler’, less computationally intensive classifiers under various scenarios.

Section 2.2 presents the SPREAD II methodology, which we refer to hereafter as “vanilla SPREAD II” to distinguish it from variants that incorporate other ancillary data or techniques data as presented in other chapters. We also present an overview of the SPREAD methodology, noting some deficiencies identified during the production of version 2 of the 2000–01 NLUM. Section 2.3 describes the data and preparation required to produce that map. Section 2.3.4 com-
parses the outputs of SPREAD II to those of SPREAD and the CLUM. Section 2.4 presents the methodological differences and the data used in the work presented here and Section 2.4.3 compares these results to the CLUM. Section 2.5 presents and discusses the simulation study and finally, Section 2.6 summarises the results of this chapter.

2.2 Methods

A map comprises of a number of regions. A region is a lattice of \( N \) pixels that we will generally index by \( n \). There are \( K \) LUs that we will index by \( k \) and pixel \( n \) has an actual LU \( C_n \in \{1, \ldots, K\} \). Here, a region is a reporting region for the AgStats. The total number of pixels with LU \( k \) is

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

where \( I(a = b) \) is the indicator function taking the value one if its argument is true and zero otherwise. We use bold notation to represent the vector version of arguments over their natural dimension. For example \( \mathbf{T} = (T_1, \ldots, T_K) \), i.e. over the index \( k \).

The aim of the inference is to infer \( C \), the LU of all individual pixels. We have \( M \) measurements \( \mathbf{Y}_n = (Y_{1n}, \ldots, Y_{Mn}) \) for each pixel, and a sample of locations \( j = 1, \ldots, H \) where the LU is known, which we shall refer to as control sites. In our work to date with both SPREAD and SPREAD II, \( \mathbf{Y}_n \) are fortnightly or four-weekly NDVI values based on NOAA-AVHRR data over a twelve month period (see Ramsey et al. (1995) for example).

2.2.1 SPatial REallocation of Aggregated Data (SPREAD)

SPREAD performs the above inference using a linear programming algorithm that disaggregates LU within a region while ensuring that the total area of each LU is equal to that reported by the ABS AgStats for the corresponding time period. It was used in Stewart et al. (2001) to produce version 1 of the NLUMs. While Walker and Mallawaarachchi (1998) consider a range of practical issues in the generation of LU maps (such as non-coincident target and statistical zones and pixels of mixed LU) in their presentation of SPREAD, we will focus on the core scientific basis of the technique.

The Gower metric (Gower, 1971) between the \( n \)th pixel and the \( j \)th control site is given by

\[
g_{nj} = \frac{1}{|S_k|} \sum_{s \in S_k} \frac{|Y_{sn} - Y_{sj}|}{\text{Range}(Y_s)},
\]

where \( S_k \) is the set of all control sites with LU \( k \) and \( Y_s \) is the set of measurements of the \( s \)th variable in the entire dataset (i.e. all control sites and pixels).

The SPREAD algorithm assumes that \( \mathbf{T} \) is known and fixed. It seeks to allocate the LUs to the pixels to form predictions \( \hat{C}_n \) such that the areal constraints given by equation (2.1) are satisfied and the objective function

\[
G = \sum_{n=1}^{N} \prod_{k=1}^{K} h_{nk} I(C_n = k),
\]
where $h_{nk} = \min_{j \in S_k} g_{nj}$, is minimised.

SPREAD is a novel and significant research achievement but exhibits a number of statistical flaws that became evident while using it to produce the 1996–97 NLUM (Stewart et al., 2001). These are as follows:

1. The algorithm cannot incorporate uncertainty in the AgStats ($\bf{T}$). This is an important issue as administrative records are rarely without error. We note that in the periods considered herein, the AgStats are based on censuses and hence the error is hard to quantify, but if the methods were applied to survey (as opposed to census) data, the estimates will contain, potentially significant, sampling error.

2. The algorithm is an optimisation algorithm and not a statistical model or framework and hence does not provide a framework for estimating and expressing uncertainty arising from uncertainty in the inputs and the estimation process in its outputs. In practice there is considerable ambiguity in the classification of pixels due to failures of assumptions (e.g. pixels of mixed LU) and limitations in the ability to discriminate between LU classes.

3. The classifier based on the Gower metric is inefficient. This metric defines a geometry that may not discriminate between the LUs effectively. As an extreme example, if a single measurement $s$ has all the discriminatory power, the other measurements simply add noise and could swamp this information. The authors note, however, that it “could be replaced with any of the measures commonly used in supervised classification of imagery”.

SPREAD II attempts to address these problems by using a very different approach that allows assessment of uncertainty, incorporation of any probabilistic classifier and extension to incorporate of a wide range of ancillary data of varying forms.

### 2.2.2 Spatial REallocation of Aggregated Data Version II (SPREAD II)

Vanilla SPREAD II is based on the same data as SPREAD and uses an Markov Chain Monte Carlo (MCMC) algorithm to produce a posterior probability distribution over the various agricultural LU classes for each pixel.

Let $\Theta$ be a vector of parameters and $[\bullet]$ denote a density function or probability. The algorithm uses a categorical prior over the various LUs that occur within a region at each pixel within the region. We denote the parameters of these priors with $\Pi$. As described below, in the vanilla form presented here it is important that categorical prior be the same at every pixel. The posterior distribution integrates the AgStats data with the outputs of a probabilistic classifier $[Y_n|C_n]$ using the conditional distribution $[T|\Theta]$ as a stochastic constraint. A ‘final allocation’ (a single layer map) is then produced that maximises the posterior mean while respecting the areas reported in the areal constraint data.

The joint distribution of the parameters and data in SPREAD II is:

$$[C, T, Y, \Theta, \Pi].$$  \hspace{1cm} (2.4)

At each step in the MCMC procedure, we select a LU for a given pixel and hence require the
full conditional distribution \( [C_i, T|C_{-i}, T_{-i}, Y, \Theta, \Pi] \), where \( C_{-i} \) is \( C \) with \( C_i \) excluded and \( T_{-i} \) is \( T \) calculated excluding \( C_i \). This can be decomposed as

\[
[C_i, T|C_{-i}, T_{-i}, Y, \Theta, \Pi] \\
\propto [Y_i|C_i][C|T, \Theta, \Pi][T|T_{-i}, \Theta, \Pi][T_{-i}, \Theta, \Pi] \\
\propto [Y_i|C_i][C_i|\Pi][T|\Theta] \\
\propto \frac{[Y_i|C_i][C_i|\Pi][T|\Theta]}{[T|\Pi]}, \tag{2.5}
\]

Here we have asserted that the pixels are independent, the satellite imagery for a pixel depends only on the LU of that pixel and that the LU of a pixel is independent of \( \Theta \) given \( T \). We treat \( \Theta \) and \( \Pi \) as constants and hence \( [T_{-i}, \Theta, \Pi] \) is constant for all \( C_i \in K \) and can be ignored.

We work with areas rather than pixel counts by substituting \([A|\Theta]\) for \([T|\Theta]\), where \( A \) is a vector of areas allocated for each of the LUs. The area of pixel \( n \) is \( a_n \).

The terms in (2.5) are quite general and, provided the conditional dependence is respected, a wide range of models could be used for each. We have used:

- \([Y_i|C_i]\) is estimated using a kernel density smoother from metrics calculated from the temporal NDVI profiles of the control site; specifically: the mean, range and time of year of the maximum NDVI value. These metrics are very similar to those chosen in DeFries et al. (1998) except we use time of year of maximum NDVI, which is subject to seasonality and hence not meaningful in the context of global study such as theirs.

- \([C_i = k|\Pi]\) is a constant \( 1/K \forall k \in \{1, \ldots, K\} \), and hence \([C]\) is the constant \( 1/K^N \) and \([T]\) is multinomial.

- \( \Theta \) is a constant vector of length \( K \) containing estimates, based on the AgStats, of the area of each LU (in hectares).

- \([A|\Theta]\) is multivariate normal with a mean vector \([\mu_1, \mu_2, \ldots, \mu_K]\) and a diagonal \( K \times K \) covariance matrix with diagonal entries \([\sigma^2_1, \sigma^2_2, \ldots, \sigma^2_K]\). In the published versions of the NLUMs produced to date and the current work, \( \sigma_k = 200 \) (hectares).

When these terms are substituted into (2.5) we get

\[
[C_i, T|C_{-i}, T_{-i}, Y, \Theta, \Pi] \propto \frac{[Y_i|C_i][C_i = k](\tilde{T}_k + 1)\phi(\tilde{A}_k + a_n; \mu_k, \sigma^2_k)}{\phi(\tilde{A}_k; \mu_k, \sigma^2_k)} \tag{2.6}
\]

where \( a_n \) is the area of the pixel under consideration and \( \phi(x; \mu, \sigma^2) \) is the normal density function with mean \( \mu \) and variance \( \sigma^2 \). Quantities with a tilde represent values before the new LU is selected (i.e. excluding the current pixel).

Note that (2.6) does not depend on \( \Pi \). This occurs because \([C|\Pi]\) in the numerator of (2.5) cancels with the product of probabilities in the multinomial distribution in the denominator and implies that all prior information must enter the model through the constraints. This is computationally convenient, but restrictive, as one could introduce all kinds of data into the model via \( \Pi \), which is the standard way of doing so in the Bayesian context. The restriction
could be removed if we let the probabilities for the categorical prior differ between pixels, but the specification that they be the same for all pixels is required to make (2.5) tractable. In full generality \([T|\Pi]\) is
\[
[T] = \sum_{S} \prod_{n=1}^{N} \prod_{k=1}^{K} \pi_{kn}^{I(C_n=k)},
\]
where \(\pi_{kn}\) is the probability that pixel \(n\) has LU \(k\) and \(S\) is the set such that (2.1) holds for all \(k \in K\). The cardinality of \(S\) is \(N!/T_1!T_2!\cdots T_k!\). If, however, each pixel has the same categorical prior (that is \([C_i=k]=\pi_k\)), then \([T|\Pi]\) is multinomial and easy to work with. We present an approximation in Chapter 5 in the context of producing a sequence of maps that could be used to remove this restriction, but do not pursue this further here.

When producing version 3 of the NLUMs we added further constraints, based on spatial data describing zones where irrigation and horticulture occur. Letting \(S\) be the total area of agricultural land inside the irrigation zones, \(A_I\) be the total area of land allocated to irrigated LUs inside the irrigation zones, \(\alpha\) be the mean proportion of the total area of agricultural land inside the irrigation zones that is used for irrigated LUs, and \(\sigma^2\) be the variance of the constraint, we multiply (2.5) by
\[
\phi(A_I; \alpha S, \sigma^2).
\]
An analogous expression is used for the horticulture zones and the complementary non-irrigation and non-horticulture zones.

This modification was introduced to deal with horticultural crops being allocated where they were known not to occur. The introduction of these terms is based on a heuristic argument, as \([T]\) is no longer a simple multinomial distribution. This issue is resolved in Chapter 3.

Through the running of the MCMC algorithm, counts of how many times each pixel is allocated each LU \((c_{kn})\) are maintained. Once the algorithm has completed, the posterior probabilities of each pixel having each LU are calculated using the posterior mean of the sample,
\[
\hat{\pi}_{kn} = \frac{c_{kn}}{R},
\]
where \(R\) is the number of iterations (excluding burn-in) for which the MCMC algorithm was run.

The Gibbs sampling algorithm produces a distribution of maps (i.e. updates sufficiently separated in the sequence can be taken as samples from the joint posterior of the pixels). While this is attractive theoretically, users generally want a (single layer) thematic map. An issue that has been noted of Bayesian approaches is that proportional prior probability may be used; but in this case large classes tend to be overestimated and small classes are underestimated or even disappear (Carfagna and Gallego, 2005). This was also raised by Strahler (1980). This was our initial experience with SPREAD II when producing a thematic map from the posterior. In order to combat this, we apply Algorithm 1 which successively chooses the least abundant LU class, as reported in the to the AgStats, and allocates that class to the pixels with the highest posterior probabilities for that class. This process is then repeated on the remaining LU classes until all pixels are allocated. This was chosen as a practical approach and ensures that the final allocation map is consistent with the agricultural statistics at the regional scale. This is appealing from a technical point of view (assuming that the areal constraints are accurate) and
may also make the map more appropriate for uses which are concerned largely with regional totals as is often the case for planning purposes.

**Algorithm 1** Final allocation Algorithm

1. Create a list of all pixels in the region being classified.
2. Sort this list in descending order based on the probability value for the rarest LU.
3. Allocate the first \( m \) pixels such that area constraint for the rarest LU is satisfied and remove the allocated pixels from the list.
4. Repeat steps 2 and 3 until all pixels are allocated.

This describes the mathematical foundation of SPREAD II. Some implementation details are given in Section 2.3.2.

### 2.2.3 A Simple Bayesian Classifier

The probability model (2.4) is a generalisation of:

\[
[C, Y, \Theta, \Pi] = [Y|C][C|\Pi][\Pi, \Theta] = \prod_{n=1}^{N} [Y_n|C_n][C_n|\Pi][\Pi, \Theta],
\]

in which the second line assumes that the LU class of a pixel is independent of those of the other pixels. A relatively early appearance of this in the context of “maximum likelihood classification” (specifically QDA) is Strahler (1980). A simple way of incorporating the AgStats data into the classification is to let \([\Pi, \Theta]\) be degenerate and \([C_n = k|\Pi]\) = \(\pi_k = \mu_k / A\).

That is, the prior probability of a pixel having LU \( k \) is the proportion of all (agricultural) land \((A = \sum_{k=1}^{K} \mu_k)\) that has (agricultural) LU \( k \) according to the AgStats. This corresponds to \([C_i = k|\Pi]\) = \(\pi_k\) and \([T, \Pi, \Theta]\) being a multinomial distribution over the counts of pixels with each LU in (2.5). We present results for this model for comparison with the vanilla SPREAD II.

We also present results for SPREAD II with \(\sigma_k^2 = A\pi_k(1 - \pi_k)\), which corresponds to the variance of \(T_k\) under the corresponding multinomial distribution. This will provide insight into the impacts of ignoring the off-diagonal elements of the covariance matrix when we compare it to this ‘simple’ Bayesian model.


SPREAD II was used to produce the agricultural components of the NLUMs for the years 1992–93, 1993–94, 1996–97, 1998–99, 2000–01 and 2001–02 (Smart et al., 2006). The approach was based on the same LU classes and input data types as used by Stewart et al. (2001). An overview of the NLUMs and the major steps prior to running SPREAD II in the context of producing the 2000–01 NLUM are given here and the reader is referred to Stewart et al.
All maps were produced by using regions corresponding to Statistical Local Areas (SLAs) which are the smallest reporting regions for the AgStats. The currency of each map is the one year period to 31 March in the corresponding collection year. This is the same as the reference period for the AgStats up to and including the 1998–99 survey. From the 1999–00 survey onward the AgStats reference period was changed to the year ending 30 June in the year of collection. As well as facilitating comparison of the data, the original reference period was retained for two reasons. First, the change in reference period would have had little effect on the specific agricultural statistics used and, second, using satellite imagery corresponding to the old reference period is better suited to discriminating winter crops.

The maps were produced in geographic coordinates referred to the Geocentric Datum of Australia 1994 (GDA). They have a pixel size of 0.01 degrees. The area of the pixels ranges from approximately 1.2km\(^2\) in the far north of Australia to approximately 0.9km\(^2\) in the far south. This coordinate system, pixel size and the pixel alignment were chosen to match those of the NDVI imagery.

The LU classification used for the maps is version 5 of the Australian Land Use and Management (ALUM) classification.

### 2.3.1 Data

Here we describe the data used in the construction of these maps and the preparation thereof. This data was prepared by Robert Smart of Australian Bureau of Agricultural and Resource Economics and Science (ABARES). The description provided here was also largely prepared by him and further developed through personal communications with him through the course of this study.

#### 2.3.1.1 Preparing the Agricultural Land Use Mask

Agricultural land is the complement of non-agricultural land, which was identified from a range of existing digital maps that varied from year to year. For the 2000–01 NLUM these include topographic data, the CLUM data, and a range of spatial data provided by various state and federal government agencies. See Smart et al. (2006) for the full details.

#### 2.3.1.2 Agricultural Statistics (Areal Constraint) Data

The agricultural data used to produce the 2000–01 national map were based on the 2000–01 agricultural census. The commodity classes available in the agricultural statistics data vary from year to year and must be transformed into standard classes. The agricultural statistics data were first summarised into commodity groups with vegetable areas corrected for multiple cropping and orchard tree numbers converted to areas. The data for pastures, cereals, legumes and oilseeds were then adjusted to compensate for double cropping using data from the 1996–97 Farm Survey (Australian Bureau of Agricultural and Resource Economics, 1997). The data
were then aggregated to 20 commodity groups and each of these disaggregated into dryland and irrigated (for a total of 40 commodity groups) using irrigation area data from the AgStats. Finally, six commodity groups for which the AgStats contained no data were dropped and the data for each SLA scaled to accord with the total area of agricultural land identified within that SLA. For a small number of SLAs it was necessary to make minor modifications to the agricultural statistics data to avoid aberrant results from the scaling procedure. These modifications generally appeared to be due to the LU reported by one or more reporting units being misallocated to the SLA corresponding to the address of the reporting unit, rather than to the SLA where the LU occurred.

2.3.1.3 NDVI Data

At present, the geographic extent considered herein limits the choice of suitable satellite imagery, due to the cost of acquiring the large volumes of fine scale imagery that would be required. The exception to this is, of course, Landsat. However, as noted above, Landsat’s 16-day return time combined with the high frequency of cloud cover over many agricultural regions makes it difficult to build classifiers to distinguish between agricultural LU classes. The other datasets available at these scales (in Australia) are MODIS and AVHRR and hence we are limited to coarse (1km) to moderate (250m) spatial resolutions. Further, the training data we use here was collected in the 1990s before MODIS data was available and consequently we are limited to using AVHRR data.

Maps were produced using a time series of 13 composite cloud-corrected NDVI images prepared by the Environmental Resources Information Network (ERIN) from the AVHRR NDVI archive covering the reference period. An image was created for each 28 day period in the year to 31 March (ignoring 31 March and 29 February); for each pixel, a spline was used to create an NDVI value for each day and the value for each output image was the average of the daily values for the period represented.

2.3.1.4 Control Site Data

Control sites were selected from a database of ground control points compiled over the period 1997 to 1999 for the production of the 1996–97 SPREAD map. The control site data were collected by state and territory agencies (both government and private) using a mixture of field work and analysis of aerial photos and satellite imagery (Stewart et al., 2001). Control sites for a given commodity were sought in areas where that commodity was abundant according to the 1996–97 agricultural census. This database contains around 2871 suitable control sites, each of which has location, LU and year attributes specifying the location of the site and the LU occurring at the site in the year of observation. Observations were made for the years ending 31 March 1997, 1998 and 1999 at each control site. Table 2.1 shows the total number of control sites for each LU in each state (i.e. over all years). All control sites from all years were used in mapping each LU in every SLA. We recognise that it is less than ideal to use training data from different time periods to the one under consideration, or from locations as geographically distant from the pixel under consideration as occurs with this limited database, but this was unavoidable due to the very small number of control sites for various LUs.
<table>
<thead>
<tr>
<th>Description</th>
<th>Irrigated</th>
<th>NSW</th>
<th>Vic</th>
<th>Qld</th>
<th>SA</th>
<th>WA</th>
<th>Tas</th>
<th>NT</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residual/Native pastures</td>
<td>No</td>
<td>21</td>
<td>70</td>
<td>155</td>
<td>93</td>
<td>10</td>
<td>0</td>
<td>55</td>
<td>404</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0</td>
<td>20</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>Agroforestry</td>
<td>No</td>
<td>0</td>
<td>27</td>
<td>0</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sown pastures</td>
<td>No</td>
<td>36</td>
<td>91</td>
<td>151</td>
<td>121</td>
<td>10</td>
<td>44</td>
<td>7</td>
<td>460</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>17</td>
<td>37</td>
<td>35</td>
<td>17</td>
<td>0</td>
<td>10</td>
<td>2</td>
<td>118</td>
</tr>
<tr>
<td>Cereals excluding rice</td>
<td>No</td>
<td>48</td>
<td>94</td>
<td>169</td>
<td>235</td>
<td>0</td>
<td>14</td>
<td>1</td>
<td>561</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>10</td>
<td>12</td>
<td>31</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>59</td>
</tr>
<tr>
<td>Rice</td>
<td>Yes</td>
<td>2</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Legumes</td>
<td>No</td>
<td>11</td>
<td>35</td>
<td>32</td>
<td>73</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>151</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>1</td>
<td>9</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>26</td>
</tr>
<tr>
<td>Oilseeds</td>
<td>No</td>
<td>11</td>
<td>35</td>
<td>13</td>
<td>39</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Sugar cane</td>
<td>No</td>
<td>12</td>
<td>0</td>
<td>65</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0</td>
<td>0</td>
<td>134</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>134</td>
</tr>
<tr>
<td>Non-cereal forage crops</td>
<td>No</td>
<td>1</td>
<td>14</td>
<td>13</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Cotton</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0</td>
<td>0</td>
<td>53</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>Other non-cereal crops</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0</td>
<td>3</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>Other vegetables</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>10</td>
<td>18</td>
<td>48</td>
<td>12</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Potatoes</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>4</td>
<td>13</td>
<td>22</td>
<td>9</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>Citrus fruit</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>3</td>
<td>3</td>
<td>27</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>57</td>
</tr>
<tr>
<td>Apples</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0</td>
<td>12</td>
<td>9</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>Pears</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Stone fruit</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0</td>
<td>21</td>
<td>27</td>
<td>12</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>67</td>
</tr>
<tr>
<td>Nuts</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Plantation fruit</td>
<td>No</td>
<td>4</td>
<td>0</td>
<td>61</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Grapes</td>
<td>No</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>6</td>
<td>18</td>
<td>6</td>
<td>77</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>107</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td>205</td>
<td>564</td>
<td>1146</td>
<td>753</td>
<td>20</td>
<td>108</td>
<td>75</td>
<td>2871</td>
</tr>
</tbody>
</table>

Tab. 2.1: Number of control sites available for each commodity in each state.
2.3.1.5 Irrigation and Horticulture constraints

The data used for the irrigation constraint was based on an irrigation boundaries dataset published by the National Land and Water Resources Audit (NLWRA) (National Land and Water Resources Audit, 1999) and some additional Victorian irrigation areas. The horticulture constraints were based on various land cover datasets (Bureau of Rural Sciences, 1999, 2005). These were converted to rasters of the same topology as the NDVI data. See Bureau of Rural Sciences (2010a) for full details.

2.3.2 Running SPREAD II

The constant $\alpha$ from (2.2.2) was set to 70 percent for the irrigation constraint where there was sufficient area reported in the AgStats to achieve this. This was based on the observation that 70 percent of the control sites with irrigated LU fell inside the irrigation mask. For the horticulture mask $\alpha$ was set to 90 percent, where there was sufficient area reported in the AgStats to achieve this, based on the judgement that the horticulture mask is a reasonably accurate representation of horticulture regions.

LUs for which the area prescribed in the AgStats was very small, or for which we had no control sites, were removed and the areas of the remaining LUs were scaled up to retain the same total area of agricultural LU according to the AgStats within the region.

The first stage of the SPREAD II algorithm was to allocate an initial LU to each pixel in the SLA being analysed. This allocation was done by setting the LU for $m$ pixels to $k$, where $\sum_{n=1}^{m} a_n \approx \mu_k$, where $a_n$ is the area of pixel $n$ and $\mu_k$ is the area of LU $k$ according to the AgStats. Note that any allocation of LUs to the pixels will suffice, but this method was chosen so the algorithm starts close to the expectation of the ‘constraint’ distributions, $[A|\Theta]$. We note that the burn-in period for the algorithm could potentially be reduced further by allocating the initial LU of each pixel based on $[C_n = c_n|Y_n]$ (which can be calculated from $[Y_n|C_n = c_n]$ in the obvious way assuming $[Y_n]$ is constant) and assigning LUs to the pixels as described above.

All pixels $n \in \{1, \ldots, N\}$ were then checked to ensure that $[Y_n|C_n = c_n] > 0$, where $c_n$ is the LU initially allocated to pixel $n$, and if it was not, the LU for the pixel was swapped with that of another such that both had non-zero probability for their allocated LU. In some rare cases, a pixel had zero probability for all LUs. In this case, $[Y_n|C_n = k]$ was set to $1/K$ for all $k \in \{1, \ldots, K\}$.

