School of Earth and Planetary Sciences (EPS)
Remote Sensing and Satellite Research Group (RSSRG)

Predicting And Mapping Of Honey Flow From *Corymbia Calophylla* Trees With Remote Sensing

Tristan Campbell

0000-0002-3796-9582

This thesis is presented for the Degree of

Doctor of Philosophy

of

Curtin University

September 2020

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Tristan Campbell

Date: 22/09/2020

Abstract

Pollination is vital for ecosystem health, with 60 – 90% of plant species globally requiring animal pollination. This has a direct impact on humans, with over 80% of all crops depending on insect pollination to some extent. This is estimated to be worth a total of €153 billion annually (approximately 9.5% of the total production value of all crops). Honeybees (*Apis mellifera*) have a globally significant role in crop pollination and are the most economically valuable pollinator for many monoculture crops. With honeybee nutrition a key factor in pollination efficacy, Western Australian apiarists heavily utilise the *Corymbia calophylla* forests to both manage the condition of their bees as well as produce some of the highest antimicrobial honey in the world. As these forests cover an area of more than 84,000 km^2, apiarists travel thousands of kilometres every year to inspect apiary sites and move bees between optimal locations during the *Corymbia calophylla* flowering period of February to March.

As mature *Corymbia calophylla* trees are over 10 m tall and honeybees typically seek food sources up to 1 km from their hives, it is not feasible to accurately determine the flower abundance of an apiary site from the ground. With the rapid increased availability of high-quality, recreational UAVs in recent years, an algorithm has been developed to measure relative differences in flower abundance both between apiary sites and over the flowering period. The recommended UAV survey specifications from a clustering analysis of *Corymbia calophylla* images using standard red-green-blue multispectral sensors can classify flower abundance to 90% accuracy.

To assist apiarists with apiary site assessment remotely, analysis of ground-based spectroradiometer data was done to determine the optimum satellite-based sensor to detect changes in *Corymbia calophylla* flowers. The Marri Flowering Index was developed based on the spectral bands with the highest sensitivities to flower abundance. This index is calculated by dividing the green visual band data by the near-UV band (Band 10 / Band 8 for MODIS for example). The index was able to reliably discriminate good honey producing years versus moderate to poor years for a mature *Corymbia calophylla* forest site near Perth, Western Australia. However, further work is required to assess the index's efficacy in areas of lower forest coverage and its ability to measure changes in flower abundance during a single season.

To develop a predictive model of honey harvest for a given apiary site, honey harvest data over 8 years from 16 apiary sites were compared with vegetation indices. The MODIS NDVI

was reliably able to discriminate between good versus moderate to poor harvest yearsusing a Decision Tree model but unable to distinguish between moderate and poor years. Addition of weather data for the month immediately prior to flowering improved the discrimination of good harvest years and was able to discriminate between good, moderate and poor years with 70% accuracy.

With the combination of satellite-derived vegetation indices and weather proven able to classify honey harvest quality, machine learning algorithms were applied to larger set of weather and satellite vegetation index data to create a quantitative prediction model. After several stages of analysis, a Gradient Boosted Regression model was developed, capable of predicting the honey harvest per hive with a Mean Average Error of +/- 8.9 kg, against a harvest range of 0 to 71 kg. This model used the January Evapotranspiration data from the MODIS sensor and the average maximum temperature for November. With *Corymbia calophylla* typically starting in February, an additional Gradient Boosted Regression model was created using only the November average maximum temperature. While the Mean Average Error was higher (+/- 11.7 kg), this model gives apiarists a harvest prediction two months prior to the onset of flowering, giving more time to adjust apiary management plans accordingly with a more accurate prediction given closer to the start of flowering.

While each of the three algorithms developed in this research have been specific to *Corymbia calophylla* trees, the development method for each algorithm can be readily applied to other honey producing species. The machine learning prediction model in particular has used datasets that are widely available across the honey producing areas of the world. As a result, the methodology can be readily used to create predictive models for other species or regions with sufficient availability of honey harvest records.