Once the initial allocation was done, SPREAD II was run for a burn-in period of 2,000 iterations, then a further 10,000 iterations from which the posterior probabilities of each LU for each pixel were estimated using (2.7). Analysis of the outputs suggests that the burn-in of 2,000 iterations is more than sufficient for the MCMC to ‘forget’ its initial state, and that a run length of 10,000 is much more than sufficient to get good estimates of the probability; good estimates could be achieved with run lengths of around 1,000 in most cases. An analysis of the burn-in period is presented in Section 2.4.1.4.

The variances contained in $\Theta$ were chosen to allow ‘room to move’ in the algorithm, rather than realistic estimates of the variances of the AgStats estimates. This is partially justified in agricultural census years as the areal estimates contain negligible sampling error, though the
issue of non-sampling error still remains and is hard to quantify. For survey years, the variances
used in the constraints should be based on the variances of the AgStats estimates.

2.3.3 Software

A range of software was used in constructing these versions of the NLUMs. Workstation
ARC/INFO software, Version 9.1 under SunOS was used for all Geographic Information Sys-
tems (GIS) operations, including:

- construction of the non-agricultural LU mask,
- construction of the irrigation and horticulture masks,
- scaling the AgStats to accord with the area of agricultural land identified in each region,
- extracting NDVI profiles for the control sites and the agricultural pixels and exporting
  them to a format that can be read by the SPREAD II software, and
- importing the SPREAD II outputs from ASCII Comma Separated Values (CSV) files
  back into ESRI grid format.

The statistical package R (R Core Team, 2015) was used to do most of the preparation of the
data before running SPREAD II. This included:

- loading and preprocessing the AgStats data,
- calculating $[Y_n|C_n]$ by comparing the NDVI profiles of the control sites with those of the
  target pixels,
- setting up the parameters of (2.2.2), and
- serialising the outputs of SPREAD II for post-processing using ARC/INFO.

SPREAD II itself was written in C++ and compiled as a shared library called from R. A small
PERL script was used to coordinate the running of the various scripts and programs over all
regions.

2.3.4 Results

SPREAD and SPREAD II were run across the entire continent of Australia. In this section we
present examples of the output of SPREAD II and quantitative and qualitative comparisons of
the NLUMs produced using SPREAD, SPREAD II, and the CLUM data.

2.3.4.1 Quantitative Comparison to the Catchment Scale Land Use Map of
Australia (CLUM)

While the CLUM has its limitations, it was the only source of ground truth data available. It is
important to note that the results presented here are confounded, as the horticulture constraint
was developed from the CLUM data. The preparation of the figures shown in this section was
done by Robert Smart of ABARES. Many of the details contained in the following description
are also due to Robert and correspondence with Australian Collaborative Land Use Mapping
Program (ACLUMP) members.

The NLUMs were overlaid on CLUM data with the same currency and the percentage of pixels
with matching LUs calculated. Two methods of matching were used:

1. pixels were awarded a match if both the NLUM and CLUM data had identical LU codes,

2. pixels were awarded a match if the NLUM and CLUM LU codes matched under one or
more of a set of rules that compensate for known limitations affecting the comparability
of these data. The rules used and the reason for applying them are given in tables 2.2
and 2.3, which were prepared by Robert Smart when preparing the validation data.

We refer to these two matching methods as exact and relaxed hereafter.

The main limitations alluded to in the second matching method are:

- In the construction of the NLUMs for years up to and including 2001–02, it is assumed
  that no grazing occurs in forested areas where the crown cover exceeds 50 percent. This
  assumption leads to some misclassification of grazing (of natural vegetation and of min-
  imally modified pastures) as conserved (mainly remnant native vegetation). Figure 2.2
  illustrates this; much land is shown in the national scale maps as conserved that should
  be shown as grazing natural vegetation according to the catchment scale map.

- Distinguishing native and modified pasture has always been a problem for LU mapping.
  The AgStats data that we have used for the construction of NLUMs, from the 1992–93
  agricultural census data up to and including the 2001–02 agricultural survey data, give
  areas for sown pastures and native pastures, but always with a large shortfall. We have
  assumed this is due to under-reporting of native pasture. Comparison of the NLUMs based
  on these AgStats data with CLUM data suggests that the sown pasture areas reported in
  the AgStats data do not account for all of the modified pasture. Some of what we have
  mapped as native pasture in the NLUMs is really modified pasture, albeit pasture that
  is only minimally modified, such as mosaics of native and exotic pastures. This is also
  illustrated in Figure 2.2.

- The NLUMs and CLUM cannot be expected to show a high level of agreement with
  respect to irrigation status as this is not well mapped in the CLUM, where it is generally
  inferred from the presence of irrigation infrastructure.

We chose CLUM data with currency 1996–97 covering a collection of land parcels in south-east
Victoria, and CLUM with currency 2000–01 covering a collection of land parcels in central
New South Wales (NSW) for comparison with NLUM data of the same currency. These two
collections of land parcels and are shown in Figure 2.1.

The areas of each LU are provided in Table 2.4. Non-agricultural LUs make up a small pro-
portion of the central NSW test regions and a large proportion of the south-east Victoria test
Fig. 2.1: Test regions used for quantitative comparisons. Top panel: central NSW, bottom panel: south-east Victoria. Black lines show SLA boundaries in the top and bottom panels and state boundaries in the middle panel.
2. Spatial REallocation of Aggregated Data

Version II

---

<table>
<thead>
<tr>
<th>Tertiary codes used in the catchment scale land use maps</th>
<th>Keys to comments in Table 2.3</th>
<th>Allowed matches in the national scale land use map—only non-identical matches shown</th>
</tr>
</thead>
<tbody>
<tr>
<td>110</td>
<td>a, b</td>
<td>110, 111, 112, 113, 117, 651</td>
</tr>
<tr>
<td>111</td>
<td>b</td>
<td>111, 651, 661</td>
</tr>
<tr>
<td>112</td>
<td>b</td>
<td>112, 611, 651</td>
</tr>
<tr>
<td>113</td>
<td>b</td>
<td>113, 611, 651</td>
</tr>
<tr>
<td>120</td>
<td>b</td>
<td>120, 611, 651</td>
</tr>
<tr>
<td>121</td>
<td>b, e</td>
<td>120, 121, 611, 651</td>
</tr>
<tr>
<td>130</td>
<td>a</td>
<td>130, 133</td>
</tr>
<tr>
<td>133</td>
<td>c</td>
<td>133, 210</td>
</tr>
<tr>
<td>210</td>
<td>d</td>
<td>133, 210</td>
</tr>
<tr>
<td>312</td>
<td>e</td>
<td>310, 312</td>
</tr>
<tr>
<td>320</td>
<td>f, g</td>
<td>133, 210, 420</td>
</tr>
<tr>
<td>321</td>
<td>e, f, g</td>
<td>133, 210, 320, 321, 420</td>
</tr>
<tr>
<td>330</td>
<td>a, g</td>
<td>330, 331, 334, 338, 431, 436</td>
</tr>
<tr>
<td>350</td>
<td>a, g</td>
<td>350, 454</td>
</tr>
<tr>
<td>420</td>
<td>f, g</td>
<td>133, 210, 320, 420</td>
</tr>
<tr>
<td>430</td>
<td>a, g</td>
<td>331, 334, 430, 431, 436</td>
</tr>
<tr>
<td>450</td>
<td>a</td>
<td>450, 454</td>
</tr>
<tr>
<td>540</td>
<td>a, e</td>
<td>500, 540, 541, 542</td>
</tr>
<tr>
<td>542</td>
<td>e</td>
<td>500, 542</td>
</tr>
<tr>
<td>550</td>
<td>h</td>
<td>541, 550</td>
</tr>
<tr>
<td>552</td>
<td>h</td>
<td>541, 552</td>
</tr>
<tr>
<td>553</td>
<td>h</td>
<td>500, 541, 553</td>
</tr>
<tr>
<td>581</td>
<td>i</td>
<td>500, 580, 581</td>
</tr>
<tr>
<td>582</td>
<td>i</td>
<td>500, 582</td>
</tr>
<tr>
<td>610</td>
<td>a, j</td>
<td>610, 611, 620</td>
</tr>
</tbody>
</table>

Tab. 2.2: Concordance between ALUM codes mapped in SPREAD II and accepted matches in the CLUM used for validation.

---

regions. In the south-east Victoria test regions, the agricultural land is almost entirely used for grazing whereas in the central NSW test regions cropping and horticulture are almost as prevalent as grazing.

Comparisons were based on pixel counts rather than actual areas since the geographic extent of each region is small and consequently the area of each pixel within each region is approximately constant across the region. CLUM polygons were converted to a raster format with the same coordinate system, cell size and cell alignment as the NLUM data. The LU assigned to each CLUM pixel was that which occupied the largest proportion of the area covered by the pixel.

The percentage of pixels that matched based on the two criteria described above are given in Table 2.5. The differences between SPREAD and SPREAD II appear to be minor, with SPREAD doing slightly better in the case of exact matches and SPREAD II doing slightly better in the relaxed matches.

2.3.4.2 Qualitative Comparisons in Selected Regions

The following comparisons present two regions where there are large differences between SPREAD, SPREAD II and the CLUM.

Figure 2.2 shows the SPREAD data, the SPREAD II final allocation in and around “Cobar” for
2. Spatial REallocation of Aggregated DataVersion II

Key  Comment
a  Land use specified only to secondary level in the catchment scale maps—any corresponding tertiary level in the national scale maps is considered an acceptable match
b  Conservation and natural environments categories in the catchment scale maps—any water category with the tertiary level conservation qualifier is considered an acceptable match in the national scale maps
c  1.3.3 (remnant native vegetation) can also be mapped as 2.1.0 (grazing natural vegetation) in the national scale maps
d  2.1.0 (grazing natural vegetation) can also be mapped as 1.3.3 (remnant native vegetation) in the national scale maps
e  Land use specified to tertiary or secondary level in the catchment scale maps—the corresponding secondary or primary level is considered an acceptable match in the national scale maps
f  3.2.0 (grazing modified pasture) and its tertiary level categories can also be mapped as 1.3.3 (remnant native vegetation) or 2.1.0 (grazing natural vegetation) in the national scale maps
g  The irrigated equivalent of a matching dryland land use and the dryland equivalent of a matching irrigated land use are considered acceptable matches
h  5.5.0 (services) and its tertiary level categories are mapped as 5.0.0 (intensive uses) or as 5.4.1 (rural residential) in the national scale maps
i  5.8.1 (mines) and 5.8.2 (quarries) are mapped as 5.0.0 (intensive uses) or 5.8.0 (mining) in the national scale maps
j  6.1.0 (lake)—since 6.2.0 (reservoir/dam) is not offered as an alternative in the catchment scale maps concerned, 6.2.0 in the national scale maps is considered an acceptable match

Tab. 2.3: Comments for Table 2.2.

<table>
<thead>
<tr>
<th>Land use</th>
<th>Area of region (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Central NSW</td>
</tr>
<tr>
<td>Conservation and natural environments</td>
<td>6.4</td>
</tr>
<tr>
<td>Forestry</td>
<td>0.0</td>
</tr>
<tr>
<td>Grazing</td>
<td>47.8</td>
</tr>
<tr>
<td>Cropping and horticulture</td>
<td>35.7</td>
</tr>
<tr>
<td>Intensive uses</td>
<td>5.4</td>
</tr>
<tr>
<td>Water</td>
<td>4.7</td>
</tr>
<tr>
<td>Total</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Tab. 2.4: Mix of LUs included in the test polygons in the two test regions based on the AgStats and other input datasets.

<table>
<thead>
<tr>
<th>Map</th>
<th>Method</th>
<th>Test region</th>
<th>Exact (%)</th>
<th>Relaxed (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996–97</td>
<td>SPREAD</td>
<td>SE Victoria</td>
<td>60.7</td>
<td>83.8</td>
</tr>
<tr>
<td>1996–97</td>
<td>SPREAD II</td>
<td>SE Victoria</td>
<td>59.7</td>
<td>87.7</td>
</tr>
<tr>
<td>2000–01</td>
<td>SPREAD</td>
<td>Central NSW</td>
<td>4.1</td>
<td>56.2</td>
</tr>
<tr>
<td>2000–01</td>
<td>SPREAD II</td>
<td>Central NSW</td>
<td>4.1</td>
<td>60.3</td>
</tr>
</tbody>
</table>

Tab. 2.5: Classification accuracy of SPREAD and SPREAD II using both exact matching and relaxed matching.
2000–11, and the CLUM, which has currency 2004–05. SPREAD allocates much of the cropping to the Darling River flood plain in the north-west of the SLA while SPREAD II allocates it predominantly the south. Investigation revealed that the distribution of cropping shown in the SPREAD map is incorrect and that shown in the SPREAD II map is broadly correct and consistent with the distribution of cropping in the CLUM (Bureau of Rural Sciences, 2004).

It is interesting to note that SPREAD II produced a similar distribution of cropping to that produced by SPREAD when the set of control sites used was restricted to be geographically close to the SLA, as it was when running SPREAD.

![Fig. 2.2: Qualitative comparison for “Cobar”: (a) SPREAD (2000–11), (b) SPREAD II (2000–11), (c) CLUM (2004-05).](image)

In some regions SPREAD II produced very different results to those of SPREAD. An extreme example is shown in Figure 2.3, which shows the SPREAD data, SPREAD II final allocation

---

1 When running SPREAD, for each commodity group, only the three control sites geographically closest to the SLA were used; had SPREAD been run with control sites that better characterised native pasture in the Cobar SLA, a more consistent result may have been observed.
and CLUM data in and around “Esperance” in Western Australia (WA) for 1996–97. The discrepancies between the NLUMs and CLUM data are due primarily to differences in the source data and their interpretation between these two products.

Much of the land classified as conservation in the NLUMs is classified as grazing of natural vegetation in the CLUM data. This discrepancy results from differences in the source data and its interpretation. Land classified as grazing and cropping in the NLUMs is classified almost exclusively as cropping in the CLUM. There are two reasons for this. First, the original CLUM data used a LU category called crop pasture rotations that for purposes of comparison with the NLUMs, has been shown as cropping in the figure. Second, the CLUM in WA was compiled using a method that resulted in biases towards higher valued LUs; relatively coarse spatial regions were mapped and in each region the LU was assigned based on the dominant LU by value of production. This renders the comparison with the CLUM relatively uninformative in this region, but we include it for completeness.

By far the most interesting aspect of this comparison is the different spatial distributions of cropping, grazing of modified pastures and grazing of natural vegetation between SPREAD and SPREAD II. Investigation showed that the NDVI profiles of the control sites in this region were quite different to those of the same crops in other regions and the pattern observed in the SPREAD results is a consequence of it only using information from local control sites. When SPREAD II was run using only local control sites, similar patterns were observed. Discussion with researchers familiar with LU in this region revealed that SPREAD II shows the correct pattern (Richardson, 2013).
Fig. 2.3: Qualitative comparison for “Esperance”. (a) SPREAD (1996-97), (b) SPREAD II (1996-97), (c) CLUM (1996–97).
2.4 SPREAD II Maps for 2005–06 and 2010–11

SPREAD II was used to produce the agricultural components of the NLUMs for the years 2005–06 and 2010–11. These time points correspond to agricultural censuses and hence the ABS publishes the AgStats at the SLA level. We have used the same input data as was used for the production of those maps to explore different aspects of and modifications to SPREAD II. The approach used was broadly the same as that described in Section 2.3, though the NDVI data used and LU classes classified using SPREAD II differ. The geographic coordinate system is the same, the geographic coverage is roughly the same and each map covers the same period relative to each year. The LU classification used for these maps is version 7 of the ALUM classification (Department of Agriculture, Fisheries and Forestry, 2010). For full details on the preparation of the published version of the 2005–06 map, see Bureau of Rural Sciences (2010a) and Bureau of Rural Sciences (2010b).

As described below, the AgStats data available for these periods did not distinguish between irrigated and non-irrigated LUs and hence only the ‘primary’ LU was considered. The classification scheme used for the published maps included both irrigated and non-irrigated agricultural LU classes which were derived via post-hoc adjustments. We only consider the primary classification herein since we are focusing on the performance of SPREAD II, not the overall accuracy of the NLUMs.

It is likely that considering irrigated and non-irrigated crop types together will reduce the overall classification accuracy. While the information in the NDVI profiles may not be able to discriminate between the irrigated instances of the same crop, combining the control site profiles will likely decrease the ability to discriminate either from other LUs.

While it would be natural to use the same regions as those used in the previous work mentioned above, for these more recent periods, we do not, because validation would be problematic. First, the currency of the CLUM data used for the validation varies spatially, and hence regions chosen for analysis have to have relatively recent CLUM data. Second, we wanted to consider regions which had a significant amount of agricultural land and a ‘reasonable’ mix of agricultural LUs. This is not the case for the regions presented above, which were chosen for presentation because they contained interesting aberrant results.

2.4.1 Data

The data used for the production of the maps for the periods considered in this section are broadly the same as those described in Section 2.3.1. Only differences of interest are described here.

2.4.1.1 Agricultural Statistics Data

The AgStats used to produce the 2005–06 and 2010–11 NLUMs were based on the 2005–06 and 2010–11 agricultural censuses respectively. As noted in Section 2.3.1.2, the commodity classes available in the AgStats vary from year to year and must be transformed into standard classes. For the periods considered here, the LU classes available did not allow separation of irrigated
and non-irrigated LUs for classification using SPREAD II. Table 2.6 shows the LU classes used within SPREAD II and the concordance between these classes and ALUM version 7.

### 2.4.1.2 NDVI Data

As with the earlier NLUMs, the NDVI imagery enters the construction of a map via the conditional distributions \( Y_n | C_n \). The profiles are summarised by their mean, range and the time of year of maximum NDVI and the conditional distributions \( Y_n | C_n \) were estimated from pixels with known LU over the same period relative to a calendar year using a three dimensional kernel smoother with a gaussian kernel.

For the periods considered in this section, leave-one-out Cross Validation (CV) (Stone, 1974) was used to select the bandwidths used in the kernel smoother. The primary reason for choosing leave-one-out CV in over k-fold CV is the imbalance between the number of control sites available for each LU. Three optimisation procedures were tried:

1. the Nelder-Mead algorithm (Nelder and Mead, 1965),
2. the Self-Organising Migrating Algorithm of Zelinka (2004), and
3. the charged system search of Kaveh and Talatahari (2010).

The first was used via the R function “optim” in package “stats”, the second via the function “soma” in package “soma”, and the third was a bespoke implementation in C++. The latter two consistently outperformed the former and discovered broadly consistent optima.

The CV was performed giving equal weight to all LU classes and across all LU classes simultaneously. However, since we consider each region separately and hence are only concerned with discriminating between LUs which occur within that region, CV could be performed separately within each region to produce a classifier which is optimal for separating only the relevant set of LUs. A further modification would be to include weights in the objective function that reflect the area of each LU occurring within the region, or perhaps some other criteria reflecting the costs (benefits) of incorrect (correct) classification of each LU class. Such values might be based on the economic value of the commodity in question or the importance of the LU in some decision theoretic framework.

### 2.4.1.3 Control Site Data

Control sites were selected from the same database of ground control points compiled for the production of the 1996–97 SPREAD map. Some LUs and hence their control sites which were not included in the LU classification submitted to SPREAD II in production of the earlier NLUMs could be included under the coarser LU classification used for the periods considered here. As mentioned in Section 2.3.1.4, the control site data available is minimal for many LUs and the spatial distribution is patchy for all LUs. For the periods considered in this section, the control site data are also very old and the prevailing climate is quite different from when the control site data were collected. This is likely to lead to poorer and perhaps (more) biased classification results.
Figure 2.4 shows the NDVI profiles of the control site used for creating the NLUM for the years 2005–06 and 2010–11. Visual inspection of the profiles suggests that many of these LU classes will be hard to separate even within a given year. The confusion matrix shown in Figure 2.5 for the classifier used here confirms this, showing a percentage match rate of 28 percent and many LU classes being classified as some other LU class more often than they are classified as themselves.

In Figure 2.4 it is hard to discriminate between the different years. Figure 2.6 shows the means of these profiles for each year. It is clear in this plot that there are marked differences in the mean profile for some crops between years, some of which will affect the metrics we have chosen. However, when we run the CV to determine the match percentages for each year separately, based on the same bandwidths, we get percent match rates of 27 percent, 17 percent, 23 percent and 31 percent for the years 1996–97, 1997–98, 1998–99 and 1999–00 respectively (an average of 24 percent). This implies that it is better to use all available control sites from the current database rather than only those from a specific year, even for those years represented in that database. While this is a surprising result, we have not explored it further here.
Fig. 2.5: Confusion matrix based on the NDVI profiles of the control sites used throughout this thesis.

2.4.1.4 Running SPREAD II

SPREAD II was run using the same settings described in Section 2.3.2, with some minor differences. In the published maps sub-regional constraints were applied for cultivation and horticulture (see Section 2.3.1.5). Here we do not apply these constraints but consider them separately and in detail in Chapter 3. We also add another post-classification step in addition to Algorithm 1, which further maximises the expected value of the final map with respect to the posterior. The full post-classification procedure is described by Algorithm 2.

Algorithm 2 Extended Final allocation Algorithm

1: Apply Algorithm 1.
2: Iterate over all pixels \((i)\) in the region and for each pixel iterate over all other pixels \((j)\) in the region. For each pair of pixels, if \(Pr(C_{ik}) + Pr(C_{jl}) > Pr(C_{il}) + Pr(C_{jk})\), where \(k\) and \(l\) are the LUs assigned to pixels \(i\) and \(j\) in step 1 or a previous iteration of this step, then swap the LUs of the pixels.
3: Repeat step 2 until no swaps occur.

2.4.2 Software

The process of running SPREAD II using the software described in Section 2.3.3 was cumbersome. It was difficult to coordinate the end-to-end process as much of the pre-processing using ARC/INFO was manual and managing the plethora of input and output files was error prone.
Fig. 2.6: Means temporal profiles of the NDVI data used throughout this thesis.
and time consuming.

While R is well suited to statistical analysis, it was not particularly well suited to the data manipulations required for preparing the data for SPREAD II. Many of the regions are quite large and because of R’s pass by value semantics, it was quite slow and memory use became an issue.

For these reasons, most of the system was re-written in C++. The application now works directly with the spatial inputs and writes the outputs directly to ASCII grids, avoiding many of the fiddly pre- and post-processing steps which were both error prone and time consuming. C++ proved to be far more efficient in terms of both memory use and performance and far better suited to the looping and aggregation involved in these pre- and post-processing steps. The memory footprint was reduced by around 60 percent and the pre-processing time was halved. ARC/INFO is still used in the preparation of the inputs and for the published maps, in some of the post-processing steps and for combining the agricultural layer with the other layers in the final construction of a map.

2.4.3 Results

Accuracy is again assessed against CLUM data. In this section we only present results for “Murrumbidgee (A)”. Results for the other regions are presented in Chapter 6. We only consider the LU classes directly mapped by SPREAD II. Here we have not applied the sub-regional constraints described in Section 2.2, so the confounding between the results and the CLUM noted in Section 2.3.4.1 does not occur here. As in Section 2.3.4.1, we base our comparisons on pixel counts rather than physical areas. A pixel is awarded a match based on the concordance shown in Table 2.6 either if it matches the ALUM code or if the CLUM data has one of the additional ALUM codes.

We do not make the distinction between exact and relaxed matches used in Section 2.3.4.1 because of the lack of irrigation status in the data published in the AgStats and the nature of the data present in the CLUM for the regions we are considering. The reasons for allowing matches for the classes given in the ‘additional ALUM codes’ of Table 2.6 were identified through discussion with Robert Smart and are as follows:

- The AgStats data used in this work does not allow separation of irrigated and non-irrigated areas, so these cannot be treated separately in SPREAD II. Hence, a class matches either its irrigated or non-irrigated ALUM code.