Acknowledgements

This thesis began several years ago with a conversation with Dr Peter Fearns regarding the potential for a geophysicist and beekeeper to do a doctorate in the remote sensing field. Dr Fearns has been with me every step of the way since, assisting with the transition from geoscience consulting to the academic sector. After many, many edits and reviews, here we are! Without the advice and support from Dr Fearns this thesis would never have even started.

After the work began to progress, it became apparent how many disciplines were involved with both the research and the applications. So thank you to people who helped with their input and review of the papers and this thesis; Dr Kenneth Dods (food scientist with ChemCentre at the WA Honey Research and Development Group), Dr Kingsley Dixon (botanist with the School of Molecular and Life Sciences at Curtin University) and Dr Rebecca Handcock (GIS and data scientist with the Curtin Institute for Computation).

The goal of this thesis has always been to combine my professional geoscience expertise with my pastime of beekeeping to deliver a result to the beekeeping industry. The advice from countless beekeepers, and their decades of experience, has been invaluable to the project, as has support from the Bee Industry Council of Western Australia (BICWA) and the WA Honey Research and Development Group (WAHRDG) and the Cooperative Research Centre for Honey Bee Products (CRC HPB).

And of course the support from my family has been invaluable. From putting up with 'family bushwalks' that doubled as data collection excursions, to being woken up in the middle of the night by me bashing away at the computer after having an idea that has to be put down before I forget it, to the myriad of other trials and tribulations that family members of a PhD candidate endure, your support has been enduring, unwavering and made this possible.

List of Publications Included as Part of the Thesis

Campbell T, Fearns P (2018) Simple remote sensing detection of *Corymbia calophylla* Flowers using common 3 –band imaging sensors. Remote Sensing Applications: Society and Environment, 11, pp. 51- 63. https://doi.org/10.1016/j.rsase.2018.04.009

Campbell T, Fearns P (2018) Honey crop estimation from space: Detection of large flowering events in Western Australian forests. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-1, pp. 79–86. https://doi.org/10.5194/isprs-archives-XLII-1-79-2018

Campbell T, Dods K, Dixon K, Fearns P, Handcock R (2020) Machine Learning Regression Model for Predicting Honey Harvests. Agriculture, 10 (4), pp. 118. https://doi.org/10.3390/agriculture10040118

I warrant that I have obtained, where necessary, permission from the copyright owners to use any third party copyright material reproduced in the thesis, or to use any of my own published work in which the copyright is held by another party. Copyright statements for each publication are presented in Appendix B – Publication Copyright Statements.

Tristan Campbell

Date: 22/092020

Statement of Contribution of Others

Tristan Campbell's contribution to this thesis, and the publications presented herein, included conception of the study, acquisition of field data, identifying and sourcing other data sources, data analysis, drafting of manuscripts and work associated with publication.

Co-authors made contributions of sufficient significance to warrant co-authorship of the resulting articles. Details of these contributions are as follows:

- Dr Peter Fearns provided overall research supervision, critical review of study design and manuscript review for all publications
- Dr Kenneth Dods and Dr Kingsley Dixon provided manuscript review of the third paper, using their expertise in fields complimentary to remote sensing to improve the quality of the publications
- Dr Rebecca Handcock provided manuscript review of the third paper, using her expertise in fields complimentary to remote sensing to improve the quality of the publication

Statements of contribution from co-authors regarding Tristan Campbell's contribution to each publication are provided in Appendix A - Statements of Contribution from Co-Authors.