- There are several issues relating to pasture due to the way in which grazing is reported in the AgStats. First, the 2005–06 census only reports areas of grazing and the 2010-11 census reports grazing on improved pastures and grazing on other land, though it is not clear how farmers would interpret these categories. However, to concord with the ALUM, we would ideally map native grazing excluding native-exotic pasture mosaics (aligning with ALUM code 2.1.0) and modified grazing, including native-exotic pasture mosaics (aligning with ALUM codes 3.2.0 or 4.2.0). In production of the 2005–06 and 2010–11 NLUMs, additional datasets were used in post-classification procedures to allocate land to the appropriate class. In the context of the current work (evaluating the accuracy of SPREAD II), we felt it better to simply group all grazing together.
Tab. 2.6: Condordance between LU classes mapped in SPREAD II and ALUM version 7 classes used for validation.

- The AgStats does not distinguish between forestry types, so all forestry types have been aggregated in the CLUM.

- In the regions considered herein, there is a significant area of land classified as cropping in the validation data. Generally, in those same regions there is very little land allocated to the tertiary level codes or one tertiary code dominates, so not much is lost by classifying to catcropping. Conversely, a lot of pixels would be excluded from the validation by retaining the tertiary level.

The one region where the aggregation will affect the validation results is “Horsham (RC) Bal”. Here there is a reasonable area of both LUs 331 and 338.

In the current work we have been less restrictive about the currency of the CLUM data as there are few areas with the same currency as our maps, and we also allow validation against data falling in the year either side of the year under consideration. Bearing in mind that the currency of the CLUM data is based on calendar year, the inclusion of 2006 and 2011 validating the 2005 and 2010 maps respectively is reasonable. Further, discussion with ACLUMP members suggests that there has been relatively little LU change in the regions we consider here between the periods we map and the period of the corresponding CLUM data. Only pixels that are allocated by SPREAD II and have a corresponding LU in the CLUM are considered, so some pixels allocated by SPREAD II will not be included in the validation. There is some mismatch in the pixels that are agricultural land between the NLUM and CLUM, but these mismatches

---

2 One could argue that 2006 and 2011 might, in fact, be better years to choose given that the NLUM reference years are nominally 1/4/05 to 31/3/06 and 1/4/10 to 31/3/11 and the AgStats reference years are 1/7/05 to 30/6/06 and 1/7/10 to 30/6/11.
are introduced in the construction of the agricultural mask, which is outside the scope of SPREAD II.

2.4.3.1 Quantitative Comparison to CLUM

It is clear from Figure 2.7 and Table 2.8 that including the AgStats data makes a dramatic difference to the final classification. The application of Algorithm 1 reduces the percent match rate, which is most likely due to the dominance of pasture at 80 percent of the area of agricultural land in the CLUM data and 75 percent in the AgStats data. Hence, allocating more pixels to pasture will increase the percent match rate (if one allocated all pixels to pasture then a match rate of 80 percent would be achieved). This is discussed further in Section 2.5.

When the AgStats data are used to form a prior (panel e), we get an extremely poor result. Similarly to the results we saw for “Esperance” in Section 2.3.4.2, the regions allocated to pasture and other cropping are almost switched compared to the CLUM data. Comparison of the NDVI profiles in these regions with those in the control site database, revealed that those
2. Spatial REallocation of Aggregated Data Version II

<table>
<thead>
<tr>
<th>Land Use</th>
<th>LU code</th>
<th>AgStats area</th>
<th>CLUM area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pasture</td>
<td>210</td>
<td>203280</td>
<td>237863</td>
</tr>
<tr>
<td>Agroforestry</td>
<td>310</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>Other cropping</td>
<td>330</td>
<td>70091</td>
<td>58229</td>
</tr>
<tr>
<td>Hay &amp; silage</td>
<td>333</td>
<td>2566</td>
<td></td>
</tr>
<tr>
<td>Oil seeds</td>
<td>334</td>
<td>1964</td>
<td></td>
</tr>
<tr>
<td>Cotton</td>
<td>336</td>
<td>0</td>
<td>2140</td>
</tr>
<tr>
<td>Tree fruits</td>
<td>341</td>
<td>568</td>
<td>509</td>
</tr>
<tr>
<td>Tree nuts</td>
<td>343</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>Citrus</td>
<td>348</td>
<td>279</td>
<td></td>
</tr>
<tr>
<td>Grapes</td>
<td>349</td>
<td>328</td>
<td></td>
</tr>
<tr>
<td>Seasonal vegetables &amp; herbs</td>
<td>354</td>
<td>233</td>
<td>813</td>
</tr>
<tr>
<td>Rice</td>
<td>439</td>
<td>20443</td>
<td></td>
</tr>
</tbody>
</table>

Tab. 2.7: LU areas reported in the AgStats and validation data for region “Murrumbidgee (A)”.

<table>
<thead>
<tr>
<th>Description</th>
<th>Match (%)</th>
<th>Average Posterior (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Raw classifier: unconstrained</td>
<td>5.19</td>
<td>11.81</td>
</tr>
<tr>
<td>(b) Raw classifier: constrained</td>
<td>57.17</td>
<td>11.81</td>
</tr>
<tr>
<td>(c) Raw classifier: AgStats as priors, unconstrained</td>
<td>20.05</td>
<td>36.31</td>
</tr>
<tr>
<td>(d) Raw classifier: AgStats as priors, constrained</td>
<td>54.67</td>
<td>36.31</td>
</tr>
<tr>
<td>(e) SPREAD II: unconstrained</td>
<td>72.31</td>
<td>57.11</td>
</tr>
<tr>
<td>(f) SPREAD II: constrained</td>
<td>58.15</td>
<td>57.11</td>
</tr>
<tr>
<td>(g) SPREAD II: AgStats as priors, unconstrained</td>
<td>72.04</td>
<td>57.32</td>
</tr>
<tr>
<td>(h) SPREAD II: AgStats as priors, constrained</td>
<td>57.95</td>
<td>57.32</td>
</tr>
</tbody>
</table>

Tab. 2.8: Classification accuracy of SPREAD II and alternatives.

of the pixels labelled as other cropping were indeed generally more similar to those of pasture in the database, and vice versa.

When Algorithm 1 is applied, the final classification produced by the raw classifier and by SPREAD II are quite similar, both in terms of the number of pixels matched and the spatial distribution of the LUs. The areas classified to each LU are, of course, very similar to the areas given in Table 2.7, with pasture and cereals dominating, and rice occurring along the northern edge of the region. From Table 2.8 we can see that the average posterior of the true class is much larger for the SPREAD II map than for the raw classifier.

Comparison of match percentages shown in Table 2.8 suggests that SPREAD II is not contributing significantly to the categorical classification within this region if the AgStats data are used: regardless of the manner in which they are used. However, the posterior distribution is very different to the raw classifier in terms of expected match rate even when the priors for the raw classifier are based on the AgStats. The question remains as to whether SPREAD II improves the classification accuracy compared to a random allocation of LUs based on the areas reported in the AgStats. We can calculate the expected proportion of pixels that will be correctly allocated exactly as

\[
\bar{p} = \frac{\sum_{k=1}^{K} N_{kc} N_{ka}}{N^2},
\]

where \(N_{kc}\) is the number of pixels of land use \(k\) in the region according to the CLUM and \(N_{ka}\) is the number of pixels of land use \(k\) in the region according to the AgStats. As we are dealing with a large number of pixels, the mean number that get correctly allocated will be approximately normal and we can use a binomial approximation to calculate its variance as
Using this approximation, the expected match rate is 58.37 percent with a standard deviation of 0.91 percent. The expected match rate based on the posterior using SPREAD II and applying Algorithm 1 (which allocates the same number of pixels to each LU as would be done by the random allocation respecting the AgStats data) is within two standard deviations of this, suggesting that SPREAD II is not improving the classification given the AgStats, but is not reducing it either.

\[ N\bar{p}(1 - \bar{p}) \]

2.4.3.2 Burn-In

As SPREAD II is an MCMC based algorithm, we need to allow for a burn-in period. In SPREAD II the posterior estimate of the probability that a pixel has a given LU is based on the proportion of times it was classified as having that LU. Figure 2.8 shows this average against the number of iterations for a randomly selected subset of pixels from this region. These plots do not include the burn-in period. It appears from this figure that the algorithm converges fairly quickly and that a burn-in of 1,000 iterations followed by a run of 2,000 iterations will produce reliable posterior estimates. These are the numbers of iterations used to produce the results above.
Fig. 2.8: Proportion of times a pixel has been allocated each land use for six randomly selected pixels through the running of SPREAD II.
2.5 Simulation Study

The sparsity of the control sites available for this study, the variation in the LU areas we observe within regions, and the nature of our validation data makes it hard to assess classifiers, compare them or understand what factors affect their performance. In this section we develop synthetic datasets over which we have complete control and assess the performance of and robustness of SPREAD II. We do this by varying the amount of noise contained in the remotely sensed imagery and introducing biases into the areal constraints data.

As described in Section 2.3.1.4, the database of control site data used in these studies was compiled in the late 1990s, and all suitable control sites in the database are used in every region at every time point. Climatic conditions, planting times, management practices, phenology and potentially many other factors influence the spectral signature of a given LC (and hence LU). Over an area the size of Australia these differences are likely to be important for at least some LUs. Ideally the control sites used for training a classifier for a particular region of a map would be collected in the same period for which the map is being produced and from the same geographic region. Such a database of control sites was not available for this work and the results are likely to suffer because of it.

It is clear from Section 2.4.1.3 that the database of profiles available for this study is inadequate for separating between the classes used in the classification scheme. This is likely at least partly due to the reasons noted above (the database itself is old, collected over four years and covers the whole continent), but also to the coarse scale of the imagery relative to many of the LU classes included in the classification scheme. Many of these LUs occur at resolutions well below that of the satellite imagery; hence, the imagery and control sites are dominated by mixed pixels for most LUs. This affects both the ability of the imagery to discriminate between LU classes and the interpretability of the validation; how do we interpret a match when we know that the pixel is not pure?

Even with finer scale imagery, the classification scheme used here would likely be ambitious. The profiles shown in Figure 2.4 and confusion matrix in Figure 2.5 show that the combination of satellite data and classifiers used herein are inadequate for discriminating between the LUs of ALUM version 7.

The CLUM data are produced from polygons that do not cleanly partition over our 1km resolution raster. For the validation performed herein, the LU for each pixel is chosen based on the largest overlap with a polygon. Alternatives would be to downsample the NLUM pixels or consider the intersections of the CLUM polygons with the pixels. Clearly the method used to compare the classifications will have a potentially strong bearing on any measure of accuracy and how it can be interpreted.

Third, the CLUM itself has a nominal accuracy of 80 percent (Australian Bureau of Agricultural and Resource Economics, 2011).

From the results presented in Section 2.4.3 it is clear that incorporating the AgStats data into the classification process, whether it be via SPREAD II, applying the post-classification to the raw classifier or using them to form pixel-level priors, improves the final classification significantly. In order to gain some understanding of how much the AgStats data contributes to the classification process we perform several analyses in which we vary the signal-to-noise ratio
of synthetic satellite imagery, the relative areas of the different LU classes within the region, and the LU areas reported in the areal constraints (i.e. introduce a bias into the AgStats). The performance of the various classifiers is assessed based on percent match rates and the probability of observing the true LU averaged across all pixels in the map.

Each of these analyses compare the performance of:

1. the raw classifier without the post-classification procedure (i.e. the kernel density smoother alone),
2. the raw classifier with the post-classification procedure (Algorithm 1),
3. the raw classifier using the AgStats as priors without the post-classification procedure,
4. the raw classifier using the AgStats as priors with the post-classification procedure,
5. the posterior generated by SPREAD II without the post-classification procedure, and
6. the posterior generated by SPREAD II with the post-classification procedure.

The profiles are generated from a set of hypothetical end members that are created from sinusoidal and constant functions. We use 10 LU classes in this study, as this is typical of the number of LUs in many regions in practice. The end members are shown in Figure 2.9.

The synthetic profiles for pixels of LU type \( k \) are generated by mixing each (temporal) component of the end member with a random normal variate using

\[
y^*_{nt} = \alpha y_{t(k)} + (1 - \alpha)\phi(\mu, \sigma^2),
\]

(2.10)

where the parameters \( \alpha, \mu \) and \( \sigma^2 \) are constant for all land use classes, time points and pixels. \( y^*_{nt} \) is the value of the synthetic profile at time \( t \), \( y_{t(k)} \) is the value of the end member for LU \( k \) at time \( t \) and \( \alpha \) is the mixing coefficient that determines how much noise is mixed with the end member. We vary \( \alpha \) between zero and one in each simulation. The end members are the same for all scenarios, but the synthetic profiles are generated separately for each.

![Fig. 2.9: Synthetic end members used for generating synthetic NDVI data in simulation study.](image)
The simulation is run on a ‘region’ of 20×20 pixels. The spatial configuration of the pixels is irrelevant in these analyses as the classifiers do not take any account of spatial context. We use 10 LUs and have 40 control sites for each LU class.

Four analyses follow in which the total area of the each LU class is varied within the region (that is, the total area of pixels with each LU), in the areal constraints or both. The overall goal of this set of analyses is to understand how well SPREAD II and the post-classification procedure work, relative to the raw classifier on which it is built, in the presence of biases in the areal constraint data. In all analyses, the raw classifier without the post-classification procedure should (and does) perform identically. There are minor differences in the results because the synthetic profiles are generated separately in each analysis. For the raw classifier without the post-classification procedure, when \( \alpha \) is zero the number of LUs fully determines the asymptotic (with respect to the number of pixels and control sites) accuracy; the proportion of pixels of each LU does not matter. In the analyses presented here, the expected hit rate with \( \alpha \) equal to zero is 10 percent. As the profiles become more informative, we expect the hit rate to converge to 100 percent.

The top panel of Figure 2.10 shows how the classification accuracy changes based on hit rates when each LU has the same number of pixels and the areal constraint data accurately reflect this. Clearly there is little difference between the classification accuracies. The values when \( \alpha \) equals zero are as we would expect: there is no information in the profiles and, hence, the allocation is completely random. With 10 LU classes we expect that one allocation in ten will be correct giving a hit-rate of 10 percent. As \( \alpha \) approaches 1 we expect all variants to converge to a hit rate of 100 percent, as the classification based on the profiles approaches perfection and the areal constraint data agrees with actual number of pixels of each LU.

In the second panel, 50 percent of the pixels in the region have LU 1, 10 percent of have LU 2 and each of the remaining LUs have 5 percent of the pixels. The areas reported in the areal constraint data accurately reflect this. Here the methods that incorporate the AgStats are clearly performing better when \( \alpha \) is small. The results when \( \alpha \) equals zero are as we expect:

- The raw classifier randomly assigns a LU to each pixel and as described above and the expected hit rate is 10 percent.
- Without the application of the post-classification procedure, SPREAD II allocates nearly all pixels to the most abundant (LU 1). Since half of those are LU 1, it gets about 50 percent correct.
- When the post-classification procedure is applied to either the raw classifier or SPREAD II, 50 percent of the pixels get set to LU 1 and have a 50 percent chance of being correct, and each of the other 50 percent have a \( 100 \times 0.5/9 \) percent chance of being correct, giving a hit rate of 27.7 percent.

Once again, as \( \alpha \) approaches 1 we expect all variants to converge to a hit rate of 100 percent since the areal constraint data agrees with the actual numbers of pixels.

In the third panel the numbers of pixels of each LU are the same as in the second. The areal constraint data, however, is biased: all LUs are reported to have the same area. In this case, as we would expect, the scenario that does not incorporate the areal constraint data performs
Fig. 2.10: Percent correctly matched for the trialled classifiers for profiles generated using (2.10) for various values of $\alpha$. Top panel: the total area of pixels with each LU is identical and this is reflected in the areal constraints. Second panel: the total area of pixels with each LU varies but is accurately reflected in the areal constraints. Third panel: the total area of pixels of each LU is the same as in the second but all LUs are reported to have the same area in the areal constraints. Fourth panel: the total area of pixels with each LU is identical but LU 1 is reported to have half the total area and the remaining LUs have equal area in the areal constraints.
best. When $\alpha$ equals zero we expect a hit rate of approximately 10 percent for all variants, as every pixel has a 10 percent chance of being allocated the correct LU based on the areal constraint data. As $\alpha$ approaches 1, SPREAD II without the post classification procedure also approaches 1, but more slowly, since as $\alpha$ approaches 1, the raw classifier approaches perfection and dominates the areal constraints.

In the fourth panel all LUs have the same number of pixels, but in the areal constraint data LU 1 is reported to have half the total area and the remaining LUs have equal area. Again, as we would expect, the scenario that does not incorporate the areal constraint data performs best. When $\alpha$ equals zero we still expect a hit rate of 10 percent. For the scenarios that include the areal constraint data (that is, all scenarios except for the raw classifier without the post-classification procedure), 10 percent of the pixels (those with LU 1) have a 50 percent chance of being correct and the other 90 percent of pixels have a $100 \times 0.5/9$ percent chance of being correct. In this analysis, the scenario for SPREAD II with the post-classification procedure performs worst. In this scenario, the use of the constraints while running SPREAD II increases the probability of LU 1 significantly so that when the post-classification procedure is applied and the final map produced, some pixels get allocated to LUs other than 1 or their true land use, reducing the hit rate to below that of the raw classifier with the post-classification procedure.

Figure 2.11 shows the probability of observing the true LU averaged across all pixels from the same analysis. The results are consistent with those presented for the match percentages, but the differences between SPREAD II and the raw classifier are not so pronounced. One interesting difference between these results, however, is the relative similarity of the bottom two panels in Figure 2.11 compared to Figure 2.10. This similarity implies more of a symmetry between results coming from the ‘opposing’ biases present in the areal constraint data and demonstrates the strength of the constraint enforced by Algorithm 1. When the areas of the LUs are the same, we don’t see any improvement from the AgStats data.

2.6 Discussion

The analyses presented in this chapter show that SPREAD II can outperform the raw classifier on which it is built under certain conditions; specifically, if the areas of different LUs differ within a region and the AgStats data are not biased. When these conditions do not hold, SPREAD II can perform poorly compared to the raw classifier, particularly if the post-classification procedure is used. This implies that to justify the use of SPREAD II, some confidence in the areal constraint data is required. However, in most regions there will be multiple LUs, their areas will vary, and biases in the order of those present in the pathological examples used in the simulation study are unlikely to occur in practice and can be checked for given the existence of data such as the CLUM. Our results also suggest that vanilla SPREAD II performs at least as well as SPREAD, if not marginally better.

In the region considered in Section 2.4.3, the simple Bayesian approach produced a pathological result. We will see later that the simple Bayesian approach tends to perform as well or better than all other approaches considered in this thesis across our test regions. This region presented here appears to be unusual in this regard. In our simulation study, the simple Bayesian approach appeared to be more robust to biases in the areal constraints data and performed well compared to all other approaches in most cases. The simple Bayesian approach is also much faster to
Fig. 2.11: Probability of observing the true LU averaged across all pixels for the trialled classifiers for profiles generated using (2.10) and various values of $\alpha$. 
compute and requires significantly less computer memory, meaning it is able to handle larger regions and/or higher resolution maps.

Both here and when producing version 3 of the NLUMs, we used a constant variance in the normal distributions used to implement the constraints (that is, the variance was the same for all LU classes). This is justified by the estimates of the areas of each LU being based on agricultural censuses and thus, theoretically at least, being free of error. When we used constraint variances equal to $\sigma_k^2 = A\pi_k(1 - \pi_k)$ (panels g and h of Figure 2.7), we get almost identical maps to the those produced using $\sigma_k = 200$. This is not surprising when one considers that in this region, where there are around 3,500 pixels, the maximum possible standard deviation, corresponding to a LU with a reported area equal to half the total area of agricultural land within the region, is approximately 295, which is relatively close to 200 considering the number of pixels we are dealing with. This choice of variance structure was made based on consideration of the multinomial structure of the simple Bayesian model and for comparative purposes. In practice, the variances used in the constraints should reflect uncertainties in the estimates of the areal totals.

Whichever analysis is used, the results presented in this chapter demonstrate that the use of areal constraints data is of great value. Many LUs will result in similar or identical LCs and hence the extent of a given LU within a region cannot be determined using remotely sensed data alone. Of course, if LUs with similar or identical LCs occur within a region, then the spatial distribution of these will be hard if not impossible to determine using only the methods and data considered here, and the posterior distributions for pixels within the region will be roughly proportional to the areas stated in the areal constraints data. Here, due to mismatches in the LU classifications used by the ABS when producing the AgStats and those of the ALUM, many of the LUs which have similar NDVI profiles are combined into joint validation classes and this confusion is hidden. We see, however, in some of the less aggregated classes, which generally occur with lower frequency in the region considered here, that many classes are poorly identified by all methods considered. The set of classes mapped is ambitious to say the least and many of the classes occur at spatial scales well below that of the 1km resolution imagery which has been used here. Hopefully the work of subsequent chapters can resolve some of this remaining confusion.
3. INCLUDING SUB-REGIONAL CONSTRAINTS IN SPREAD II

3.1 Introduction

As mentioned in Section 1.1.2, many studies have used areal constraint data to aid in LU and/or LC mapping. Such constraints were the basis of the techniques explored in Chapter 2, where the total areas allocated to each (agricultural) LU were constrained to be close to the areas reported in the areal constraint (AgStats) data. In this chapter we extend this approach to incorporate spatial information describing where within a region given classes of LUs occur. By way of introduction, Smart et al. (2006) used the datasets described in Section 2.3.1.5 data in the production of version 3 of the NLUMs to identify zones where irrigation and horticulture occur with specified intensities.

We refer to subregions where such data exist as “zones” to distinguish them from “regions”, which is used as previously to refer to areas over which areal constraint data is available. Similarly, a constraint over a zone is referred to as a “zonal constraint”. An example of a zone can be seen in panel (h) of Figure 3.1. A label such as “irrigated” is referred to as a “super type”. While we only consider estimates based on digital datasets herein, estimates could also be derived from, for example, expert opinion, the location of infrastructure or environmental assets, cadastral data or any other data that can inform where LUs of a given super type occur, provided that data can be turned into a spatial extent.

Many techniques for including sub-regions have been explored previously. Most techniques stratify a region into subregions in which different classification techniques are applied. Strahler et al. (1978) use elevation to stratify a region into elevation zones, then use prior probabilities based on observed frequencies of each species type within each elevation zone to aid in the classification based on remotely sensed data. Manandhar et al. (2009) use different post-classification corrections in different zones to refine an initial classification, based on the LU change classes and corresponding misclassification errors that occur in the zones. Judex et al. (2006) use sub-regions defined by elevation data two ways, one of them being a post-classification procedure and the other augmenting the feature vector used in the primary classifier. Stefanov et al. (2001) use digital city and reservation boundaries to assist LC classification by noting what types of LC occur within those boundaries via an expert based system. Stuckens et al. (2000) use urban/nonurban masks to post-classify pixels which were out of context; e.g. classified as high-density urban but were outside of an urban zone according to the mask. You et al. (2009) and You and Wood (2006, 2005) present a method of disaggregating agricultural statistics and various refinements thereof, based on choosing the fractions of “production systems” which are used for various LUs, where production systems are distinguished by the technology/cultivar mixed used for agricultural production and are expressed as a spatial extent.

The methods presented in this chapter incorporate independent data sources that contain new
information that is independent of the other data used in vanilla SPREAD II. As such, if these
data are reliable, we expect that these methods will improve overall classification performance
regardless of the performance of the underlying classifier.

3.1 Possible Approaches

There are at least three methods of including data of this form in SPREAD II:

1. modifying the model for $[Y_i|C_i = k, ...]$, 
2. adjusting of the priors ($\Pi$) for pixels within a zone and/or its complement, and
3. introducing a constraint over the area of land within a zone that is allocated to LUs with
   the super type associated with the zone.

Adjusting $[Y_i|C_i = k, ...]$ was unsuccessfully trialled, when producing version 3 of the NLUMs,
as a way of dealing with the 'salt and pepper' patterns of LUs observed in the results of
SPREAD II in some regions. This approach avoids the computational complexities of adjusting
the priors as it does not affect $[T]$, but has no conceptual or mathematical justification.