Tristan Campbell	Dr Peter Fearns
PhD Candidate	Supervisor

List of Abbreviations

ACRONYM	DEFINITION
ADAM	Australian Data Archive for Meteorology
ANOVA	Analysis of Variance
AppEEARS	Application for Extracting and Exploring Analysis Ready Samples
ASD	Analytical Spectral Devices
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
BBCH	Biologische Bundesanstalt, Bundessortenamt, Chemische Industrie
BICWA	Bee Industry Council of Western Australia
BOM	Bureau of Meteorology
BRT	Boosted Regression Tree
CASA	Civil Aviation Safety Authority
CBERS	China-Brazil Earth Resources Satellite
CSIRO	Commonwealth Scientific and Industrial Research Organisation
DT	Decision Trees
DPIRD	Department of Primary Industries and Regional Development
DSLR	Digital Single Lens Reflex
EM	Electromagnetic
ENSO	El Nino-Southern Oscillation
EOS	End of Season
EPBC	Environmental Protection and Biodiversity Conservation
EPS	School of Earth and Planetary Sciences
EROS	Earth Resources Observation Satellites
ET	Evapotranspiration
EVI	Enhanced Vegetation Index
FOV	Field of View

ACRONYM	DEFINITION
FPAR	Fraction of Photosynthetically Active Radiation
GAMLSS	Generalised Additive Model for Location, Scale and Shape
GBR	Gradient Boosted Regression
GIS	Geographic Information System
GLS	Global Land Survey
GPP	Gross Primary Productivity
GSL	Growing Season Length
GVP	Gross Value of Production
IBRA	Interim Biogeographic Regionalisation for Australia
JM Distance	Jeffries-Matusita Distance
KNMI	Royal Netherlands Meteorological Institute
LAI	Leaf Area Index
LE	Average Latent Heat Flux
MAE	Mean Average Error
MFI	Marri Flowering Index
MODIS	Moderate Resolution Imaging Spectroradiometer
MSS	Multi-Spectral Scanner
NASA	National Aeronautics and Space Administration
NDVI	Normalised Difference Vegetation Index
NDWI	Normalised Difference Water Index
NDYI	Normalized Difference Yellowness Index
PCA	Principal Component Analysis
PET	Total Potential Evapotranspiration
PLE	Average Potential Latent Heat Flux
PSN	Net Photosynthesis

ACRONYM	DEFINITION
RGB	Red Green Blue
RMSE	Root Mean Squared Error
RSSRG	Remote Sensing and Satellite Research Group
SAVI	Soil Adjusted Vegetation Index
SD	Standard Deviation
SDGs	Sustainable Development Goals
sMFI	Scaled Marri Flowering Index
SOS	Start of Season
TCI	Temperature Condition Index
UAV	Unmanned Aerial Vehicle
USA	United States of America
UV	Ultraviolet
VCI	Vegetation Condition Index
VI	Variable Importance
VI	Vegetation Index
VIIRS	Visible Infrared Imaging Radiometer Suite
VYT	Variable Yield Technology
WA	Western Australia
WAHRDG	WA Honey Research and Development Group
WGS84	World Geodetic System 1984

Table of Contents

List of Figures

List of Tables

List of Equations

1 General Introduction

1.1 Overview

The beekeeping industry in Western Australia has grown rapidly in the past decade, from 660 registered beekeepers in 2010 to over 3,000 in 2019 (Thomson 2019). In addition, honey produced from Western Australia has some of the highest antimicrobial properties known for honey (Irish et al. 2011). As these high antimicrobial honey varieties are produced from *Corymbia calophylla* (marri) and *Eucalyptus marginata* (jarrah) trees that occur across a large area (see Figure 1-1) of approximately 84,000 km^2 (Herbarium 2015), beekeepers often travel long distances to inspect apiary sites and manage their beehives. Access to a geospatial tool to map and predict areas of higher and lower honey production would make apiary management more efficient and improve industry safety by reducing the amount of rural driving required.

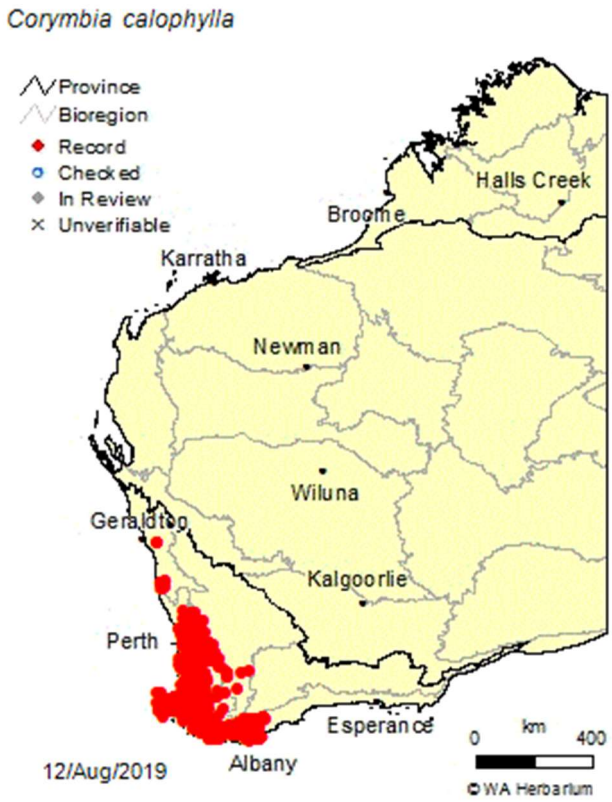


Figure 1-1: Geographic extent of Corymbia calophylla *(Herbarium 2015)*

While there have been several studies into the potential for remote sensing data to provide this information for south-east Australia (for example Webber (2011) and Arundel et al. (2016)), there has been limited work in this research in the south-west of Australia, nor has

there been significant progress in developing the qualitative relationships identified in these studies into a quantitative model.

This thesis presents the research conducted to develop methods to map and predict the volume of *Corymbia calophylla* honey produced per beehive for a given apiary site from a combination of remote sensing and weather station inputs. This includes:

- The development of a drone imagery flower detection algorithm for detailed inspection of specific apiary sites
- Spectral separation analysis of marri flowers compared with their environment
- Development of a machine learning predictive model, the development process of which may be applied to other species or regions around the globe.

This chapter provides general background of the economic and ecological importance of honey bees, some of the recent research pertaining to the medicinal and nutritional properties of honey from Western Australia, an outline of the research presented in this thesis and how the research objectives may assist with improving honey bee health, honey production and assessing ecosystem health.

1.2 Background

Pollination is a vital role in ecosystem services around the world, with 60 – 90% of plant species requiring animal pollination (Kremen et al. 2007). Humans are direct beneficiaries of these pollination services, with over 80% of all crops depending on insect pollination to some extent (Aizen et al. 2019). This was estimated to be worth a total of €153 billion by Gallai et al. (2009), with this figure representing 9.5% of the total production value of all crops grown for consumption by humans. A later study by Lautenbach et al. (2012) showed that while the value of pollinated crops was fairly stable at the start of the 21st century, the value has risen noticeably since 2009 when consistent increases in producer prices began to be seen. In addition to the direct benefits of pollination, Patel et al. (2020b) has suggested that bees can contribute to 15 of 17 of the United Nation's Sustainable Development Goals (SDGs).

While there are over 20,000 different species of insects that contribute to pollination (Lieutier et al. 2017), honey bees (*Apis mellifera*) have a globally significant role in crop pollination (Potts et al. 2016), and are indeed the most economically valuable pollinator for many monoculture crops around the world (Klein et al. 2007).

Considering Australia in particular, Gill (1989) established that the economic value of honey bees was between A$600 million and A$1.2 billion. This figure was updated by Karasinski (2018) to be between A$8.35 billion and A$19.97 billion for the 2014-2015 financial year (1st July 2014 to 30th June 2015), compared with a gross value of production (GVP) for the Australian honey bee industry of A$101 million over the same period (van Dijk et al. 2016). With the economic value of honey bees in Australia being over 100 times the GVP of the honey bee industry, speculation by Voorhies (1933) that the economic value of honey bees is of far greater value than the products from the beehives is well and truly proven as correct.

For Western Australia, which is the focus of this thesis, the economic value of honey bee pollinators is estimated to be between A$800 million and A$1.4 billion for 2014-2015 (Karasinski 2018), compared with total agricultural production of A$5.7 billion in Western Australia for the same period (Value of Agricultural Commodities Produced, Australia, 2014-15 2016). This represents between 14 and 25% of the state's total agricultural production.