Adjusting the priors ($\Pi$) seems a natural choice for the case of a single zonal constraint. In
this case one could modify the priors such that

$$\frac{\sum_{n \in P^*} \sum_{k \in K^*} \pi_{nk}}{\sum_{n \in P^*} \sum_{k \in \bar{K}^*} \pi_{nk}} = c,$$

(3.1)

for pixels inside the zone for some appropriate constant $c$. Here, $K^*$ and $\bar{K}^*$ are the sets of
LUs of the super type pertaining to the zone and its complement respectively and $P^*$ is the
set of pixels within the zone. A natural choice for $k$ is the ratio of the expected area of land
inside the zone which is allocated to LUs in the set $K^*$ and the expected area of land inside
the zone, which is allocated to LUs in the set $\bar{K}^*$. While this approach is reasonable for a
single zone, it is not clear how one would proceed if there are multiple zones, especially if they
overlap spatially. Forming the pixel priors this way implies that the priors for a given LU vary
between pixels inside and outside the constraint and hence that the distribution counts of pixels
of each LU is no longer multinomial. Consequently, the simplification that makes (2.5) easy to
sample from no longer holds. This is not an issue using the “simple Bayesian” method described
in Section 2.2.3 and approximations introduced in Chapter 5 could be exploited when using
SPREAD II, but this is not explored here.

The method introduced here of forming constraints over the area of land allocated to LUs with
a common super type is very similar to that used for the constraints based approach used in
vanilla SPREAD II. One of the main advantages is that it avoids the computational issues
associated with modifying $\Pi$. This method was used when producing version 3 of the NLUMs.
However, the method was then based on a heuristic and the posterior did not fully take account
of the assumed prior. The current chapter formalises the inclusion of such constraints and notes
issues that need to be considered in practice.

Section 3.2 presents the theory used for including a single zonal constraint. It also introduces
the theory for the inclusion of multiple zonal constraints, but stops short of a full solution due to
difficulties in calculating the required expressions. Section 3.3 contrasts the results with those presented for vanilla SPREAD II and other ‘simpler’, less computationally intensive techniques that incorporate the same data. In that section we also trial an approximation in the single zonal constraint case that ignores the additional terms analogous to those that are difficult to calculate in the multiple zonal constraint case. This is similar to the approach used when making version 3 of the NLUM, except there we applied constraints both inside and outside the zone. We trial this approach in order to get some indication of whether the inclusion of those terms is important from a practical perspective. Section 3.4 summarises and discusses our results and presents some caveats.

3. Methods

In the current work it is assumed that there are available data describing where within a region LUs with a specific super type occur with a given intensity. These data are used to form a binary partition of the region into two zones, $\alpha$ and $\bar{\alpha}$. For example, if we have a dataset containing polygons that describe where irrigation is known to occur with a given intensity, we can partition the region into an “irrigation zone” and a “non-irrigation zone”. Every LU would be assigned a super type of either “irrigated” or “non-irrigated”. Note that a pixel with super type $\alpha$ can occur both inside and outside the zone $\alpha$ and vice versa.

3.2.1 A Single Zonal Constraint

We first consider the case where we have a single zonal constraint and corresponding zones of types $\alpha$ and $\bar{\alpha}$. Following the same development as for (2.5), we can express and decompose the conditional posterior distribution for a pixel as

$$
[C_i, T, S, L|C_{-i}, T_{-i}, S_{-i}, L_{-i}, Y, \Theta, \Pi, \Omega] 
\propto [Y_i|C_i][C|T, S, L, \Theta, \Pi, \Omega][T|S, L|\Theta, \Pi, \Omega] 
\propto [Y_i|C_i][C|T, S, L, \Theta, \Pi, \Omega][T|S, L|\Theta, \Pi, \Omega] 
\propto [Y_i|C_i][C|T, S, L, \Theta, \Pi, \Omega][T|S, L|\Theta, \Pi, \Omega] 
\propto [Y_i|C_i][C|T, S, L, \Theta, \Pi, \Omega][T|S, L|\Theta, \Pi, \Omega] 
\propto [Y_i|C_i][C|T, S, L, \Theta, \Pi, \Omega][T|S, L|\Theta, \Pi, \Omega] 
\propto [Y_i|C_i][C|T, S, L, \Theta, \Pi, \Omega][T|S, L|\Theta, \Pi, \Omega] 
\propto [Y_i|C_i][C|T, S, L, \Theta, \Pi, \Omega][T|S, L|\Theta, \Pi, \Omega] 
\propto [Y_i|C_i][C|T, S, L, \Theta, \Pi, \Omega][T|S, L|\Theta, \Pi, \Omega]
$$

(3.2)

where $\Omega = \{\alpha \mu_\alpha, \alpha \sigma_\alpha^2\}$ contains the mean ($\alpha \mu_\alpha$) and variance ($\alpha \sigma_\alpha^2$) of the area within the zone that is used for LUs of super type $\alpha$, $S_b^a$ is the number of pixels with LU super type $b$ in zone $a$, $S_b^a$ is the number of pixels with LU super type $b$ in the entire region, and $L$ is the total number of pixels with LU super type $b$ in the entire region, and $L = \{L_b^a\}$. $\Pi$ and $\Theta$ are defined as they were in Chapter 2.

The denominator of (3.2) can be written as

$$
[T, S, L|\Pi] = [S|T, L, \Pi][T|\Pi],
$$

(3.3)

since $L$ is degenerate given $T$. $[T|\Pi]$ is the same as in Chapter 2.

In the development that follows we require that $\pi_{ak} = 1/K$ and drop the term $\Pi$. Since $[S|S_b^a, L, T]$ is degenerate, within zone $a$ $[S|T, L]$ is hypergeometric with probability mass func-
When evaluating this for a pixel we must take into account its initial LU (i.e. the LU it was given in the previous iteration), $\tilde{k}$. Without loss of generality, let LU $\tilde{k}$ have super type $b$. Note that we can work with either $L_b$ and $S_{ba}$ or $L_\bar{b}$ and $S_{\bar{b}a}$ due to the degeneracy noted above.

Dividing (3.4) by its value when the pixel has this initial LU and evaluating for LU $k$ gives

$$[S|T, L] \propto \begin{cases} 1 & \text{if the super type of } k \text{ is } b \\ L^b(S^b_{a} + 1) & \text{otherwise.} \end{cases}$$

(3.5)

The form of (3.5) is intuitive; $S_{ba}/L_b$ is the proportion of pixels of LU zone type $b$ that falls in zone $a$, so $[S|T, L]$ is proportional to the odds of a randomly selected pixel in zone $a$ having LU super type $b$ relative to LU super type $b$ if the number of pixels with LU super type $b$ was increased by one.

For the zonal constraints we work with areas instead of pixel counts, as described for the regional constraints in Chapter 2. We specify set $a \mu_a = a \alpha A$, where $a \alpha A$ is the area of land inside the constraint and $\alpha$ is the intensity of LUs with super type $a$ within the zone (i.e. the proportion of that land expected to be used for LUs of super type $a$ within the zone). The following three cases need to be considered:

1. the pixel is in zone $a$, the super type of the initial LU of the pixel is $a$ and the super type of LU $k$ is $\bar{a}$,

2. the pixel is in zone $a$, the super type of the initial LU of the pixel is $\bar{a}$ and the super type of LU $k$ is $a$, and

3. the pixel is in zone $\bar{a}$ or the super type of the initial LU of the pixel is the same as the super type of LU $k$.

The combined areal and zonal constraints in (3.2) are given by

$$[T, S, L|\Phi, \Omega] = \begin{cases} \phi \left( \tilde{A}_k + a_n; \mu_k, \sigma_k^2 \right) \phi \left( \alpha \tilde{A}_a - a_n; a \mu_a, a \sigma_a^2 \right) & \text{case 1} \\ \phi \left( \tilde{A}_k + a_n; \mu_k, \sigma_k^2 \right) \phi \left( a \tilde{A}_a + a_n; a \mu_a, a \sigma_a^2 \right) & \text{case 2} \\ \phi \left( \tilde{A}_k + a_n; \mu_k, \sigma_k^2 \right) & \text{case 3.} \end{cases}$$

(3.6)

Here $a \tilde{A}_a$ is the area inside zone $a$ which is allocated to LUs with super type $a$ when the current pixel has its initial LU, $\tilde{k}$. All other quantities are defined in Chapter 2.

Note that the additional constraint is only applied within the zone. One could potentially apply a complementary constraint in zone $\bar{a}$, as was done when producing version 3 of the NLUMs, but this does not seem appropriate. First, the regional constraints already implicitly apply this constraint, and second, knowing what proportion of land within the constrained region is used for LUs of super type $a$ does not imply direct knowledge of the proportion in the complementary zone (though it is implied given the regional areal constraints).
In Chapter 2 we could express the posterior probability of the map given the LUs of all pixels, excluding the one under consideration, in terms of quantities that exclude the current pixel, and hence use expression (2.6) for all LUs. When we include the zone constraints, five distinct cases need to be to considered. These are:

1. the super type of the initial LU of the pixel is the same as the super type of LU \( k \),
2. the pixel is in zone \( a \), its initial LU has super type \( a \) and the super type of LU \( k \) is \( \bar{a} \),
3. the pixel is in zone \( a \), its initial LU has super type \( \bar{a} \) and the super type of LU \( k \) is \( a \),
4. the pixel is in zone \( \bar{a} \), its initial LU has super type \( a \) and the super type of LU \( k \) is \( \bar{a} \), and
5. the pixel is in zone \( \bar{a} \), its initial LU has super type \( \bar{a} \) and the super type of LU \( k \) is \( a \).

The resulting expressions for the posterior distribution are

\[
[C, T, S, L | C_{-i}, T_{-i}, S_{-i}, L_{-i}, Y, \Theta, \Omega] = \begin{cases}
P_k &\text{case 1} \\
P_h \frac{\sum c}{L^*(S^* + 1)}(\sum \sum) &\text{case 2} \\
P_h \frac{\sum c}{L^*(S^* + 1)}(\sum \sum) &\text{case 3} \\
P_h \frac{\sum c}{L^*(S^* + 1)}(\sum \sum) &\text{case 4} \\
P_h \frac{\sum c}{L^*(S^* + 1)}(\sum \sum) &\text{case 5}
\end{cases}
\]

where \( P_k \) is given by (2.6).

### 3.2.2 Multiple Constraints

Equations (3.4) and (3.5) hold for a single zone constraint. The equivalent expressions for multiple zone constraints are much more complicated, as all zones and their intersections need to be taken into account. Consider the case for two zones; let \( c S_{ab}^b \) be the number of pixels with LU super type \( b \) in zone \( a \) in some arrangement of pixels \( c \), the analogue of (3.4) is

\[
|S|T, L = \sum_{c \in C} \left( \frac{L_{ab}^b}{S_{ab}^b} \right) \left( \frac{S_{ab}^b}{c S_{ab}^b} \right) \left( \frac{S_{ab}^b}{c S_{ab}^b} \right) \left( \frac{S_{ab}^b}{c S_{ab}^b} \right)
\]

\[
\propto \sum_{c \in C} \frac{L_{ab}^b c S_{ab}^b}{c S_{ab}^b} \frac{L_{ab}^b c S_{ab}^b}{c S_{ab}^b} \frac{L_{ab}^b c S_{ab}^b}{c S_{ab}^b} \frac{L_{ab}^b c S_{ab}^b}{c S_{ab}^b} \]

\[
(3.8)
\]
where the usual constraints for the multinomial hypergeometric distribution

\[ N_{ab} = c_{ab}^a + c_{ab}^\bar{a} + c_{ab}^{\bar{a}a} + c_{ab}^{\bar{a}\bar{a}} \]

apply and \( C \) is the set of all combinations such that constraints

\[
\begin{align*}
S_a &= c_{ab}^a + c_{ab}^\bar{a} + c_{ab}^{\bar{a}a} + c_{ab}^{\bar{a}\bar{a}} \\
S_b^\bar{a} &= c_{ab}^a + c_{ab}^\bar{a} + c_{ab}^{\bar{a}a} + c_{ab}^{\bar{a}\bar{a}} \\
L^{ab} &= c_{ab}^a + c_{ab}^\bar{a} + c_{ab}^{\bar{a}a} + c_{ab}^{\bar{a}\bar{a}} \\
L^\bar{a}b &= c_{ab}^a + c_{ab}^\bar{a} + c_{ab}^{\bar{a}a} + c_{ab}^{\bar{a}\bar{a}} \\
L^{\bar{a}}\bar{b} &= c_{ab}^a + c_{ab}^\bar{a} + c_{ab}^{\bar{a}a} + c_{ab}^{\bar{a}\bar{a}}
\end{align*}
\] (3.9)

hold. Note that we can ignore the last zone (in this case \( \bar{a}\bar{b} \)) and that the choice of which zone we ignore is arbitrary.

The cardinality of the set \( C \) will generally be large and the constraints (3.9) are difficult to deal with. We have not found a way of approximating (3.8), so including multiple zonal constraints is an open problem. However, our experience is that using multiple constraints, albeit with a mathematically flawed implementation, produces a better map. While we cannot directly assess the importance of including the term \( S[T, L] \) when there are multiple zonal constraints, we can assess its impact in the case of a single constraint. If its omission does not adversely affect the results too dramatically, then it may be reasonable to omit it and apply multiple constraints anyway.

### 3.2.3 A Simple Bayesian Classifier

As mentioned in the introduction, with a single constraint we can adjust the prior probabilities for pixels inside zone \( a \) such that the expected area allocated to LUs with super type \( a \) inside the zone is equal to the areas implied by the zonal constraints. This can be seen as an extension to the simple Bayesian model described in Section 2.2.3. Using equation (3.1) we set the prior for each LU at each pixel to be

\[
\pi_{nk} = \begin{cases} 
\frac{\alpha \mu_k}{\sum_{k \in K^*} \mu_k} & \forall k \in K^*, n \in P^* \\
\frac{(1 - \alpha) \mu_k}{\sum_{k \in \bar{K}^*} \mu_k} & \forall k \in \bar{K}^*, n \in P^*,
\end{cases}
\]

where \( K^* \) and \( \bar{K}^* \) are the sets of LUs with super type \( a \) and \( \bar{a} \) respectively and \( P^* \) is the set of pixels inside zone \( a \).

In some cases the combined area of LUs with super type \( a \) will be too small to fill the proportion
α of the area of zone a. In this case we set

$$\pi_{nk} = \begin{cases} \frac{\mu_k}{\sum_{n \in P^*} a_n} & \forall k \in K^*, n \in P^* \\ \frac{1 - \sum_{k \in K^*} \pi_{nk} \mu_k}{\sum_{k \in \bar{K}^*} \mu_k} & \forall k \in \bar{K}^*, n \in P^*. \end{cases}$$

where $a_n$ is the area of pixel n. We would also like the expected area of each LU over the entire region to be equal to the area specified by the AgStats, or as close to it as possible, so we set

$$\pi_{nk} = \frac{\mu_k - \sum_{n \in P^*} a_n \pi_{nk}^*}{\sum_{k \in K} \mu_k - \sum_{n \in P^*} a_n} \quad \forall n \in \bar{P}^*.$$ 

where $\bar{P}^*$ is the set of pixels outside zone a.

As noted in the discussion at the end of this chapter, this approach has limitations, but it is consistent with the methods used for calculating the areas used by the zonal constraints used in SPREAD II and hence is reasonable for assessing the merits thereof.

### 3.3 Results

We consider the same study region as was used in Chapter 2. The input data is the same, with the addition of the mask which defines the extent of the zone. The only additional parameters are $\Omega$ and $\alpha$. The values of these parameters were chosen when producing the early NLUMs by looking at various horticulture zones and determining what proportion of the area within those zones were used for horticultural purposes and through discussion with people knowledgeable on LU patterns within those regions. This is discussed in more detail in Section 2.3.2. Based on that analysis we used the same value across all regions and continue to use those values here. In reality it is likely that this would vary between regions and ideally one would choose a value specific to each region.

<table>
<thead>
<tr>
<th>Description</th>
<th>Match (%)</th>
<th>Average Posterior (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Raw classifier: AgStats as priors with zonal constraints, unconstrained</td>
<td>63.32</td>
<td>50.95</td>
</tr>
<tr>
<td>(b) Raw classifier: AgStats as priors with zonal constraints, constrained</td>
<td>73.34</td>
<td>50.95</td>
</tr>
<tr>
<td>(c) SPREAD II: zonal constraints, unconstrained</td>
<td>85.21</td>
<td>73.16</td>
</tr>
<tr>
<td>(d) SPREAD II: zonal constraints, constrained</td>
<td>74.29</td>
<td>73.16</td>
</tr>
<tr>
<td>(e) SPREAD II: zonal constraints and no ratio constraints, unconstrained</td>
<td>85.34</td>
<td>73.27</td>
</tr>
<tr>
<td>(f) SPREAD II: zonal constraints and no ratio constraints, constrained</td>
<td>74.57</td>
<td>73.27</td>
</tr>
<tr>
<td>(g) SPREAD II: unconstrained</td>
<td>72.31</td>
<td>57.11</td>
</tr>
</tbody>
</table>

**Tab. 3.1: Classification accuracy of the zonal constraints methods contrasted with SPREAD II.**

Figure 3.1 shows the results for “Murrumbidgee (A)”, the spatial extent of constraint used and the validation data. The spatial distribution of the LU allocations is clearly very different when the zonal constraints are applied, with most of the cultivated LUs being allocated in the eastern half of the region rather than the west, which is in agreement with the zonal constraints. Comparison of Tables 2.8 and 3.1 shows that both the percentage match and average posterior have increased significantly compared to vanilla SPREAD II using an unconstrained final allocation. For both SPREAD II including the zonal constraints and the simple Bayesian approach using an unconstrained final allocation, the percentage match increases in the order of 10 percent. For SPREAD II using an unconstrained final allocation, the average posterior improves in the
order of 16 percent and for the simple Bayesian approach it increases in the order 6 percent.

While these results are positive, and the spatial configuration of LU within the region appears closer to reality (see Figure 3.2), we must be a little circumspect as the spatial extent of the cultivation constraint appears so similar to the extent of cultivated LUs in the CLUM data, that they almost certainly have some common lineage, at least in the region considered here.

As such, the agreement between the two is likely to be high and given that the CLUM data themselves have a nominal spatial accuracy of 80 percent, the accuracy of the maps may be overstated.

It is interesting to note that, rather than adversely affecting the percentage match or average posterior, omission of \([S|T, L]\) actually produces the best, though not significantly different, results.
3.4 Discussion

A zonal constraint is likely to help where its spatial extent covers a moderate to large proportion of the agricultural land within a region. When it covers only a small proportion, it cannot have a significant impact on the overall accuracy of the map. If the constraint covers all the agricultural land, then if all LUs present in the areal constraint data are of the super type corresponding to the constraint, the zonal constraint introduces a constant scale factor to the posterior at every pixel and consequently has no effect. If not all LUs are of the super type corresponding to the constraint, then it acts as a modification to the areal constraint data. In the latter case, one should probably be cautious about including the zonal constraint(s); if the modification is small, it will have little impact and, if it is large, it implies a contradiction between the areal constraint data (i.e. the regional constraints) and the zonal constraint. In this case, further consideration of the accuracy of one or both datasets would be advisable.

In the region presented here the improvement over vanilla SPREAD II is significant. We observe, however, that using the simple Bayesian model presented in Section 3.2.3 we get very similar results in the final classification but that the expected match rate is around 10 percent lower, suggesting the posterior distribution is inferior. One might conclude, therefore, that the simple Bayesian approach, with its reduced computational burden, is adequate for producing a categorical map but, if one is interested in the posterior probabilities themselves, SPREAD II may be a superior technique, depending on one’s confidence in the various data.

One issue with the constraints is completeness. Figure 3.2 shows the cultivation mask used here overlaid on the Google Maps satellite layer extracted in late 2015. It is clear that the cultivated constraint does not include some subregions that are obviously cultivated. One must bear in mind, when considering this image, that the constraint was derived from data with a currency of 2004 while the satellite imagery is likely to have been acquired in 2014 or 2015. On the other hand, the constraint was used in making maps for 2010, so, given the differences we observe here, it is unlikely to have been complete at that time. It appears, however, that the area covered by the constraint is almost all cultivated and hence it is unlikely to do harm, despite being incomplete.

Another issue with the inclusion of zonal constraint(s) is that they may not be consistent with the areal constraint data and/or result in pathological posteriors. For example, in the SLA “Carnamah (S)” in Western Australia, we have the spatial area and the area reported in the AgStats given in Table 3.2. Applying the methods described in Section 3.2.3 using $\alpha$ set to 0.9, we get the expected areas and prior probabilities, rounded to the nearest hectare and five decimal places respectively, shown in Table 3.3.

<table>
<thead>
<tr>
<th>Extent or Statistic</th>
<th>Area (ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All pixels</td>
<td>185,049</td>
</tr>
<tr>
<td>All pixels in cultivated zone</td>
<td>161,689</td>
</tr>
<tr>
<td>Area of LUs with super type of cultivated reported in AgStats</td>
<td>170,920</td>
</tr>
</tbody>
</table>

Tab. 3.2: Total areas of: pixels, pixels in the cultivated zone, and reported in the AgStats for “Carnamah (S)”.

Three issues are apparent in Table 3.3. First, the expected area of LUs with super type $\bar{a}$ is greater than the area reported in the AgStats. Second, the prior probability for LUs with super
Fig. 3.2: Cultivated constraint overlayed on Google Maps: extracted 20/10/2015.

<table>
<thead>
<tr>
<th>LU</th>
<th>Super Type</th>
<th>Expected</th>
<th>Reported</th>
<th>( \pi_k ) Inside Cultivated Zone</th>
<th>( \pi_k ) Outside Cultivated Zone</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>cultivated</td>
<td>69,521</td>
<td>81,656</td>
<td>0.42997</td>
<td>0.51946</td>
</tr>
<tr>
<td>3</td>
<td>cultivated</td>
<td>57,418</td>
<td>67,440</td>
<td>0.35511</td>
<td>0.42902</td>
</tr>
<tr>
<td>6</td>
<td>cultivated</td>
<td>10,166</td>
<td>11,940</td>
<td>0.06287</td>
<td>0.07596</td>
</tr>
<tr>
<td>8</td>
<td>cultivated</td>
<td>7,367</td>
<td>8,653</td>
<td>0.04556</td>
<td>0.05504</td>
</tr>
<tr>
<td>11</td>
<td>cultivated</td>
<td>998</td>
<td>1,172</td>
<td>0.00617</td>
<td>0.00746</td>
</tr>
<tr>
<td>13</td>
<td>cultivated</td>
<td>48</td>
<td>57</td>
<td>0.00030</td>
<td>0.00036</td>
</tr>
<tr>
<td>14</td>
<td>cultivated</td>
<td>3</td>
<td>3</td>
<td>0.00002</td>
<td>0.00002</td>
</tr>
<tr>
<td>0</td>
<td>non-cultivated</td>
<td>16,128</td>
<td>1,4094</td>
<td>0.09975</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>non-cultivated</td>
<td>40</td>
<td>35</td>
<td>0.00025</td>
<td>0</td>
</tr>
</tbody>
</table>

Tab. 3.3: Summary of areas of each LU in “Carnamah (S)”. The expected areas and probabilities are calculated as per Section 3.2.3 using \( \alpha \) set to 0.9.

As noted in Section 3.2.2, correctly including multiple constraints is difficult. Experience in making the NLUMs suggests, however, that no single constraint yields ideal results. We have found that the horticulture constraint is valuable for separating Pasture and Agroforestry, which are the only uncultivated LUs, from the others. Unfortunately, when only the horticulture constraint is applied, irrigation is poorly mapped. The results of our experiment where we omitted \([S|T,L]\) in the case of a single constraint allows us to conjecture that one may be able to ignore that term and include multiple constraints successfully anyway.

We have used fixed variances in the normal distributions that implement the zonal constraints. As with the regional constraints, these should reflect uncertainty in the estimates of the areas to which they apply. The uncertainty here arises from uncertainty about \( \alpha, \Theta \) and any inaccuracies in the data used to form the zonal constraint. As noted previously, in our case \( \Theta \) is derived...
from an agricultural census and hence is, theoretically at least, free of error. The former was
derived from an analysis of several regions and expert opinion, but as with the estimate of $\alpha$
itsel, is likely to vary between regions and is likely to be hard to quantify.
4. INCLUDING SPATIAL CONTEXT IN SPREAD II

4.1 Introduction

In vanilla SPREAD II, each pixel is considered independently. Depending on the resolution of the remotely sensed data used for producing a map, we may expect that pixels that are close spatially may share the same LU, especially in the case of broad acre LUs (Kiiveri and Campbell, 1992), or that some LUs occur in similar regions to others (Cihlar and Jansen, 2001). We also expect that for two pixels of the same LU, the satellite imagery observed will be more similar if they are close spatially due to similarities in climatic conditions, phenology and management practices. However, a major constraint in the production of LU maps is the prohibitive cost of collecting control site information, and it may not be feasible to obtain a sufficient amount of training data to allow the luxury of using only control sites that are close spatially to the region being mapped and from the same time period. The methods explored in this chapter consider modifications to SPREAD II that attempt to incorporate spatial context and, if successful, may compensate for the scarcity of training samples that are geographically and temporally close to the pixel being considered.

There are rich literatures on models that incorporate spatial context in several fields, including GIS, machine learning, image analysis and both frequentist and Bayesian statistics. In the image classification literature, early work on classifiers that include spatial context includes that of Hassner and Sklansky (1980), who introduced MRFs into image modelling, Richards et al. (1982), who use “label relaxation” in an iterative post-classification procedure, and Geman and Geman (1984) and Besag (1986), who discuss the use of local neighbourhoods in a more formal yet general image classification/sharpening context. Some other examples of the use of local neighbourhoods include Melgani and Serpico (2002), who present an iterative image fusion algorithm that combines the class labels at eight surrounding pixels at both the present and previous times and that of the pixel itself at the previous time. Kiiveri and Caccetta (1998) present a similar methodology, based on a longer series of images, for detecting salt-affected regions. Most methods make the assumption, perhaps implicitly, that the relationships between pixels can be described by an MRF, which is the basis for the first approach we consider in this chapter.

The other approach we explore here is augmentation of the training data with “partially-labelled” pixels; using the label (or selection probabilities) assigned to a pixel on the previous iteration of the MCMC procedure. Similar ideas are explored in Zhou et al. (2004) and Camps-Valls et al. (2007), who present a method of iteratively combining the information from a labelled sample and the labelling of all other pixels in an image using the recursion $F(t + 1) = \alpha SF(t) + (1 - \alpha)Y$, where $S$ is a normalised affinity matrix, $F(t)$ is an $n \times c$ matrix where each row corresponds to the probabilities that a pixel has each of the $c$ possible LUs,
and $Y$ is an $n \times c$ matrix where $Y_{ik}$ equals one if pixel $i$ is in the training sample and has LU $k$ and zero otherwise. Bruzzone et al. (2006) present a semi-supervised classification based on Transductive Support Vector Machines (TSVMs), which introduces the most informative, yet most likely to be correctly classified, unlabelled pixels into the training sample. Zhu (2005) presents a review of semi-supervised learning literature.

The two methods considered here are:

1. Use the class labels of neighbouring pixels in a manner similar to Kiiveri and Caccetta (1998) using a MRF.
2. Use similarities (and dissimilarities) between the NDVI profile of the current pixel and those of neighbouring pixels.

Method 1 results in a change in the form of the posterior distribution through $[C_i | C_{-i}, ...]$ and method 2 changes the way the we estimate $[Y_i | C_i]$. In method 1, the posterior distribution over the LUs at a pixel depends on the LU of a pixel and the LUs of the pixels in its neighbourhood. We model the spatial relationships in this case using a MRF. In method 2, we ‘model’ the spatial relationships by augmenting the set of control sites used in calculating $[Y_i | C_i, C_{-i}, Y_{-i}, \Psi]$, where $\Psi$ represents the parameters of the model used for the spatial relationships. We refer to this method as “the transductive approach”, following the terminology in Vapnik (1998).

Unlike the modification considered in Chapter 3, the methods explored here do not introduce a new independent source of ancillary data and hence we expect that their potential for improvement will be dependent on the accuracy of the underlying raw classifier; if it performs poorly, then any methodology that builds on its results is also likely to perform poorly. In the context of method 2, if the global information is poor, then local pixels will be poorly classified and including their partial labels is unlikely to improve things; worse, it may degrade the results. On the other hand, even if the global information is good, then adding the local information may or may not result in improvements, depending on whether the global information already captures most or all of the spatial dynamics. Method 1, on the other hand, is more about the spatial distribution of LU and hence is only likely to improve things if there is spatial clustering. In the context of the relative coarseness of the satellite imagery used here compared to the spatial scale of many of the LUs under consideration, it is dubious from the outset that we are likely to see improvement in the classification results.

Section 4.2 presents the theory underpinning the methods. Section 4.3 presents and describes the results. In Section 4.3.1 we present an analysis of how much the neighbourhood size and free parameters of the methods impact on the results for our study region, and in Section 4.4 we discuss the results.

4.2 Methods

Although there is no formal reason not to use both methods simultaneously, we limit ourselves to applying them separately in the current work. This is partly supported by Kiiveri and Campbell (1992), who found that including explicit models for the spatial correlation of the
imagery (specifically, Gaussian Conditional Autoregressive (CAR) models) achieved little gain over MRF models based purely on the labels of the immediately surrounding pixels and note that With hindsight, intuitively we might say that we are “double dipping” [by including the CAR component of the model] and most of the neighbouring information is extracted with the model for the pixel labels. They did not mention results that only used the CAR model, which is in some ways analogous to the second method we explore here.

4.2.1 Markov Random Field Approach

For the MRF approach we replace the term \([C|T, \Pi]\) in (2.5) with

\[
[C|T, \Pi] = \left[ C_i = k \right| C_{-i}, \Pi][C_{-i}|T, \Pi]
\]

\[
= \frac{\sum_{s \in R_{-i}} I(C_s = k) + 1}{|R_i| + K} \sum_{k \in K} [C|C_i = k, T, \Pi][C_i = k|T, \Pi]
\]

\[
\propto \frac{r_{ik} + 1}{|T, \Pi|} \left( \tilde{T}_k + 1 \right) \pi_{ik} \pi_k \n R\pi_k + 1 / \n R + K , \tag{4.1}
\]

where \(R_{-i}\) is the set of pixels in the neighbourhood of pixel \(i\), \(N_R = |R_{-i}|\) and \(r_{ik} = \sum_{n \in R_{-i}} I(C_n = k)\). The ‘+1’ in the numerator and ‘+K’ in the denominator ensure that every LU (including those not present in the neighbourhood) has a chance of being selected. Letting \(R_{ik}\) be a random variable over the possible values of \(r_{ik}\), we get the last line by limiting ourselves to the case that the categorical prior over the LUs is identical at each pixel and note that \(R_{ik}\) is binomial, giving

\[
[C = k|\Pi] = \sum_{r_{ik} = 0}^{N_R} [C = k|R_{ik} = r_{ik}, \Pi][R_{ik} = r_{ik}|\Pi]
\]

\[
= \sum_{r_{ik} = 0}^{N_R} \frac{r_{ik} + 1}{N_R + K} [R_{ik} = r_{ik}|\Pi]
\]

\[
= E(R_{ik}|\Pi) + 1 \n R + K = \n R\pi_k + 1 / \n R + K \tag{4.2}
\]

If we ignore the differences in neighbourhood size for boundary pixels we can assume that \([T, \Pi]\) has a multinomial distribution. Hence, including spatial context using this method is as simple as multiplying (2.5) by \((r_{ik} + 1)/(N_R\pi_k + 1)\).

An alternative to using the LU of surrounding pixels is to use the selection probabilities calculated in the previous iteration (\(\tilde{\pi}_{sk}\)). In this case we would use

\[
[C|T, \Pi] = \sum_{s \in B_{-i}} \tilde{\pi}_{sk} [C_{-i}|T, \Pi]
\]

\[
\propto (\tilde{T}_k + 1)\tilde{\pi}_{ik} / \pi_k ,
\]

where \(\tilde{\pi}_{ik} = \sum_{s \in B_{-i}} \tilde{\pi}_{sk}\). In this case \([C = k|\Pi] = \pi_k\), which is the same for both boundary and internal pixels. We do not explore this method further here.
4.2.2 The Transductive Approach

In previous chapters, the densities \( \mathbb{Y}_i | C_i = k \) are calculated using a kernel density smoother as

\[
\mathbb{Y}_i | C_i = k = \frac{1}{|S_k|} \sum_{s \in S_k} \prod_{l=1}^{L} \phi \left( \varphi_l (Y_i, Y_s); 0, \sigma_l^2 \right)
\]

(4.3)

where \( L \) is the number of metrics, \( S_k \) is the set of control sites with LU \( k \), \( \phi \left( x; \mu, \sigma^2 \right) \) is a normal distribution with mean \( \mu \) and variance \( \sigma^2 \), and \( \varphi_l (Y_i, Y_s) \) is the distance between the satellite imagery for pixels \( i \) and \( s \) under the metric \( l \).

Under the transductive approach we augment the set of control sites used for estimating the probability of observing the satellite imagery for a pixel given its LU to include pixels in its neighbourhood and the LU they were allocated on the previous iteration. The term for the satellite imagery is

\[
\mathbb{Y} | C, \Psi = \mathbb{Y}_i | C, \mathbb{Y}_{-i} | \Psi | \mathbb{Y}_{-i} | C, \Psi,
\]

where \( \mathbb{Y}_{-i} \) is the satellite imagery for all pixels excluding \( i \). The terms in the product on the right-hand side depend on each other and the evaluation of the product would require recursive evaluation of each—which would involve all pixels in the entire region—until convergence. This would be computationally expensive and we have not checked that this recursion would actually converge. Hence we drop the second term. We suspect that this may also make the procedure more stable, but have not explored this.

We combine the estimates of \( \mathbb{Y}_i | C_i = k \) (the probability of observing the satellite imagery based on the control sites alone) with \( \mathbb{Y}_i | C_i = k, C_{-i}, \mathbb{Y}_{-i}, \Psi \) (the probability of observing the satellite imagery using the neighbouring pixels) using the convex combination

\[
\mathbb{Y}_i | C_i = k, C_{-i}, \Psi = \alpha \mathbb{Y}_i | C_i = k + (1 - \alpha) \mathbb{Y}_i | C_i = k, C_{-i}, \mathbb{Y}_{-i}, \Psi.
\]

(4.4)

We continue to use (4.3) for evaluation of \( \mathbb{Y}_i | C_i \) and

\[
\mathbb{Y}_i | C_i = k, C_{-i}, \mathbb{Y}_{-i}, \Psi \propto \sum_{s \in R_{-i}} w (D (i, s), k) \prod_{l=1}^{L} \phi \left( \varphi_l (Y_i, Y_s); 0, \sigma_l^2 \right),
\]

(4.5)

where \( D (i, s) \) is a distance function and \( w (x, k) \) is a weighting function. In the current work, we use

\[
w (\bullet, k) = \frac{I (\hat{C}_s = k)}{\sum_{s \in R_{-i}} I (\hat{C}_i = k)}
\]

where \( \hat{C}_s \) is the LU chosen for pixel \( s \) at the previous iteration. Another possibility is to include pixels in the neighbourhood of pixel \( i \) with a distribution over their possible LUs. One method of doing this would be to use the selection probabilities calculated on the last iteration, which would amount to making \( w (\bullet, k) \) the probability that the pixel \( s \) has LU \( k \), but has not been explored here.

It appears from Figures 4.2 and 4.3 that the algorithm converges less quickly when we include spatial context. When using either of the methods described here, we first burn vanilla SPREAD II in for 4,000 iterations using (2.5), for another 4,000 iterations including the spatial
context. We finally run SPREAD II for 8,000 iterations to calculate the posterior probabilities.

### 4.3 Results

We use the same study region and data as described in Section 2.4. Figure 4.1 shows the results for “Murrumbidgee (A)”, the validation data, and the zonal constraint.

![Figure 4.1: Results for “Murrumbidgee (A)”](image)

Fig. 4.1: Results for “Murrumbidgee (A)”.

<table>
<thead>
<tr>
<th>Description</th>
<th>Match (%)</th>
<th>Average Posterior (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPREAD II: MRF, unconstrained</td>
<td>64.13</td>
<td>56.33</td>
</tr>
<tr>
<td>SPREAD II: MRF, constrained</td>
<td>56.40</td>
<td>56.33</td>
</tr>
<tr>
<td>SPREAD II: Local profiles, unconstrained</td>
<td>78.08</td>
<td>59.50</td>
</tr>
<tr>
<td>SPREAD II: Local profiles, constrained</td>
<td>62.99</td>
<td>59.50</td>
</tr>
<tr>
<td>SPREAD II: unconstrained</td>
<td>72.31</td>
<td>57.11</td>
</tr>
<tr>
<td>SPREAD II: constrained</td>
<td>58.15</td>
<td>57.11</td>
</tr>
</tbody>
</table>

Tab. 4.1: Classification accuracy of the MRF and transductive approaches contrasted with SPREAD II.

We can see in Figure 4.1 that the MRF approach results in a very different spatial pattern to that of vanilla SPREAD II. The LUs are allocated in similar locations to those we observe for vanilla SPREAD II, but regions of pixels with the same LU type are far more contiguous. The area of land allocated to each LU is quite similar when Algorithm 2 is applied and when it is not, than it is for the other analyses.
When the transductive approach is used, the area of land allocated to the dominant LU in the AgStats data, pasture, is significantly larger than under vanilla SPREAD II. This improves the percent match rate significantly for the reasons noted in Section 2.4.3.1. Visual comparison to the CLUM suggests that this may not indicate better performance of the method. The spatial distribution of the LU allocations is more scattered when Algorithm 2 is applied than they are for vanilla SPREAD II. While this increased the percent match rate compared to vanilla SPREAD II, visual comparison to the CLUM again suggests that this may not indicate better performance.

![Fig. 4.2: Proportion of times a pixel has been allocated each land use for six randomly selected pixels through the running of SPREAD II using the transductive approach.](image)

4.3.1 Neighbourhood Size

Under both the MRF and transductive approaches we need to choose a the neighbourhood size. Under the transductive approach we also need to choose a value for the parameter $\alpha$. In practice, these could be chosen using cross validation. We do not pursue this here but instead assess the impact of these parameters through simulation for the sake presentation.
Fig. 4.3: Proportion of times a pixel has been allocated each land use for six randomly selected pixels through the running of SPREAD II using the MRF approach.

<table>
<thead>
<tr>
<th>Neighbourhood Diameter</th>
<th>Match (%)</th>
<th>Average Posterior (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>75.40</td>
<td>62.31</td>
</tr>
<tr>
<td>5</td>
<td>71.30</td>
<td>61.62</td>
</tr>
<tr>
<td>7</td>
<td>69.30</td>
<td>61.31</td>
</tr>
<tr>
<td>13</td>
<td>68.78</td>
<td>60.89</td>
</tr>
<tr>
<td>25</td>
<td>74.48</td>
<td>60.82</td>
</tr>
</tbody>
</table>

Tab. 4.2: Percent match and average posterior for various parameter combinations using the MRF approach.
Including Spatial Context in SPREAD II

4. Including Spatial Context in SPREAD II

<table>
<thead>
<tr>
<th>Neighbourhood Diameter</th>
<th>Alpha</th>
<th>Match (%)</th>
<th>Average Posterior (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.20</td>
<td>79.57</td>
<td>64.14</td>
</tr>
<tr>
<td>3</td>
<td>0.40</td>
<td>79.30</td>
<td>64.01</td>
</tr>
<tr>
<td>3</td>
<td>0.60</td>
<td>79.21</td>
<td>63.85</td>
</tr>
<tr>
<td>3</td>
<td>0.80</td>
<td>78.94</td>
<td>63.60</td>
</tr>
<tr>
<td>5</td>
<td>0.20</td>
<td>79.76</td>
<td>64.22</td>
</tr>
<tr>
<td>5</td>
<td>0.40</td>
<td>79.60</td>
<td>64.06</td>
</tr>
<tr>
<td>5</td>
<td>0.60</td>
<td>79.24</td>
<td>63.86</td>
</tr>
<tr>
<td>5</td>
<td>0.80</td>
<td>79.01</td>
<td>63.59</td>
</tr>
<tr>
<td>7</td>
<td>0.20</td>
<td>79.86</td>
<td>64.20</td>
</tr>
<tr>
<td>7</td>
<td>0.40</td>
<td>79.50</td>
<td>64.03</td>
</tr>
<tr>
<td>7</td>
<td>0.60</td>
<td>79.27</td>
<td>63.95</td>
</tr>
<tr>
<td>7</td>
<td>0.80</td>
<td>79.11</td>
<td>63.67</td>
</tr>
<tr>
<td>13</td>
<td>0.20</td>
<td>79.63</td>
<td>64.01</td>
</tr>
<tr>
<td>13</td>
<td>0.40</td>
<td>79.47</td>
<td>63.91</td>
</tr>
<tr>
<td>13</td>
<td>0.60</td>
<td>79.27</td>
<td>63.83</td>
</tr>
<tr>
<td>13</td>
<td>0.80</td>
<td>78.88</td>
<td>63.62</td>
</tr>
<tr>
<td>25</td>
<td>0.20</td>
<td>79.50</td>
<td>64.07</td>
</tr>
<tr>
<td>25</td>
<td>0.40</td>
<td>79.17</td>
<td>64.03</td>
</tr>
<tr>
<td>25</td>
<td>0.60</td>
<td>78.75</td>
<td>63.83</td>
</tr>
<tr>
<td>25</td>
<td>0.80</td>
<td>78.22</td>
<td>63.63</td>
</tr>
</tbody>
</table>

Tab. 4.3: Percent match and average posterior for various parameter combinations using the transductive approach.

Table 4.2 shows the results for various neighbourhood sizes under the MRF approach. The average posterior is insensitive to the neighbourhood size and the percentage match is higher for the neighbourhood sizes of 3 and 25 than for the neighbourhood sizes of 5, 7 and 13. The insensitivity of the average posterior to the neighbourhood size suggests that it is unlikely these differences in the percentage match rates are significant and we conclude that, for these data, the neighbourhood diameter is not significantly impacting classification accuracy.

Table 4.3 shows the results for various neighbourhood sizes and values of $\alpha$. The results are insensitive to the neighbourhood diameter but both the percentage match and average posterior diminish marginally as $\alpha$ increases, though not enough to conclude that the neighbourhood diameter is significantly impacting classification performance.

4.4 Discussion

It appears from the results of this chapter that, for the data and classification scheme we have used here, including spatial context using the methods presented adds very little to the classification accuracy and, in fact, diminishes it. The simulation results presented in Section 4.3.1 show that the neighbourhood size and, in the case of the transductive approach, the tuning parameter $\alpha$, have little effect for this region and data.

These results are not surprising given the quality of discrimination achieved using this satellite imagery as presented in Figure 2.5. Further, for the classification schemes and image resolutions used in this thesis, the ‘expectation’ that neighbouring pixels share common LUs is also dubious. Many, if not most, of the LUs in the ALUM classification occur at scales of less than 1km$^2$ and hence many pixels will contain multiple LUs and will often contain land that does not belong to any class included in the classification scheme.
We have added only pixels in the immediate neighbourhood of a pixel to the training set. There is, however, no reason to limit the additional pixels to these nearest neighbours. We also only use pixels within the same region. This is primarily due to the computational overhead of dealing with different sets of LUs that may exist in neighbouring regions. We have not assessed whether this makes a significant difference to pixels near the edge of a region, which will have different neighbourhood sizes to those that are not.

We have also given all pixels in a neighbourhood the same weight. If one were to use larger neighbourhoods, it may be desirable to give more weight to pixels which are close in space and/or time than those further away, as they are likely to have more similar phenology and be exposed to similar climatic, environmental and potentially management conditions. The weighting could also take into account the confidence with which a pixel has been classified. Under the MRF approach, a weighting could be applied to (4.1) based on the distance between pixels \( i \) and \( s \).

Using this approach, provided that all pixels share the same categorical prior, (4.2) would still hold. For the transductive approach, two further approaches could be viable. The first is to base the variances of the kernel density smoother applied to the individual metrics in (4.5) on the distance between pixels, and the second is to change the functions \( w \) or \( D \).

Weighting the contribution to \( Y_i|C_i \) based on the spatial and/or temporal distance between a control site and pixel \( i \) could also be applied to the original training sample and was explored during the course of the current work. It was found, however, that the differences between profiles in our training sample did not vary consistently, either spatially or temporally. This could be due to the relatively coarse scale of the imagery used in this work, in which control sites are rarely pure pixels, and/or because the spectral overlap of the LU classes under the classification scheme we have used is too great.

We have chosen to use the LU of each pixel in the neighbourhood that was allocated on the previous iteration. Another possibility, as described in Section 4.2, is to use the selection probabilities that were used to perform that allocation. This would simplify the form of the posterior in the case of the MRF approach and resolve the issue of differing neighbourhood sizes. This was not pursued here since the results observed for the analyses presented showed no improvement over vanilla SPREAD II, and we felt further modifications would be unlikely to perform significantly better using the given data and classification scheme.

We treat \( \alpha \) as fixed free parameter which must be chosen by an analyst. Bruzzone and Marconcini (2009) present a domain-adaptation algorithm for TSVMs, which gradually introduces the imagery for a new time point for which there is no labelled data and removes the original training data. In the current work, we could achieve this by gradually decreasing \( \alpha \), which would decrease the influence of the ‘global’ training data and increase the influence of the local profiles. A schedule for doing this would need to be considered carefully and the samples used for calculation of \( \hat{\pi}_{ik} \) (as per (2.7)) should only use samples drawn under the final value of \( \alpha \), as changing the model parameters violates the assumptions underpinning the MCMC algorithm. The validity of this approach, and its sensitivities to the performance of the underlying raw classifier, would also require investigation.
5. PRODUCING A TIME SERIES OF MAPS USING SPREAD II

5.1 Introduction

The production of a temporal series of LU maps is an important problem in GIS. Such a series of maps is essential for a wide range of scientific and policy studies. Richards et al. (1982) note that in principle, it would seem pointless to perform routine classification of crops yearly using spectral data alone when records of previous growing patterns may be available to guide or assist the classification. In the Australian context, we have such records in the form of systematic high resolution mapping in the CLUM and previous NLUMs.

A series of LU maps can be created in a variety of ways. Most simply, maps can be produced independently of each other, as is the case for the NLUMs. Alternatively, information available for multiple time points can be considered. Approaches of the latter type can further be classified into two types: those based on change detection (Singh, 1989; Coppin et al., 2004; Lu et al., 2004; Radke et al., 2005) and those based on exploiting temporal relationships between images and potentially other data at multiple time points (Kuiveri and Caccetta, 1998; Kuiveri et al., 2003; Melgani and Serpico, 2002). The methods presented in the current work fall into the latter class, which can be further categorised into “cascade classification”, when the maps are classified sequentially (Swain, 1978), and “compound classification” when the maps are classified simultaneously (Duda and Hart, 1973). We note for clarity that some change detection methods include a cascade or compound classification step which exploits the temporal relationships between images. For example, Bruzzone and Serpico (1997) present a method for the detection of LC transitions in which a compound classification is followed by a post-classification. Under this method, a land-cover change in the considered couple of pixels is detected if the two classes to which such pixels are assigned, are different.

While we have not explored the use of change detection techniques in the current context, we felt that it was unlikely to perform well given the observed performance of the raw classifier we are working with and have not pursued it further here.

Swain (1978) describes a Bayesian framework for cascade or compound classification of images. The main result is based on the approximation that the probability of a particular series of LU types at a pixel is given by

\[
[C_1, C_0| Y_1, Y_0] \approx [Y_1| C_1][Y_0| C_0][C_1| C_0][C_0]. \tag{5.1}
\]

where the subscripts 0 and 1 denote two adjacent times, with 0 being the former. The author uses this expression in a Bayes minimum risk classifier for \( C_1 \), by summing over all values of \( C_0 \). The main benefit of this factorisation is that one only requires separate training sets for each
LU class, rather than a training set that contains examples from each possible LU transition type, which is rarely available and would be expensive to collect. Of course, if \( Y_1 \) and \( Y_0 \) are conditionally dependent given \( C_1 \) and \( C_0 \), then this approach will be sub-optimal. However, the simplification it induces makes it highly attractive and it forms the basis of many techniques, including one of the two considered herein.

Bruzzone and Serpico (1997) and Bruzzone et al. (1999) describe methods for estimating LC maps at two time points simultaneously using (5.1) and independent training samples for each time point. The former method estimates \([C_1|C_0]\) using an iterative technique and estimates \([C_1]\) directly from the sample. The second uses the Expectation Maximisation (EM) algorithm to estimate \([C_1, C_0]\) directly. For both methods it is assumed that, for each of the training samples, the number of pixels belonging to each class is approximately proportional to the prior probability of that class, and that the probability distribution of training data is the same as that of the data in the whole image. They note that in the absence of such a training sample these terms could be estimated from other pairs of maps or based on expert opinion, but that it would be difficult to estimate all required probabilities well using these means. It has been shown, however, that efficiency gains could be expected from using such a prior (Richards et al., 1982).