With the proportion of crops effectively pollinated by insects linearly related to honey bee population size in or near the producing landscape (Vaissière 1991), it is important to have strong and healthy colonies available for pollination. Honey bee nutrition is crucial for strong and healthy colonies (Haydak 1970), with poor nutrition directly linked to reduced brood production for a given colony size (Sagili and Pankiw 2007) and subsequent reduction in hive population. Healthy diets for honey bee colonies require a minimum suite of 10 different amino acids (De Groot 1952), typically sourced from pollen when honey bee colonies are in a natural environment. However, the nutritional value of pollen varies considerably between species (Somerville 2005). A lack of any of any of these nutritional inputs can lead to a decrease in hive health and therefore reduced pollination services. The effects of a poor nutritional diet for a colony initially results in reduced brood production, followed by cannibalization of young larvae as a source of protein for more developed larvae, then death of the hive from starvation (Brodschneider and Crailsheim 2010).

Research into the nutritional value of pollen from Western Australian plants (Manning 2008) has shown that pollen from marri trees has a high concentration of protein and exceptionally high concentration of some amino acids. More recent research comparing the amino acid concentrations of marri pollen to established requirements for bee health (Kratz et al. 2019) has shown that marri pollen meets all of the amino acid nutritional requirements for honey bees and has high concentrations of three out of the ten essential

amino acids. Therefore, marri pollen is an excellent source of nutrition for honey bees. This can be either for during the honey production season, or marri pollen can be harvested during the marri flowering season, and fed to the bees later in the form of pollen patties (Somerville 2005) prior to commercial pollination contracts to improve their pollination effectiveness.

In addition to the economic value of pollination services to the Western Australian agricultural sector, many honey varieties produced by honey bees from flora endemic to the state can have antimicrobial properties (Roshan et al. 2017). Indeed, multiple studies have shown that honey from marri (*Corymbia calophylla*) and jarrah (*Eucalyptus marginata*) trees often have a higher antimicrobial effect than honey from manuka (*Leptospermum scoparium*) plants, that are endemic to south-east Australia and New Zealand (for example Manning (2011) and Irish et al. (2011)). Jarrah honey has also been found to be significantly more effective against three species of *Candida* fungus than manuka honey and manuka-derived medical products (Irish 2006). The medicinal value of Western Australian honey extends to improved wound treatment. For example, use of honey on horse leg wounds has been reported to reduce the time required for veterinary gels and sterile pads (Camarda 2007).

These high levels of medicinal properties have resulted in the value of honey from *Corymbia calophylla* forests increasing in recent years, with beekeepers receiving as little as A$3 per kilo a decade ago and up to A$30 per kilo in recent years at the farm gate (Hastie 2018). These high values for honey, combined with the volume of the honey produced from a typical apiary site in a *Corymbia calophylla* forest, means that in a producing year an apiary site in a *Corymbia calophylla* forest can yield a beekeeper upwards of A$50,000 of honey from 100 bee hives in less than two months (Barbour et al. 2019).

With the *Corymbia calophylla* forests of southwest Australia being unique to the geographic region described as the South West Australia Floristic Region (Beard 1980), there are a number of endemic fauna species that rely on the *Corymbia calophylla* trees as a food source. This includes both the nectar from the flowers and seeds from the nuts produced post flowering. Some of these species have been listed as nationally threatened under the Federal Environmental Protection and Biodiversity Conservation (EPBC) Act, such as the Red-tailed Black Cockatoos (*Calyptorhynchus banksii*). These birds are listed as vulnerable under the EPBC Act (Beeton 2009) and rely heavily on the seeds from the marri trees as a food source (Cooper et al. 2003).