One of the strengths of the SPREAD II algorithm is the ease with which other forms of spatial and temporal information can be integrated into the underlying probability model and hence contribute to the classification process. In this chapter, we consider making a series of maps by incorporating the results of other LU maps and a transition matrix into SPREAD II. We develop methods for both a compound and cascade classification, and extend the simple Bayesian approach presented in Section 2.2.3 such that the prior for the second time point uses both the AgStats data for that time point and the posterior from a previous map.

5.2 Methods

In the current work we only consider the estimation of two maps. Extensions to more maps would be straightforward, though tedious mathematically, and, in the case of the compound classifier, probably computationally infeasible. In the following we assume the same setup as in previous chapters with the following modifications and additions.

There are \( K \) LUs all of which are feasible at all pixels at both time points and which are indexed by \( k_1 \) and \( k_2 \) at the two time points respectively. There are \( N \) agricultural pixels within the region, and areal constraint data (AgStats) providing estimates of the total area of each LU occurring within the region is available for both time periods.

The aim of the inference is to infer \( C_{im} \), the LUs of the individual pixels at each of the time points \( m \in 1, 2 \). We use the same satellite imagery with the same temporal structure as that used in previous chapters.

In the formulation of vanilla SPREAD II, \( C \) is estimated using a MCMC algorithm that assumes each pixel has a categorical prior distribution with parameters \( \pi_{kn} = 1/K \). This assumption greatly reduces the computational complexity of sampling from the posterior, as the prior distribution \([T|\Pi]\) is multinomial and hence only depends on the total number of cells allocated to each LU, not the LUs of cells themselves. This holds more generally for \( \pi_{kn} = \pi_k \). Without
this assumption, i.e. if \( \pi_{ka} \neq \pi_{kb} \implies k \in K; a, b \in 1, \ldots, N \), we must evaluate the general form of \([T|\Pi]\):

\[
[T|\Pi] = \sum_{s \in S} \prod_{n=1}^{N} \prod_{k=1}^{K} \pi_{nk}^{I_{nk}} = \sum_{s \in S} p_s
\] (5.2)

where \( S \) is the set of all combinations LUs such that \( \sum_{n=1}^{N} I_{nk} = T_k \forall k \) and \( I_{nk} \) is an indicator function, taking the value 1 when pixel \( n \) has LU \( k \) in \( s \), and 0 otherwise. It is computationally infeasible to calculate exact values for this distribution unless the number of pixels and LUs is small, since \(|S| = N!/1!2! \ldots TK!\).

If we randomly choose a sample from \( S \) and calculate (5.2), \( p_s \) is then a product of a large number of scaled Bernoulli random variables and is hence itself a random variable, \( P \). The central limit theorem states that, as the product \( KN \) increases, \( |P| \) will approach a lognormal distribution and \( \ln(P) \) will approach a normal distribution with some mean \( \mu \) and variance \( \sigma^2 \). We use this result to develop an approximation to (5.2).

We can estimate (5.2) using Monte Carlo simulation by choosing \( R \) samples from \( S \) and forming the estimate

\[
\frac{|S|}{R} \sum_{r=1}^{R} \prod_{n=1}^{N} \prod_{k=1}^{K} \pi_{nk}^{I_{nk}},
\] (5.3)

which the law of large numbers states will converge to

\[
|S|E(P) = |S| \exp(\mu + \sigma^2/2)
\] (5.4)

as \( R \) tends to infinity.

We can estimate the parameters \( \mu \) and \( \sigma^2 \) using Monte Carlo simulation on the log scale using

\[
\hat{\mu} = \frac{1}{R} \sum_{r=1}^{R} \sum_{n=1}^{N} \sum_{k=1}^{K} \ln(\pi_{nk}) I_{nk}
\] (5.5)

and

\[
\hat{\sigma}^2 = \frac{1}{R-1} \sum_{r=1}^{R} \left[ \sum_{n=1}^{N} \sum_{k=1}^{K} I_{nk} \ln(\pi_{nk}) - \hat{\mu} \right]^2
\] (5.6)

under the constraint that \( \sum_{n=1}^{N} I_{nk} = T_k \).
We can expand the first term in (5.6) as
\[
\frac{1}{R-1} \sum_{r=1}^{R} \left[ \sum_{n=1}^{N} \sum_{k=1}^{K} I_{nkr} \ln (\pi_{nk}) \right] ^2 = \\
\sum_{n=1}^{N} \sum_{k=1}^{K} \ln^2 (\pi_{nk}) \frac{1}{R-1} \sum_{r=1}^{R} I_{nkr} ^2 \\
+ \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{l \neq k} \ln (\pi_{nk}) \ln (\pi_{nl}) \frac{1}{R-1} \sum_{r=1}^{R} I_{nkr} I_{nlr} \\
+ \sum_{n=1}^{N} \sum_{m=1}^{N} \sum_{k=1}^{K} \sum_{l \neq k} \ln (\pi_{nk}) \ln (\pi_{ml}) \frac{1}{R-1} \sum_{r=1}^{R} I_{nkr} I_{mlr}. 
\]

The limits of the terms involving $R$ in (5.5) and (5.7) as $R$ tends towards infinity are
\[
\lim_{R \to \infty} \frac{1}{R} \sum_{r=1}^{R} I_{nkr} = \frac{T_k}{N} \\
\lim_{R \to \infty} \frac{1}{R-1} \sum_{r=1}^{R} I_{nkr} ^2 = \frac{T_k}{N} \\
\lim_{R \to \infty} \frac{1}{R-1} \sum_{r=1}^{R} I_{nkr} I_{nlr} = 0 \\
\lim_{R \to \infty} \frac{1}{R-1} \sum_{r=1}^{R} I_{nkr} I_{mk} = \frac{T_k(T_k-1)}{N(N-1)} \\
\lim_{R \to \infty} \frac{1}{R-1} \sum_{r=1}^{R} I_{nkr} I_{mlr} = \frac{T_kT_l}{N(N-1)}
\]

and hence the limit of (5.5) can be written as
\[
\hat{\mu} = \sum_{k=1}^{K} \frac{T_k}{N} \sum_{n=1}^{N} \ln (\pi_{nk}). 
\]

Since the $\lim_{R \to \infty} R/(R-1) = 1$, the limit of (5.6) can be written as
\[
\hat{\sigma}^2 = \sum_{k=1}^{K} \frac{T_k}{N} \sum_{n=1}^{N} \ln^2 (\pi_{nk}) \\
+ \sum_{k=1}^{K} \frac{T_k(T_k-1)}{N(N-1)} \sum_{n=1}^{N} \sum_{m=1 \neq n}^{N} \ln (\pi_{nk}) \ln (\pi_{mk}) \\
+ \sum_{k=1}^{K} \sum_{l \neq k} \frac{T_kT_l}{N(N-1)} \sum_{n=1}^{N} \sum_{m=1 \neq n}^{N} \ln (\pi_{nk}) \ln (\pi_{ml}) - \hat{\mu}^2. 
\]

Note that the double sums over the pixels in (5.9) and (5.10) are constant given $\Pi$ and hence
that evaluation of those expressions for a given $T$ only requires summation over the LUs.

It is straightforward to show that if $\pi_{nk} = 1/K$ for all pixels and LUs that (5.9) equals $-N\ln(K)$ and (5.10) equals zero. Substitution of these values into (5.4) gives (5.2).

For small numbers of pixels we are able to calculate (5.2) exactly by computing every permutation for which the constraint holds and summing the probability of each, allowing us to evaluate the approximation directly for these cases.

Fig. 5.1: Evaluation of the approximation for a region with 17 and 4 LUs. $\pi_{nk}$ are drawn uniform distributions with ranges given by $[10, 13]$, $[5, 7]$, $[1, 2]$ and $[1, 2]$ for LUs 1 through 4 respectively and scaled for each pixel so that $\sum_{i=1}^{K} \pi_{nk} = 1$. The top 4 panels show $[\ln(P)]$ over all permutations contained in (5.2) and estimated using and approximation based on the lognormal and logskeunormal distributions for an additional pixel of each LU type. The number of pixels of each LU type is shown below each panel. The bottom 2 panels show the raw and normalised relationships between the true and approximated probabilities using the approximation based on the lognormal distribution.

For the central limit theorem to hold, all central moments of order three and higher should vanish as the number of pixels increases. While we have not determined this analytically, we
5. Producing a Time Series of Maps Using SPREAD II

can use calculations similar to those shown in (5.4) through (5.10) to calculate the third central moment and thus the skew of (5.2). We can then check to see that the skew does decrease as the number of pixels increases. This will also allow us to use a skew-lognormal distribution to approximate (5.2) and determine if it improves the approximation, thus shedding insight into the validity of using a lognormal approximation.

Figure 5.1 shows [\ln(P)] for changes in the LU of a single pixel in a hypothetical region with 17 pixels and 4 LUs, and the corresponding estimates derived from the normal and skew-normal distributions. These are compared to the true distribution, which is based on all permutations contained in (5.2). This figure shows that [\ln(P)] is reasonably well approximated by either a normal or a skew-normal distribution for this particular \( \mathbf{N} \), which was chosen to reflect the typical structure we see in the posterior of the outputs from SPREAD II. Many other values of \( \mathbf{N} \) were tested and similar results observed, though, anecdotally, the approximation seemed to perform worse when the prior distribution at each pixel was chosen independently for each pixel based on a uniform distribution over \([0, 1]\). Even in that case, however, reasonable results were observed.

In practice we deal with regions containing many more than 17 pixels, with relatively small regions containing over a thousand pixels, and we expect the approximation to improve in proportion to \(1/\sqrt{N}\) as the number of pixels \(N\) increases. It is infeasible to perform the calculations required for the exact analysis presented in Figure 5.1 for realistic numbers of pixels due to the complexity constraint imposed by the form of (5.2). Neither is it feasible to obtain reasonable estimates of (5.2) using simulation, as estimating the mean and variance of the lognormal distributions considered here requires prohibitively large samples. The variance of the estimate of the mean from a lognormal distribution is \((e^{\sigma^2} - 1)e^{2\mu + \sigma^2}/R\), where \(R\) is the number of samples. At 200 pixels we get values for \(\mu\) and \(\sigma^2\) of around \(-330\) and \(150\) respectively, giving a mean of around \(10^{-110}\) and a standard deviation of around \(10^{-80}\). We would therefore need samples of sizes above \(10^{30}\) to start getting reasonable estimates. Further, we must calculate \(\prod_{n=1}^{N} \prod_{k=1}^{K} \frac{\pi_{nk}}{\pi_{nk}}\) directly, which, for even relatively modest numbers of pixels and LUs, results in numerical underflow using double precision floating point numbers.

As noted above, if the approximation suggested above is appropriate, then we expect the skew of the distribution of \(\ln(P)\) to decrease as the number of pixels increases. Figure 5.2 shows the skew of this distribution for various numbers of pixels and it is clear that the absolute value of the skew is decreasing as the number of pixels increases.

5.2.1 Cascade Classification

We can use (5.9) and (5.10) in estimating (5.2) via (5.4) under priors that do not have the property that \(\pi_{nk} = \pi_k\), avoiding the computational complexity involved in calculating (5.2) directly. One application of this is cascade classification, where we use the posterior distribution or final classification for a map produced at one time point in forming a prior distribution for a map at another time point.

Several options are available for how to use a map in forming a prior for another map. The first is to simply use the posterior distribution over the LUs for each pixel in one map as the prior distribution for another. This may be reasonable if the probability of LU change at each pixel is small between the two time periods, which may be the case for maps that are close in
time, or for LUs which tend to persist through time (e.g. grapes).

A second possibility is to use a transition matrix, $\Psi$, where the entry $\psi_{kl}$ in row $k$ and column $l$ is the (prior) probability that pixel changes from LU $k$ in the first time period, to LU $l$ in the second time period; i.e. $\psi_{kl} = [C_{nm} = l | C_{(m-1)} = k]$, where $C_{nm}$ is the LU of pixel $n$ at time $m$. The prior probability that $C_{nm} = l$ is then

$$\pi_{lnm} = \frac{\sum_{k=1}^{K} \hat{\pi}_{kn(m-1)} \psi_{kl}}{\text{min}(\pi_{\bullet nm})/2}$$

where $\hat{\pi}_{knm}$ is the posterior probability that pixel $n$ has LU $k$ at time $m$. In the implementation we have used here, we set $\pi_{lnm} = \text{min}(\pi_{\bullet nm})/2$ if (5.11) is equal to zero to simplify the implementation.

A third approach is to use the ‘final allocation’ as produced by algorithm 1 by simply setting $\pi_{nm}$ equal to the row of $\psi_{k\bullet}$, the row of $\Psi$ corresponding to the LU assigned to pixel $n$ at time $m - 1$. This approach has not been considered here.

Whichever method is employed, the posterior for a pixel at a later time point is given by (2.5), with $\Pi$ being calculated from one of the methods just described, and hence $\pi_{kn}$ is no longer constant for all $n$.

### 5.2.2 Compound Classification

Using cascade classification we could move either forward or backward through time. If there is reason to believe that a map at a later time point is more accurate than earlier maps, then it may be better to use it as a prior for a previous map. This could occur, for example, due to
The analogous terms to (5.5) and (5.7) are when we take logs of the product terms, we get

\[ \log \left( \prod_{s=1}^{S} C_{s} \right) \]

We can develop an approximation to (5.13) using the approach developed above. In this case, if we assume that the probability of observing the imagery at a given pixel at a given point in time depends only on the LU of the pixel, and that pixels are mutually independent, the full conditional posterior distribution for two time points is proportional to

\[
\begin{align*}
[C_{s}, C_{s-1}, T_{s}, T_{s-1}, Y_{nm}, Y_{n-1,m}, \Theta] & \propto \frac{[Y_{nm}|C_{n-1,m}|[Y_{n-1,m}|C_{n-1,m-1}][C_{nm}, C_{n-1,m-1}][\Pi, \Psi][T_{s}, T_{s-1}][\Theta]}{[T_{s}, T_{s-1}][\Pi, \Psi]},
\end{align*}
\]

(5.12)

where \( C_{nm} \) is the LU of pixel \( n \) at time point \( m \), \( C_{s} = \{C_{1m}, \ldots, C_{Nm}\} \), \( T_{km} \) is the total number of pixels with LU \( k \) at time point \( m \), \( T_{s} = \{T_{1m}, \ldots, T_{Km}\} \) and \( \Theta = \{\Pi, \Phi, \Psi\} \). The denominator of (5.12) can be written as

\[
[T_{s}, T_{s-1}][\Pi, \Psi] = \sum_{s_{1} \in S_{1}} [T_{s}], \left[ C_{s} \right][\Pi] = \sum_{s_{2} \in S_{2}, s_{1} \in S_{1}} \left[ s_{2}, C_{s} \right][\Pi] = \sum_{s_{2} \in S_{2}, s_{1} \in S_{1}} \prod_{k=1}^{K} \pi^{l_{nk}}_{s_{1},s_{2}} \prod_{l=1}^{L} \psi^{l_{nkl}}_{s_{1},s_{2}},
\]

(5.13)

where \( S \) is the set of all vectors \( C_{s} \) such that (2.1) holds for all \( T_{k,m-1} \in T_{s-1} \), and \( I^{n}_{v} \) denotes an indicator that takes the value 1 if the value of \( n \) is \( k \) at time \( m \) in vector \( s \).

We can develop an approximation to (5.13) using the approach developed above. In this case, when we take logs of the product terms, we get

\[
\sum_{n=1}^{N} \sum_{k=1}^{K} I^{l}_{n_k} \ln (\pi_{nk}) + \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{l=1}^{L} I^{l}_{n_k l} \ln (\psi_{nkl}),
\]

(5.14)

The analogous terms to (5.5) and (5.7) are

\[
\frac{1}{R} \sum_{r=1}^{R} \sum_{n=1}^{N} \sum_{k=1}^{K} I^{l}_{n_k} \ln (\pi_{nk}) + \frac{1}{R} \sum_{r=1}^{R} \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{l=1}^{L} I^{l}_{n_k l} \ln (\psi_{nkl})
\]

(5.15)

\[ \sum_{n=1}^{N} \sum_{k=1}^{K} \ln (\pi_{nk}) \frac{1}{R} \sum_{r=1}^{R} I^{l}_{n_k} + \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{l=1}^{L} \ln (\psi_{nkl}) \frac{1}{R} \sum_{r=1}^{R} I^{l}_{n_k}, \]
and
\[
\frac{1}{R-1} \sum_{r=1}^{R} \left[ \sum_{n=1}^{N} \sum_{k=1}^{K} I_{nkrt}^1 \ln (\pi_{nk}) + \sum_{n=1}^{N} \sum_{l=1}^{L} I_{nktr}^2 \ln (\psi_{nkl}) \right]^2
\]
\[
= \frac{1}{R-1} \sum_{r=1}^{R} \left[ \sum_{n=1}^{N} \sum_{k=1}^{K} I_{nkrt}^1 \ln (\pi_{nk}) \right]^2
\]
\[
+ \frac{1}{R-1} \sum_{r=1}^{R} \left[ \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{l=1}^{L} I_{nktr}^1 I_{nklr}^2 \ln (\psi_{nkl}) \right]^2
\]
\[
+ \frac{2}{R-1} \sum_{r=1}^{R} \left[ \sum_{n=1}^{N} \sum_{k=1}^{K} I_{nkrt}^1 \ln (\pi_{nk}) \right] \left[ \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{l=1}^{L} I_{nktr}^1 I_{nklr}^2 \ln (\psi_{nkl}) \right]
\]
respectively. The first term in this expression is the same as (5.7). Ignoring terms that will evaluate to zero, the second term can be expressed as
\[
\frac{1}{R-1} \sum_{r=1}^{R} \left[ \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{l=1}^{L} I_{nkrt}^1 I_{nklr}^2 \ln (\psi_{nkl}) \right]^2
\]
\[
= \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{l=1}^{L} \ln (\psi_{nkt})^2 \frac{1}{R-1} \sum_{r=1}^{R} (I_{nkrt}^1)^2 (I_{nklr}^2)^2
\]
\[
+ \sum_{n_1=1}^{N} \sum_{n_2=1}^{N} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \sum_{l_1=1}^{L} \sum_{l_2=1}^{L} \ln (\psi_{n_1kt_1}) \ln (\psi_{n_2kt_2}) \frac{1}{R-1} \sum_{r=1}^{R} t_{n_1kr}^1 t_{n_1lr}^2 t_{n_2kr}^1 t_{n_2lr}^2
\]
\[
+ \sum_{n_1=1}^{N} \sum_{n_2=1}^{N} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \sum_{l_1=1}^{L} \sum_{l_2=1}^{L} \ln (\psi_{n_1kt_1}) \ln (\psi_{n_2kt_2}) \frac{1}{R-1} \sum_{r=1}^{R} t_{n_1kr}^1 t_{n_1lr}^2 t_{n_2kr}^1 t_{n_2lr}^2
\]
\[
+ \sum_{n_1=1}^{N} \sum_{n_2=1}^{N} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \sum_{l_1=1}^{L} \sum_{l_2=1}^{L} \ln (\psi_{n_1kt_1}) \ln (\psi_{n_2kt_2}) \frac{1}{R-1} \sum_{r=1}^{R} t_{n_1kr}^1 t_{n_1lr}^2 t_{n_2kr}^1 t_{n_2lr}^2
\]
\[
(5.17)
\]
and the third as

\[
\frac{2}{R-1} \sum_{r=1}^{R} \left[ \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{l=1}^{L} I_{nk,l} \ln (\pi_{nk}) \right] \left[ \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{l=1}^{L} I_{nk,l}^2 \ln (\psi_{nk,l}) \right] \\
= \frac{2}{R-1} \sum_{r=1}^{R} \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{l=1}^{L} I_{nk,l} I_{nk,l}^2 I_{nk,l}^2 \\
= \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{l=1}^{L} \ln (\pi_{nk}) \ln (\psi_{nk,l}) \frac{2}{R-1} \sum_{r=1}^{R} I_{nk,l} I_{nk,l}^2 I_{nk,l}^2 \\
+ \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{l=1}^{L} \ln (\pi_{nk}) \ln (\psi_{nk,l}) \frac{2}{R-1} \sum_{r=1}^{R} I_{nk,l} I_{nk,l}^2 I_{nk,l}^2. \\
\] (5.18)

Since we are sampling independently from the LUs at the first and second periods, we take limits in the above expressions to get

\[
\lim_{R \to \infty} \frac{1}{R} \sum_{r=1}^{R} I_{nk,l} I_{nk,l}^2 = \frac{T_{k1} T_{l2}}{N^2},
\]

\[
\lim_{R \to \infty} \frac{1}{R} \sum_{r=1}^{R} I_{nk,l} I_{nk,l}^2 I_{nk,l}^2 = \frac{T_{k1} T_{l2}}{N^2},
\]

\[
\lim_{R \to \infty} \frac{1}{R} \sum_{r=1}^{R} I_{nk,l} I_{nk,l}^2 I_{nk,l}^2 = \frac{T_{k1}(T_{k1} - 1)T_{l2}(T_{l2} - 1)}{N^2(N-1)^2},
\]

\[
\lim_{R \to \infty} \frac{1}{R} \sum_{r=1}^{R} I_{nk,l} I_{nk,l}^2 I_{nk,l}^2 = \frac{T_{k1}(T_{k1} - 1)T_{l2}T_{l2}}{N^2(N-1)^2},
\]

\[
\lim_{R \to \infty} \frac{1}{R} \sum_{r=1}^{R} I_{nk,l} I_{nk,l}^2 I_{nk,l}^2 = \frac{T_{k1}(T_{k1} - 1)T_{l2}}{N^2(N-1)^2},
\]

\[
\lim_{R \to \infty} \frac{2}{R-1} \sum_{r=1}^{R} I_{nk,l} I_{nk,l}^2 I_{nk,l}^2 = \frac{2T_{k1} T_{l2}}{N^2},
\]

and hence we can express the mean and variance of (5.14) as

\[
\hat{\mu}_t = \sum_{k=1}^{K} \frac{T_{k1}}{N} \sum_{n=1}^{N} \ln (\pi_{nk}) + \sum_{k=1}^{K} \frac{T_{k1} T_{l2}}{N^2} \sum_{n=1}^{N} \ln (\psi_{nk,l}),
\] (5.20)
respectively. If we set \( \pi_{nk} = 1/K \) and \( \psi_{nk} = \psi_{kl} \), we get significant simplification, with (5.20) becoming

\[
\tilde{\mu}_t = -N \ln(K) + \sum_{k=1}^{K} \sum_{l=1}^{L} \frac{T_{kl} T_{l2}}{N} \ln(\psi_{kl})
\]

(5.22)
\[
\hat{\sigma}_t^2 = N^2 \ln^2(K)
\]
\[+
\sum_{k=1}^{K} \sum_{l=1}^{L} \frac{T_{kl}T_{k2}}{N} \ln^2(\psi_{kl})
\]
\[+
\sum_{k=1}^{K} \sum_{l=1}^{L} \frac{T_{kl}T_{k2}}{N(N-1)} \ln^2(\psi_{kl})
\]
\[-
\sum_{k=1}^{K} \sum_{l=1}^{L} \frac{T_{kl}T_{k1}T_{k2}}{N(N-1)} \ln(\psi_{kl1}) \ln(\psi_{kl2})
\]
\[-
\sum_{k=1}^{K} \sum_{l=1}^{L} \frac{T_{kl}T_{k1}T_{k2}}{N(N-1)} \ln(\psi_{kl1}) \ln(\psi_{kl2})
\]
\[-2 \ln(K) \sum_{k=1}^{K} \sum_{l=1}^{L} \frac{T_{k1}T_{k2}}{N} \ln(\psi_{kl}) - \hat{\mu}_t^2.
\]

It is straightforward to show that (5.22) equals \(-2N \ln(K)\) and (5.23) equals zero if \(\psi_{kl} = 1/K\) for all LUs.

At first glance it would appear that a limitation of the approximation presented here is that the probability of all LU transitions must be positive at all pixels so that \(\ln(\psi_{n,m,kl}) \neq -\infty\). However, by setting the products \(T_{n,k1}T_{n,k2}\) in (5.14) to zero if \(\psi_{n,k} = 0\), it is straightforward to show that, provided that \(\pi_{nk} \neq 0\) for all pixels, we can simply exclude terms involving the product of one or more terms with \(\psi_{n,k} = 0\) and take account of the reduced number of terms in \(S_2\). For the case that \(\psi_{n,k} = \psi_{kl}\), for LU \(l\) we have
\[
|S_2| \propto \frac{N - \sum_{k \in K^*} T_{k1}}{T_{k2}}
\]
where \(K^*\) is the set of LUs for which \(\psi_{kl} = 0\).

Further, in the case where \(\pi_{nk} = 0\) for one or more pixels, we could modify the approximation such that terms in the denominators involving \(N\) would be replaced by sums over \(N_k\), the numbers of pixels which can take on LU \(k\) (i.e. have \(\psi_{n,k} \neq 0\)). However, implementing the details of handling zero values in \(\Psi\) and \(\Pi\) would be extremely fiddly and in the work presented here we simply replace any zero values with very small values.

In the formulation of the approximation it is assumed that all pixels are present in both maps and contained within the same region. However, in practice some pixels are not agricultural at all time points, or the region boundaries change and a pixel may belong to different regions at different time points. While it is possible to modify the selection probabilities (and in particular \([T_{m}, T_{(m-1)}]T_{i,m}, T_{i,(m-1)}\)) to handle these situations, doing so would significantly complicate the implementation of the algorithm and make it (more) computationally intensive to run.\(^1\) In practice, the number of pixels that fall into these two classes constitute fractions

\(^1\) The posterior would depend on the number of pixels falling into both classes across all regions that directly and indirectly intersect (indirectly being a region that intersects with an intersecting region, or intersects with a region that intersects with an intersecting region etc.) with a region.
of a percent of the total region, and the choice was made to use selection probabilities based on (2.5) for pixels that fall into these classes.

5.2.3 Transition Matrix

In the development of the mathematics above we have assumed that $\Psi$ is independent of all other parameters in the model, in particular the areal constraints, $\Theta$, and pixel priors, $\Pi$. This may be unrealistic in practice. For instance, the actual difference between the areas of land under each LU between two time points is completely determined by the pixel-wise LU transitions that occur between those two time points. As such, $\Theta$ clearly should not be considered independently of $\Psi$. Based on this observation, the relation

$$\mu_{m+1} = \Psi \mu_m,$$

(5.25)

where $\mu_m$ is the vector of areas reported in the AgStats for a region at time $m$, should approximately hold. If it does not, then there may be a ‘tension’ between the constraints based on the AgStats and the transition probabilities. Such a tension is liable to decrease the stability of the sampler and result in estimates that are less consistent with both the AgStats and the transition probabilities.

In the absence of any other options, we estimate $\Psi$ from the tables of AgStats for times one and two, which contain data for a large number of regions, using (5.25) and a constrained regression approach in which the coefficients are constrained to be positive and sum to one. Of course, this approach is not ideal. Phenomena such as crop rotations could give rise to extensive LU (or at least LC) change, but induce very little change in the overall areas under each LU. It also assumes that $\Psi$ is identical across all regions, which, as noted previously, is unrealistic.

The transition matrix used in this study is shown in Figure 5.3. The large off-diagonal elements are clearly spurious. However, the LUs they correspond to are either not present or have relatively small areas in the regions we are dealing with here. No effort has been made to remedy them. Further, given that we have been advised by experts that LU change between the times we consider here is probably in the order of five percent, we would expect the diagonal elements, especially those for the dominant LUs, to be larger.
### Fig. 5.3: Transition matrix used in this study calculated using the methods described in Section 5.2.3.

<table>
<thead>
<tr>
<th>Grazing NP</th>
<th>Grazing SP</th>
<th>Agroforestry</th>
<th>Winter Cereals</th>
<th>Summer Cereals</th>
<th>Rice</th>
<th>Winter Legumes</th>
<th>Summer Legumes</th>
<th>Winter Oilseeds</th>
<th>Summer Oilseeds</th>
<th>Sugar Cane</th>
<th>Hay</th>
<th>Cotton</th>
<th>Other Non-cereal Crops</th>
<th>Vegetables</th>
<th>Citrus</th>
<th>Apples</th>
<th>Pears &amp; Other Pome</th>
<th>Non-tropical Stone Fruit</th>
<th>Tropical Stone Fruit</th>
<th>Nuts</th>
<th>Plantation Fruit</th>
<th>Grapes</th>
</tr>
</thead>
<tbody>
<tr>
<td>89</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>73</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
5. Producing a Time Series of Maps Using SPREAD II

5.2.4 A Simple Bayesian Classifier

In Chapter 2 we introduced a simpler classifier in which the priors were based directly on the AgStats. An equivalent formulation is not so obvious for the case of cascade classification presented here, as we have two sources of prior information: the posterior from a previous map and the AgStats for the current map. If we wish to use both sources of information, some way of combining them must be derived. The approach we have used is to iteratively re-scale the posterior from the previous map, such that the expected areas match those of the AgStats for the new map using Algorithm 3:

\textbf{Algorithm 3} Rescaling of posterior

1: set $j = 0$ and $\hat{\pi}_{nkm}^{(m+1)} = \hat{\pi}_{nkm}$
2: set $\pi_{nk}^{(m+1)(j+1)} = \frac{\mu_{k}^{(m+1)}}{\sum_{i=1}^{N} \hat{\pi}_{ik}^{(m+1)(j+1)}}$
3: set $\pi_{nk}^{(m+1)(j+2)} = \frac{\pi_{nk}^{(m+1)(j+1)} \sum_{k=1}^{K} \pi_{nk}^{(m+1)(j+1)}}{\sum_{k=1}^{K} \pi_{nk}^{(m+1)(j+1)}}$
4: if $|\sum_{i=1}^{N} \pi_{ik}^{(m+1)(j+2)} - \mu_{k}^{(m+1)}| > \epsilon$, set $j = j + 2$ and repeat 2, and 3,
5: set $\pi_{nk} = \pi_{nk}^{(m+1)(j+1)}$

In Algorithm 3, $\hat{\pi}_{nkm}$ is the posterior for LU $k$ at pixel $n$ at time point $m$, and $\mu_{k}^{(m)}$ is the area of LU $k$ reported in the AgStats for time point $m$. We then use these re-scaled posteriors as the priors for the map made for time point $m + 1$.

5.3 Results

Unfortunately, for the test region we have been using throughout previous chapters, "Murrumbidgee (A)", we only have CLUM validation data for 2005. Given that we have advice from experts that LU change within the region is probably in the order of five percent, we present results for this region anyway to give the reader a qualitative impression of the impacts of the methods presented in the context of familiar data. We do, however, have validation data for 2010 for other regions, and present equivalent results for one of these for the purpose of quantitative validation.

Table 5.1 shows the match results for “Horsham (RC) Bal”. The columns “match percent” and “expected” show the average of the results for 2005 and 2010. The results for vanilla SPREAD II are included for comparison. Note that the slight difference in the results for vanilla SPREAD II and the other SPREAD II based classifiers in 2005 is due to slight differences in the AgStats; in 2005 there was no cotton reported, but in one or more of the regions which overlap with “Horsham (RC) Bal”, there was. Since we ensure that all pixels have the same LUs at both time points when we are doing either a compound or cascade classification, cotton is included as a possible LU for these methods, but not for vanilla SPREAD II.

Of all classifiers, the compound classifier performs best in both years, and on average, with respect to the match percentage. Both the simple Bayesian classifier which includes the transition probabilities, and the SPREAD II compound classifier which includes the transition probabilities, perform very similarly in 2010. The latter performs better in 2005, and hence on average. They both perform similarly to SPREAD II. The improvement from including the transition probabilities in the simple Bayesian approach is about five percent. Interestingly, SPREAD II
using the cascade classifier without the transition probabilities, performs best with respect to the average posterior of the true class both in 2010 and on average.

The spatial configuration of the vanilla SPREAD II results for “Murrumbidgee (A)” in 2010 are quite different to those for 2005. The results for the methods which include spatial context are, in general, quite similar to those we have observed in 2005, reflecting the impact of the transition probabilities.

It is interesting to note the differences in the raw classifier when the transition probabilities are included (Panels (a) and (c) in Figure 5.6). To help interpret this, Figure 5.7 shows the results for the raw classifier in 2010. We can see here that the classification based on the satellite imagery alone is very different in 2010, and more similar to that in Panel (c) of Figure 5.6, with the exception of the abundance of rice. As an aside, given the experts advice on the likely scale LU change, comparison of these two maps highlights the deficiency of the satellite imagery in this region in 2005.

![Fig. 5.4: Results for “Horsham (RC) Bal” in 2005. (q) CLUM data, see Table 5.1 for other panel descriptions.](image-url)
5. Producing a Time Series of Maps Using SPREAD II

Fig. 5.5: Results for “Horsham (RC) Bal” in 2010. (q) CLUM data, see Table 5.1 for other panel descriptions.

Fig. 5.6: Results for “Murrumbidgee (A)” in 2010. (q) CLUM data, see Table 5.1 for other panel descriptions.
### Classification Accuracy of Classifiers That Include Temporal Context in "Horsham (RC) Bat".

<table>
<thead>
<tr>
<th>Classifier Type</th>
<th>2000</th>
<th>2005</th>
<th>2010</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Raw classifier: using cascade classifier, unconstrained</td>
<td>19.51</td>
<td>35.39</td>
<td>24.06</td>
<td>32.49</td>
</tr>
<tr>
<td>(b) Raw classifier: using cascade classifier, constrained</td>
<td>64.69</td>
<td>35.39</td>
<td>69.60</td>
<td>35.39</td>
</tr>
<tr>
<td>(c) Raw classifier: using cascade classifier with transition probabilities, unconstrained</td>
<td>19.51</td>
<td>35.39</td>
<td>22.57</td>
<td>35.14</td>
</tr>
<tr>
<td>(d) Raw classifier: using cascade classifier with transition probabilities, constrained</td>
<td>64.69</td>
<td>35.39</td>
<td>60.94</td>
<td>35.14</td>
</tr>
<tr>
<td>(e) Raw classifier: using cascade classifier and AgStats as priors, unconstrained</td>
<td>56.35</td>
<td>53.41</td>
<td>73.74</td>
<td>64.55</td>
</tr>
<tr>
<td>(f) Raw classifier: using cascade classifier and AgStats as priors, constrained</td>
<td>69.17</td>
<td>59.36</td>
<td>74.38</td>
<td>69.17</td>
</tr>
<tr>
<td>(g) Raw classifier: using cascade classifier with transition probabilities and AgStats as priors, unconstrained</td>
<td>56.35</td>
<td>53.41</td>
<td>78.83</td>
<td>65.18</td>
</tr>
<tr>
<td>(h) Raw classifier: using cascade classifier with transition probabilities and AgStats as priors, constrained</td>
<td>69.17</td>
<td>59.36</td>
<td>74.48</td>
<td>69.17</td>
</tr>
<tr>
<td>(i) SPREAD II: using cascade classifier, unconstrained</td>
<td>74.01</td>
<td>74.01</td>
<td>74.01</td>
<td>74.01</td>
</tr>
<tr>
<td>(j) SPREAD II: using cascade classifier, constrained</td>
<td>72.89</td>
<td>72.89</td>
<td>72.89</td>
<td>72.89</td>
</tr>
<tr>
<td>(k) SPREAD II: using cascade classifier with transition probabilities, unconstrained</td>
<td>76.81</td>
<td>76.81</td>
<td>76.81</td>
<td>76.81</td>
</tr>
<tr>
<td>(l) SPREAD II: using cascade classifier with transition probabilities, constrained</td>
<td>75.52</td>
<td>75.52</td>
<td>75.52</td>
<td>75.52</td>
</tr>
<tr>
<td>(m) SPREAD II: compound classifier, unconstrained</td>
<td>76.81</td>
<td>76.81</td>
<td>76.81</td>
<td>76.81</td>
</tr>
<tr>
<td>(n) SPREAD II: compound classifier, constrained</td>
<td>75.52</td>
<td>75.52</td>
<td>75.52</td>
<td>75.52</td>
</tr>
<tr>
<td>(o) SPREAD II: unconstrained</td>
<td>73.84</td>
<td>73.84</td>
<td>73.84</td>
<td>73.84</td>
</tr>
<tr>
<td>(p) SPREAD II: constrained</td>
<td>69.20</td>
<td>69.20</td>
<td>69.20</td>
<td>69.20</td>
</tr>
</tbody>
</table>

**Tab. 5.1:** Classification accuracy of classifiers that include temporal context in "Horsham (RC) Bat".
5. Producing a Time Series of Maps Using SPREAD II

<table>
<thead>
<tr>
<th>LU code</th>
<th>Land Use</th>
<th>AgStats area 2005</th>
<th>CLUM area 2010</th>
<th>AgStats area 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>210</td>
<td>Pasture</td>
<td>141640</td>
<td>153331</td>
<td>134538</td>
</tr>
<tr>
<td>310</td>
<td>Agroforestry</td>
<td>283</td>
<td></td>
<td>58</td>
</tr>
<tr>
<td>330</td>
<td>Other cropping</td>
<td>170117</td>
<td>189981</td>
<td>173576</td>
</tr>
<tr>
<td>333</td>
<td>Hay + silage</td>
<td>12997</td>
<td></td>
<td>8402</td>
</tr>
<tr>
<td>334</td>
<td>Oil seeds</td>
<td>15609</td>
<td>2573</td>
<td>23979</td>
</tr>
<tr>
<td>340</td>
<td>Perennial horticulture</td>
<td>0</td>
<td>198</td>
<td>0</td>
</tr>
<tr>
<td>341</td>
<td>Tree fruits</td>
<td>563</td>
<td></td>
<td>339</td>
</tr>
<tr>
<td>349</td>
<td>Grapes</td>
<td>10</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>354</td>
<td>Seasonal vegetables &amp; herbs</td>
<td>16</td>
<td>297</td>
<td>14</td>
</tr>
</tbody>
</table>

Tab. 5.2: LU areas reported in the AgStats and validation data for region “Horsham (RC) Bal”.

Fig. 5.7: Results for raw classifier in “Murrumbidgee (A)” in 2010.

5.4 Simulation Study

As described in Section 2.5, there are many issues with the data we have available for both inputs to the analyses, and validation of the results. Further, as described in Section 5.2.3 the method we have used to estimate a transition matrix is not ideal. Here, we generate synthetic datasets based on a known transition matrix in order to assess the merits of the approaches presented herein, in the presence of non-random LU transitions, using the same technique described in Section 2.5.

We ignore biases in the AgStats and use data equivalent to that used for scenario one of Section 2.5 (corresponding to the top panels of Figures 2.10 and 2.11). In this scenario, the simulation is run on a ‘region’ of 20×20 pixels, each LU has the same number of pixels, and the areal constraint data accurately reflects this. We use (2.10) to generate synthetic end members by varying $\alpha$ between zero and one in each simulation, using the same end members in all scenarios, but generating the synthetic profiles separately for each. To make the simulation as
comparable as possible to scenario one of Section 2.5, we use the same end members shown in Figure 2.9. The transition matrix used in this analysis was chosen to be qualitatively similar in structure to that we estimated from the AgStats data, and is shown in Figure 5.8.

We only consider the classification in the second period, as for all but the compound classifier, the results will be identical in the first period.

![Graph showing percent correctly matched for profiles generated using (2.10) for various values of α.](image)

**Fig. 5.9: Percent correctly matched for profiles generated using (2.10) for various values of α.**

We can see from Figure 5.9 that including the transition matrix in the classifier is improving its accuracy by around 10 percent for moderate values of $\alpha$. When $\alpha$ is small, we get no improvement, because the pixels are poorly classified based on the imagery alone. In this case, we are often transitioning from a randomly chosen LU in the first period. As $\alpha$ gets large, the imagery alone is enough to determine the correct LU.

The exception to this is the raw classifier that uses compound classification without the transitions. In this case, for pixels which change LU between the two periods, the posterior probability of the LU in the first period dominates the prior for the second period, and the pixels that do change LU are often misclassified. While interpreting this result it is important to note that even though, when $\alpha$ is equal to one and the NDVI profiles are pure, while the raw classifier is able correctly classify them in every case, it does not assign a class membership probability of one to the true LU. For example, for the end members used here, the most extreme case is given by LUs one and nine. The raw classifier gives a probability of approximately 0.11 for LU one based on the profiles for LU nine and vice versa. Hence, in the case of the raw classifier the posterior in the second period is determined by the probabilities assigned by the raw classifier to all LUs, not just true LUs in the two periods and the transition probabilities. The degree to which this occurs will depend on the properties of the raw classifier and the specific end members.

Figure 5.10 shows broadly the same results as Figure 5.9, but more clearly demonstrates one of the nice properties of SPREAD II; that it performs well when used in a cascade classification
5. Producing a Time Series of Maps Using SPREAD II

5.5 Discussion

In the results presented here, the compound classifier performs marginally better than any of the other methods explored in this chapter and vanilla SPREAD II. Of the other classifiers, vanilla SPREAD II performs best, though only marginally better than the cascade classifiers. The simple Bayesian approach also performs comparatively well, giving match rates very similar to the corresponding SPREAD II variants. An important point to note, is that the inclusion of the transition probabilities in the raw classifier actually degrades their performance, supporting the hypothesis that these methods depend on the performance of the underlying classifier.

Extensions to longer sequences of maps are straightforward, mathematically at least. This could be done by applying the one-step transitions used herein between each pair of maps in a sequence, or the transition probabilities could include probabilities of longer sequences of LUs, allowing for much richer probability structures. One very appealing application of this would be to exclude LU sequences if they are extremely unlikely or impossible, of which there are potentially many. For example, it would be extremely unlikely that one would observe LU sequences such as: cotton to plantation fruit to rice over three consecutive periods. Another appealing application of the compound classifier is to leverage information from periods for which high quality areal constraint data exists into periods for which no areal constraints available or the areal constraints are of lower quality and/or published for different regionalisations.\(^2\) This could

\(^2\) This is the case in Australia, where an agricultural census is published every five years and published by SLA, and in other years, surveys are conducted and published by Statistical Division (SD)—which comprise of multiple SLAs.
be achieved by ‘sandwiching’ one or more low quality periods between high quality periods and classifying them simultaneously.

We note that while the incorporation of transition probabilities is relatively straightforward mathematically once a transition matrix has been obtained, actually obtaining a transition matrix is not. With the dearth of control site data available for the production of maps, it is hard enough to train classifiers for single time points, let alone estimate multi-temporal parameters such as transition probabilities. It is also likely the transition probabilities will vary geographically, and one would ideally only use control sites from the same time period and which are geographically close to the region of interest in their estimation. This is particularly problematic in broad scale applications with many LU classes, where number of control sites would be increased dramatically. To put this in context, in the case of the LU classification used in some of the previous NLUMs, there are 42 LU categories giving rise to over 1,700 possible LU sequences in two periods over 74,000 in three.

Approaches to solving this problem have been suggested. For example, Kittler and Foglein (1984) proposed a method for enumerating all spatial configurations of pixels. A temporal analogue of their approach would be to develop two lists: one containing all sequences that have probabilities of occurrence above a certain threshold, and another containing all impossible sequences. Sequences not on either of these lists would all be assigned the same probability, such that the sum of all probabilities was one. This may perform reasonably well in regions between times where total LU change is relatively small, or the set of likely transitions is small. Bruzzone and Serpico (1997) noted that it would be hard to obtain reasonable estimates for all possible transitions from experts and analyzing historical databases, and Instead, devise an iterative scheme for LC change detection in which they estimate the transition probabilities iteratively in conjunction with the LC for all pixels and both time points.

The use of the approximations developed in this chapter significantly increases the computation required to produce a map. The complexity of SPREAD II is $O(MNRK)$, where $M$ is the number of maps, $N$ is the number of pixels in each map, $R$ is the number of iterations (including the burn-in) and $K$ is the number of LUs. This becomes $O(MNRK^2)$ when we use heterogeneous priors and the cascade classification algorithm described above, and $O(NRK^{2M+1})$ when we use the compound classification algorithm. For our full set of test regions, this resulted in 10 fold increase execution time for the cascade classifier, a 1,600 fold increase for the compound classifier. Hence, while the inclusion of heterogeneous priors using the approximations presented herein avoids the presently insurmountable computational barriers presented by direct evaluation of (5.2), it is not without significant computational cost in its own right, and creating longer sequences of broad-scale maps using these techniques is still likely to be computationally infeasible.

We attempted to address this by developing a Graphical Processing Unit (GPU) implementation of (5.23), which appears to be well suited to calculation using a GPU; it involves a relatively small amount of data (the transition probabilities and the vectors of counts) and what appears to be reasonably high arithmetic intensity when considered over all possible LU transitions. However, the performance of our implementation for 10 LU classes using 32-bit floating point numbers was approximately equal to that being achieved with the Central Processing Unit (CPU) using 64-bit floating point numbers. It is possible that the relative performance may increase with larger numbers of LUs—as the arithmetic intensity should increase—or a more...
expertly crafted implementation.

We have only applied the approximation developed herein to temporal classification techniques. However, the ability to have different categorical distributions at every pixel also lends itself nicely to the incorporation of other ancillary data sources into vanilla SPREAD II, through prior distributions over the possible LU classes specific to each pixel.
6. SUMMARY AND DISCUSSION

6.1 Summary of Results

This chapter presents results for a set of regions suggested by ACLUMP members as being interesting test regions based on the following criteria. Ideally, a region should have:

- CLUM data for the year being validated or an adjacent year,
- a significant amount of agricultural land, and
- a variety of agricultural LUs.

The regions used are shown in Figure 6.1. Some of these regions did not meet the last criteria, but highlight some interesting aspects of the algorithms and were used anyway. Appendix A presents the CLUM and AgStats data for these regions to aid interpretation of the results presented here.

The percentage matches for the regions and average posterior probability of the true class for the algorithms trialled are shown in Figures 6.2 and 6.3. For the analyses which include temporal
Fig. 6.2: Match percentages for various classifiers trialled in this thesis for all test regions.

Fig. 6.3: Posterior probability (×100) of ‘true’ class averaged over the test region for various classifiers trialled in this thesis for all test regions.
context we have only shown results for regions which have CLUM data available for 2010: the second year in our sequence. The colors indicate the magnitude of the match rate with respect to all match rates for the corresponding region; that is, the colours are based on the magnitude of the corresponding value scaled within the region (column). In each column, the lowest value is dark green and the highest is almost white, with intermediate results ranging between these. A row is dominated by green corresponds to a classifier which performs relatively poorly across the regions and a row dominated by white corresponds to a classifier that performs relatively well.

The classifiers that perform worst are the raw classifiers that do not incorporate the AgStats and do not use Algorithm 2. Of the other classifiers, the simple Bayesian approach (the variants labelled “Raw classifier: using AgStats as priors . . .”) and SPREAD II, with the regional constraints included, perform best. The ordering is approximately the same for the average posterior probability of the true class. The results of all classifiers that include the AgStats in some manner are almost identical in regions that are dominated by a single LU.

In general, the use of Algorithm 2, which is indicated by the postfix “constrained”, reduces the percentage match rate over the “unconstrained” equivalent. When this algorithm is applied, the results are very similar regardless of the underlying method.

The notable exceptions to these general results are as follows.

1. The simple Bayesian classifier produces the worst result (excluding the raw classifier) in “Murrumbidgee (A)”. This was discussed in Section 2.4.3.1.

2. Quite a different set of classifiers: SPREAD II including the MRF and SPREAD II including the cascade classifier, produce the best results in “Horsham (RC) Bal” in terms of the average posterior probability of the true class. In this region; the areas reported in the AgStats and the CLUM data are quite similar, the subregions where the dominant LUs are located are quite contiguous, and all classifiers (excluding the raw classifier) are doing a reasonable job of assigning the various LUs in approximately the right locations.

Figure 6.4 shows Cohen’s $\kappa$ (Cohen, 1960) and Figure 6.5 shows $100 \times (1 - p)$, where $p$ is the one-sided p-value based on (2.9) using the binomial approximation described in Section 2.4.3.1. These figures allow us to gain some impression of where the methods are improving over a random allocation based on the areal totals reported in the AgStats, and whether that improvement is statistically significant.

In Figure 6.4, a negative value indicates that the method is performing worse than a random allocation. This occurs regularly in regions that are dominated by a single LU (where the random allocation does particularly well). Using this metric, the SPREAD methods and those that include temporal context are generally performing best. The clear distinction between the performance of the constrained and unconstrained methods seen in Figure 6.2 is no longer apparent, reflecting that constrained methods get more pixels of the less common LUs correct, at the cost of reducing the overall percentage match rate.

$^1$ There is no difference between “constrained” and “unconstrained” for the average posterior probability of the true class.

$^2$ Note that these figures are rounded to the nearest integer, so values of 100 in Figure 6.5 correspond to p-values of less than 0.005.
In Figure 6.5, a value of over 95 indicates a classifier that is statistically significant, and values of less than 50 indicate that the classifier is performing worse than a random allocation. Here, again, we can see that the SPREAD methods, some of the methods that include temporal context, and the constrained methods, are performing better relative to the others.

---

There is a one-to-one correspondence between negative values in Figure 6.4 and values less than 50 in Figure 6.5.
### 6. Summary and Discussion

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Northwest (N)</th>
<th>Central (C)</th>
<th>Northeast (NE)</th>
<th>South-Central (SC)</th>
<th>South-East (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw classifier: using cascade classifier, constrained</td>
<td>100</td>
<td>69</td>
<td>79</td>
<td>100</td>
<td>75</td>
</tr>
<tr>
<td>Raw classifier: using cascade classifier, unconstrained</td>
<td>100</td>
<td>68</td>
<td>79</td>
<td>100</td>
<td>73</td>
</tr>
<tr>
<td>Raw classifier: AgStats as priors, constrained</td>
<td>100</td>
<td>67</td>
<td>79</td>
<td>100</td>
<td>71</td>
</tr>
<tr>
<td>Raw classifier: AgStats as priors, unconstrained</td>
<td>100</td>
<td>65</td>
<td>79</td>
<td>100</td>
<td>69</td>
</tr>
</tbody>
</table>

Fig. 6.5: $100 \times (1 - p)$, where $p$ is a p-value, for various classifiers trialled in this thesis for all test regions.
6.2 Discussion

The challenge of providing systematic and cost effective broad scale monitoring of LU over a continent the size of Australia requires significant automation and the incorporation of multiple data sources into the final product. The focus of the work undertaken in this thesis has been to investigate the potential of the algorithm SPREAD II, and extensions thereof, to address this challenge and some of the issues identified during previous work.

SPREAD II was developed as a replacement for SPREAD. The posterior distribution formally expresses uncertainties associated with the input data and the assumed multinomial prior and hence, in theory at least, represents an improvement on the SPREAD method. Our empirical results suggest that vanilla SPREAD II performs at least as well as if not marginally better than SPREAD.

Vanilla SPREAD II incorporates estimates of areal totals for each LU, a categorical prior, and a probabilistic classifier. The inclusion of sub-regional constraints allows for refining the areal totals based on where broader classes of agricultural LU occur: for example, irrigated agriculture. The inclusion of spatial context through a MRF, or by adding the profiles of partially labelled neighbouring pixels to the training sample, is intended to take account of local phenology, climatic conditions and management practices over the period being mapped. For finer scale imagery (i.e. with resolution that is higher than the scale at which the LUs in the classification scheme occur) it would also hopefully capture aspects of the scale of specific LUs. Finally, the compound and cascade classifiers include information about the likelihood of LU change through the introduction of transition probabilities and allow information from all time periods to contribute to the classifications at each time period.

These methods have been benchmarked against the probabilistic classifier upon which SPREAD II and its extensions are built and simple Bayesian modifications to the classifier upon which they are built that incorporate the same data. Which methods perform best depends on the region and which method of validation is chosen.

Based on percentage match rates and the average posterior of the true class, the simple Bayesian approach performs consistently well across the regions we have considered and in our simulation studies. The one notable exception, is the study region “Murrumbidgee (A)” in 2005. Here, the NDVI profiles contained in the control site database are not representative of those observed in the region in 2005. The simple Bayesian approach appears to be robust against ‘biases’ in the AgStats and is conceptually simple, well understood and cheap to compute. The modifications to this method that include zonal constraints and temporal context also perform consistently well across our regions. These are the best performing classifiers of all those considered, in terms of percentage match rates, and the former, at least, does not suffer nearly as badly from the unrepresentative NDVI profiles contained in the control site database in the region “Murrumbidgee (A)” in 2005.

All the SPREAD II based classifiers perform consistently well across our test regions based on percentage match rates and the average posterior of the true class. In the simulation study presented in Section 2.5, however, we saw that vanilla SPREAD II can perform poorly if the AgStats data are biased. As was noted in that section, and is the case in our study regions, such pathological areal constraint data are unlikely occur frequently in practice and would
probably be detected by an analyst or reviewer in any case. Conversely, these classifiers offer stronger protection against unrepresentative NDVI profiles, as is demonstrated in our study region “Murrumbidgee (A)” in 2005. This last point is discussed further below. The SPREAD II based classifiers are, however, much more memory intensive and expensive to compute, which limits either the geographic extent or the resolution of the imagery to which they can be applied for a given geographic extent.

Based on Cohen’s $\kappa$ and p-values, the SPREAD II based classifiers and those that incorporate Algorithm 2 perform better, relative to the others, than was the case for percentage match rates and average posterior of the true class. These metrics take into account the number of pixels of each LU class that are correctly classified, rather than the aggregate over all LU classes. Qualitatively, they penalise regions that are ‘easy’—which is the case, for instance, when a region is dominated by a single LU—more than the percentage match rate and average posterior of the true class. The improvement we observe across our test regions likely reflects the phenomena noted by Carfagna and Gallego (2005) and Strahler (1980) (discussed near to bottom of Section 2.2.2), that when the priors are based on proportional probability, large classes tend to be overestimated and small classes underestimated. Where a region is dominated by a single LU, overstating the prevalence of that LU is likely to improve the percentage match rate.

6.3 Conclusions

What is clear from the results presented in this thesis is that inclusion of areal constraint data can lead to large improvements in classification accuracy. This is the case regardless of the manner in which it is incorporated. It is also clear that inclusion of the zonal constraints is beneficial. These results are not surprising. These data provide additional information to that contained in the remotely sensed imagery, and as such, are not dependent on the accuracy of the underlying classifier $([Y_i|C_i])$.

It is less clear is that including spatial or temporal context is beneficial. These methods depend on the accuracy of the underlying classifier; if that is poor, they are, at best, unlikely to improve classification accuracy and may, at worst, reduce it. In the test region “Horsham (RC) Bal”, used in Chapter 5, the compound classifier improved on vanilla SPREAD II, but only marginally. These classifiers appear to perform well across our study regions, but do not generally outperform the simpler classifiers. Bearing in mind that they are much more computationally expensive than both the other SPREAD II based classifiers and, in particular, the simple Bayesian approaches, the marginal gains we have observed on the present data do not seem sufficient to warrant their employment. The simulation study presented in Section 5.4 shows that including temporal context can improve classification accuracy if the underlying classifier is moderately informative and the transition matrix is known. However, as discussed previously, obtaining a reasonable approximation to the transition matrix for a region is likely to be difficult.

The qualitative appeal of the results presented for the MRF classifier in Figure 3.1—specifically, the spatial clustering and the similarity of the results for the constrained and unconstrained classifiers—suggests that this method should not, however, be dismissed without further assessment and discussion with experts on whether these qualitative aspects are important. It is also
likely that the methods that include spatial context would perform better on higher resolution imagery.

Obtaining full value from these techniques requires adequate control site information that is temporally and spatially close to the period and location being mapped. A limitation we have faced in our use of SPREAD II is the sparsity of our library of control sites. Though we have control sites representing LUs in a number of different years, the density of this information is low and its currency limited, which is an impediment to creating good classifiers that capture local climatic and spectral variability.

In the work presented here the imagery used had a coarse resolution, relative to the scale of many of the LU classes in the adopted classification scheme, and had minimal power to discriminate between them. Our results suggest that NDVI traces over a single growing season are poor classifiers for crops with similar life histories; it is only possible to distinguish between some summer and winter crops, and some perennial agriculture and annual cropping. The successful creation of LU maps requires remote sensing products that are reasonably capable of differentiating between the classes of interest. Use of areal constraints or other ancillary data is not a panacea for poor classifiers. While it will provide protection from serious error when included in appropriate techniques, it cannot manufacture additional information.

The promise of SPREAD II it is that it affords us some protection against truly awful classification accuracy when the satellite imagery is not informative, or worse, is misleading. In cases where the imagery cannot differentiate between two LUs, SPREAD II’s strong conformance to the constraints is likely to protect against serious error at the regional scale, distributing the probability across the LU classes in proportion to the areal totals reported in the AgStats. While this is an accurate statistical representation, whether it is practically useful depends on the application. When the distribution is reduced to the most likely LU, it is unlikely to be.

We assess that while methodological extensions to SPREAD II could improve the results marginally, remote sensing products with greater resolution and discriminating power are needed to significantly advance the accuracy of the final products, especially using the ambitious classification schemes targeted herein. With the advent of such data, designing better classifiers and/or optimising relative expenditures on imagery, classifier development, and control site acquisition seem appropriate. Having relatively cheap, automated algorithms, such as SPREAD II, that incorporate a wide range of data and can be run across broad regions, can assist in this process by reducing labour costs and allowing for relatively easy customisation.

It has been noted that little improvement was observed in the accuracy of satellite image classification experiments over the fifteen years up to 2004 (Wilkinson, 2005). It is hard, therefore, to assert that technological advancements in classifiers have improved accuracy. Whether or not this has changed since 2005, it is hard to imagine that this track record will not be improved upon in the near future. The current rapid increase in affordable computational resources, the development of new, more accurate satellite sensors and, in particular, the rapid expansion in the availability and affordability of relatively fine scale satellite imagery, will enable more powerful primary classifiers. Administrative data streams are increasing, and citizen science can potentially provide more cost effective ground truth and training data. In light of this, the future looks exciting for informatics based land use mapping and SPREAD and SPREAD II are two steps in a long and interesting journey toward cost effective, repeatable broad scale
land use mapping.
6.4 Further Research

Through the course of this work several areas for further research became apparent.

6.4.1 Applying the Techniques in Concert

We have considered the extensions to SPREAD II in isolation, which has allowed assessment of the merits of each separately. Further, given that the inclusion of neither spatial nor temporal context appeared to improve performance significantly using the combination of classification scheme and satellite imagery employed herein, it is unlikely that we would have seen improvement in combining these techniques. However, this may not be the case for other data or classification schemes.

Some combinations of the techniques presented here are mutually exclusive. For example, the development of the MRF approach presented in Section 4.2.1 depends on all pixels in a map having the same categorical prior. Theoretically, this precludes its with the cascade classifier. It would be interesting, however, to explore whether the violations of the assumptions underlying the development of the MRF approach would preclude its use in practice.

6.4.2 Producing Maps for Non-Census Years

The main reason that NLUMs are not produced for non-census years is that in those years the ABS does not release data at the same spatial or categorical level of detail. Both theoretically, and in the current implementation of SPREAD II, the regionalisations used at each time point need not be identical when including temporal context, and hence differences in spatial resolution can be accommodated. If the coarser categories for which estimates are produced are partitioned by the finer categories that are produced in census years, it would be possible to apply constraints at the coarser categorisation in the non-census years, while continuing to apply them at the finer one in the census years.

This would require transition matrices where entry $\psi_{kl}$ expresses the probability of a transition from LU $k$ to aggregated LU class $l$ in a census to non-census year, from aggregated LU class $k$ to aggregated LU class $l$ in a non-census to non-census year, and from aggregated LU class $k$ to LU $l$ in a non-census to census year. These could be derived simply by summing the appropriate rows/columns in ‘full’ transition matrix.

6.4.3 Inclusion of Textures or Objects

We have focused on how to include additional sources of data with an underlying classifier and have not focused at all on the properties of that classifier. We have implicitly assumed that the method used to develop the underlying classifier is mathematically independent of the methods we have explored. Of particular importance is the assumption that $[Y_i|C_i]$ is independent of all other pixels. We relax this assumption when we include spatial context, though we only explore a small subset of possible methods for doing so. Other methods for including spatial context may be simpler to implement and perhaps more robust, especially if they can be included in the context of the simpler Bayesian classifiers.
For instance, one could include image textures in the feature vector, \( Y_i \). This would allow the inclusion of spatial context in the simple Bayesian approaches. It also allows the spatial patterns to be estimated easily from the control sites and may obviate the need for inclusion of a local statistical model as was required for the methods presented in Chapter 4. Another possibility would be to use blob detection (Lindeberg, 1994) techniques and base the neighbourhoods used in Chapter 4 on the pixels within the same blob as the pixel under consideration.

These methods also have great potential in that they allow the incorporation of other remotely sensed datasets that may not be suitable for LU classification by themselves. For example, Landsat data have a much more appropriate resolution for discrimination of agricultural regions, as their resolution is below that of many fields. However, the 16-day return times, combined with the frequency of cloud cover, make temporal NDVI profiles sparse, especially over regions suitable for agricultural activity. However, relatively cloud free mosaics are often feasible over the period of a growing season, from which the textures could be calculated and resampled to the scale of the satellite imagery being employed in the underlying classifier.

6.4.4 Incorporation of Multi-resolution Data and Sub-pixel Classification

It would be straightforward to adapt SPREAD II to incorporate data of varying spatial scales, as the constraints are based on areas rather than pixel counts and only the total area allocated to each LU needs to be tracked. This could be one way of reducing the computational overhead of running SPREAD II and render it suitable for finer scale mapping. One simple approach would be to use finer resolution data, such as Landsat, via image textures or alike, to identify where small-scale agriculture occurs and use higher resolution imagery for those regions in parallel to coarse scale imagery in regions where broad-scale agriculture occurs.

Many methods of “pixel un-mixing” have been presented in the literature. While we have not reviewed that literature in detail, it may be possible to apply areal constraints in the context of these methods, since only the total area allocated to each LU within a region can be determined from a weighted sum of each pixels area, and its fractional area assigned to each LU.
Bibliography


John Richardson. Personal communication, March 2013.


## Appendix A

### TEST REGIONS

<table>
<thead>
<tr>
<th>Land Use</th>
<th>LU code</th>
<th>AgStats area</th>
<th>CLUM area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pasture</td>
<td>210</td>
<td>203280</td>
<td>237863</td>
</tr>
<tr>
<td>Agroforestry</td>
<td>310</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>Other cropping</td>
<td>330</td>
<td>70091</td>
<td>58229</td>
</tr>
<tr>
<td>Hay &amp; silage</td>
<td>333</td>
<td>2566</td>
<td></td>
</tr>
<tr>
<td>Oil seeds</td>
<td>334</td>
<td>1964</td>
<td></td>
</tr>
<tr>
<td>Cotton</td>
<td>336</td>
<td>2140</td>
<td></td>
</tr>
<tr>
<td>Tree fruits</td>
<td>341</td>
<td>568</td>
<td>509</td>
</tr>
<tr>
<td>Citrus</td>
<td>348</td>
<td>279</td>
<td></td>
</tr>
<tr>
<td>Grapes</td>
<td>349</td>
<td>328</td>
<td></td>
</tr>
<tr>
<td>Seasonal vegetables &amp; herbs</td>
<td>354</td>
<td>233</td>
<td>813</td>
</tr>
<tr>
<td>Rice</td>
<td>430</td>
<td>20483</td>
<td></td>
</tr>
</tbody>
</table>

**Fig. A.1:** “Murrumbidgee (A)” and LU areas reported in the AgStats and validation data.

<table>
<thead>
<tr>
<th>Land Use</th>
<th>LU code</th>
<th>AgStats area</th>
<th>CLUM area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pasture</td>
<td>210</td>
<td>134538</td>
<td>153531</td>
</tr>
<tr>
<td>Agroforestry</td>
<td>310</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>Other cropping</td>
<td>330</td>
<td>173576</td>
<td>189981</td>
</tr>
<tr>
<td>Hay &amp; silage</td>
<td>333</td>
<td>8402</td>
<td>573</td>
</tr>
<tr>
<td>Oil seeds</td>
<td>334</td>
<td>23979</td>
<td>2573</td>
</tr>
<tr>
<td>Perennial horticulture</td>
<td>340</td>
<td>0</td>
<td>198</td>
</tr>
<tr>
<td>Tree fruits</td>
<td>341</td>
<td>339</td>
<td></td>
</tr>
<tr>
<td>Citrus</td>
<td>348</td>
<td>273</td>
<td></td>
</tr>
<tr>
<td>Grapes</td>
<td>349</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Seasonal vegetables &amp; herbs</td>
<td>354</td>
<td>14</td>
<td>297</td>
</tr>
</tbody>
</table>

**Fig. A.2:** “Horsham (RC) Bal” and LU areas reported in the AgStats and validation data.

<table>
<thead>
<tr>
<th>Land Use</th>
<th>LU code</th>
<th>AgStats area</th>
<th>CLUM area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pasture</td>
<td>210</td>
<td>20483</td>
<td></td>
</tr>
<tr>
<td>Agroforestry</td>
<td>310</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>Other cropping</td>
<td>330</td>
<td>171654</td>
<td>30657</td>
</tr>
<tr>
<td>Hay &amp; silage</td>
<td>333</td>
<td>6720</td>
<td>30657</td>
</tr>
<tr>
<td>Oil seeds</td>
<td>334</td>
<td>3736</td>
<td>30657</td>
</tr>
<tr>
<td>Cotton</td>
<td>336</td>
<td>6650</td>
<td>4330</td>
</tr>
<tr>
<td>Tree fruits</td>
<td>341</td>
<td>273</td>
<td></td>
</tr>
<tr>
<td>Citrus</td>
<td>348</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Grapes</td>
<td>349</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Seasonal vegetables &amp; herbs</td>
<td>354</td>
<td>4</td>
<td>107</td>
</tr>
</tbody>
</table>

**Fig. A.3:** “Gwydir (A)” and LU areas reported in the AgStats and validation data.
### Land Use

<table>
<thead>
<tr>
<th>Land Use</th>
<th>LU code</th>
<th>AgStats area</th>
<th>CLUM area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pasture</td>
<td>210</td>
<td>122927</td>
<td>126133</td>
</tr>
<tr>
<td>Agriculture</td>
<td>310</td>
<td>1202</td>
<td></td>
</tr>
<tr>
<td>Other cropping</td>
<td>330</td>
<td>2163</td>
<td>683</td>
</tr>
<tr>
<td>Hay &amp; silage</td>
<td>333</td>
<td>12528</td>
<td></td>
</tr>
<tr>
<td>Oil seeds</td>
<td>334</td>
<td>376</td>
<td></td>
</tr>
<tr>
<td>Other crops</td>
<td>341</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Grapes</td>
<td>349</td>
<td>159</td>
<td></td>
</tr>
<tr>
<td>Seasonal vegetables &amp; herbs</td>
<td>354</td>
<td>357</td>
<td></td>
</tr>
</tbody>
</table>

**Fig. A.4:** “Glenelg (S) - Heywood” and LU areas reported in the AgStats and validation data.

<table>
<thead>
<tr>
<th>Land Use</th>
<th>LU code</th>
<th>AgStats area</th>
<th>CLUM area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pasture</td>
<td>210</td>
<td>45673</td>
<td>7749</td>
</tr>
<tr>
<td>Other cropping</td>
<td>330</td>
<td>232456</td>
<td>303158</td>
</tr>
<tr>
<td>Hay &amp; silage</td>
<td>333</td>
<td>12389</td>
<td></td>
</tr>
<tr>
<td>Oil seeds</td>
<td>334</td>
<td>17287</td>
<td>300</td>
</tr>
</tbody>
</table>

**Fig. A.5:** “Yarriambiack (S) - North” and LU areas reported in the AgStats and validation data.

<table>
<thead>
<tr>
<th>Land Use</th>
<th>LU code</th>
<th>AgStats area</th>
<th>CLUM area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pasture</td>
<td>210</td>
<td>375228</td>
<td>88378</td>
</tr>
<tr>
<td>Agriculture</td>
<td>310</td>
<td>158</td>
<td></td>
</tr>
<tr>
<td>Other cropping</td>
<td>330</td>
<td>462760</td>
<td>760636</td>
</tr>
<tr>
<td>Hay &amp; silage</td>
<td>333</td>
<td>5599</td>
<td></td>
</tr>
<tr>
<td>Oil seeds</td>
<td>334</td>
<td>13920</td>
<td></td>
</tr>
<tr>
<td>Perennials &amp; horticulture</td>
<td>340</td>
<td>13</td>
<td>2452</td>
</tr>
<tr>
<td>Tree fruits</td>
<td>341</td>
<td>175</td>
<td></td>
</tr>
<tr>
<td>Tree nuts</td>
<td>343</td>
<td>1541</td>
<td></td>
</tr>
<tr>
<td>Citrus</td>
<td>348</td>
<td>2470</td>
<td></td>
</tr>
<tr>
<td>Grapes</td>
<td>349</td>
<td>4097</td>
<td>2245</td>
</tr>
<tr>
<td>Seasonal vegetables &amp; herbs</td>
<td>354</td>
<td>334</td>
<td>4063</td>
</tr>
</tbody>
</table>

**Fig. A.6:** “Mildura (RC) - Pt B” and LU areas reported in the AgStats and validation data.

<table>
<thead>
<tr>
<th>Land Use</th>
<th>LU code</th>
<th>AgStats area</th>
<th>CLUM area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pasture</td>
<td>210</td>
<td>136687</td>
<td>134713</td>
</tr>
<tr>
<td>Agriculture</td>
<td>310</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td>Other cropping</td>
<td>330</td>
<td>11226</td>
<td>13051</td>
</tr>
<tr>
<td>Hay &amp; silage</td>
<td>333</td>
<td>5822</td>
<td>1374</td>
</tr>
<tr>
<td>Oil seeds</td>
<td>334</td>
<td>2078</td>
<td>196</td>
</tr>
<tr>
<td>Tree fruits</td>
<td>341</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Grapes</td>
<td>349</td>
<td>311</td>
<td></td>
</tr>
<tr>
<td>Seasonal vegetables &amp; herbs</td>
<td>354</td>
<td>239</td>
<td>294</td>
</tr>
</tbody>
</table>

**Fig. A.7:** “Wattle Range (DC) - West” and LU areas reported in the AgStats and validation data.
## Appendix A. Test Regions

### Land Use

<table>
<thead>
<tr>
<th>Land Use</th>
<th>LU code</th>
<th>AgStats area</th>
<th>CLUM area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pasture</td>
<td>210</td>
<td>240304</td>
<td>215686</td>
</tr>
<tr>
<td>Agroforestry</td>
<td>310</td>
<td>320</td>
<td>331</td>
</tr>
<tr>
<td>Other cropping</td>
<td>330</td>
<td>13873</td>
<td>29746</td>
</tr>
<tr>
<td>Hay &amp; silage</td>
<td>333</td>
<td>1606</td>
<td>277</td>
</tr>
<tr>
<td>Oil seeds</td>
<td>334</td>
<td>437</td>
<td></td>
</tr>
<tr>
<td>Tree fruits</td>
<td>341</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Grapes</td>
<td>349</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Seasonal vegetables &amp; herbs</td>
<td>354</td>
<td>2476</td>
<td></td>
</tr>
</tbody>
</table>

Fig. A.8: “Northern Midlands (M) - Pt B” and LU areas reported in the AgStats and validation data.

### Land Use

<table>
<thead>
<tr>
<th>Land Use</th>
<th>LU code</th>
<th>AgStats area</th>
<th>CLUM area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pasture</td>
<td>210</td>
<td>11734</td>
<td>10228</td>
</tr>
<tr>
<td>Agroforestry</td>
<td>310</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>Other cropping</td>
<td>330</td>
<td>0</td>
<td>93</td>
</tr>
<tr>
<td>Hay &amp; silage</td>
<td>333</td>
<td>1858</td>
<td></td>
</tr>
<tr>
<td>Perennial horticulture</td>
<td>340</td>
<td>0</td>
<td>93</td>
</tr>
<tr>
<td>Tree fruits</td>
<td>341</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Tree nuts</td>
<td>343</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>Grapes</td>
<td>349</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td>Seasonal vegetables &amp; herbs</td>
<td>354</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Fig. A.9: “West Tamar (M) - Pt B” and LU areas reported in the AgStats and validation data.