Baudin's Cockatoo (*Calyptorhynchus baudinii*) is another threatened species endemic to these forests. As with the Red-tailed Black Cockatoos, the Baudin's Cockatoo relies heavily on *Corymbia calophylla* trees for food, feeding off the seeds, buds, nuts and nectar at various times of the year (Johnstone and Kirkby 2008). There is some evidence that in years of poor *Corymbia calophylla* flowering, followed by low nut and seed production, these cockatoos tend to rely more heavily on cultivated crops as a food source (Cooper et al. 2003). This can inflict severe damage on viticulture crops (Proffitt 2019) as the *Corymbia calophylla* trees typically flower just prior to the viticulture harvest, providing the cockatoos with an excellent source of nectar. The absence of this nectar source causes the cockatoos to seek food from the nearly fully formed grapes.

Carnaby's cockatoo (*Calyptorhynchus latirostris*) is the third of the cockatoo species of the *Corymbia calophylla* forests. It too is a declared threatened species, and relies on the *Corymbia calophylla* trees for food (Cooper et al. 2002). Interestingly, due to the different beak shapes of the three species of cockatoos mentioned here, the *Corymbia calophylla* nuts left on the ground after feeding can be used to identify which species has been feeding in the area based on the beak marks (Fleming 2018).

With southwest Australia being recognised as a global biodiversity hotspot (Myers et al. 2000), these three species of cockatoos are just a few examples of the many endemic, and threatened, species that rely on the *Corymbia calophylla* forests for habitat in the form of – forage, roost and nesting sites. A greater understanding of the phenology of the *Corymbia calophylla* trees, and the ability to project food availability on an annual scale and the impacts of a warming and drying climate due to climate change on a long term scale (Andrys et al. 2017) may represent valuable information for conservation planning and management.

1.3 Research Objectives and Significance

The primary research objective of this thesis is to develop methods to map and predict *Corymbia calophylla* honey harvest weight, utilising readily available temporospatial input data such that the models can be used for:

1. Tools for apiarists in Western Australia to optimally manage their *Corymbia calophylla* honey apiary sites
2. A development template to develop similar tools for honey harvests from other species and regions

The development of this model comprises four distinct sequential steps:

1. Development of a simple algorithm to calculate percentage flower coverage of an apiary site's canopy using readily available Unmanned Aerial Vehicles (UAVs), such as could be deployed by a beekeeper
2. Spectral separation analysis of *Corymbia calophylla* flowers against other environmental features that are typically found in *Corymbia calophylla* forests, and use of this analysis to determine the most suitable satellite sensors for detection of marri flowers
3. Assessment of relationships between weather data, remotely sensed vegetation indices and *Corymbia calophylla* honey harvests
4. Use of machine learning tools to develop a quantitative prediction model based on the key findings from steps 1 – 3.

The direct economic significance of this research is with the potential for increased honey production, particularly in years where the *Corymbia calophylla* harvest weights exhibit significant spatial variability. With the *Corymbia calophylla* honey harvested per hive ranging from 0 kg for some apiary sites in some years to as high as 71 kg per hive (Campbell et al. 2019), and varying by as much as almost 55 kg per hive between sites in a single year, knowing in advance the expected honey production across different apiary sites each season can have a significant impact on both bee health and apiary profitability. For a commercial beekeeper operating 1,000 beehives, an additional 55 kg of high value honey per hive at a 'farm gate' price of A$30 equates to an increased revenue of A$1,650,000 over the typical *Corymbia calophylla* flowering period of six weeks (Brooker and Kleinig 2001).

With the machine learning model development process described in this thesis using feature inputs that are widely available across most of the globe, the process used to develop the *Corymbia calophylla* honey prediction model could feasibly be used to create a predictive model for other honey varieties, both in the same region or elsewhere in the world (subject to reliably accurate historic honey harvest data). This has the potential to broaden the direct economic significance of the research considerably.

As has been discussed earlier, under Section 1.2 (Background), wild and managed honey bee populations have a significant role in the pollination of commercial crops. This represents a larger economic benefit than the direct value of honey and related product sales, as well as impacting on global food security. With 'failed' honey seasons often resulting in reduce bee and hive health, and even death of entire hives (Huntsdale 2019),

being able to predict these seasons ahead of time may allow beekeepers to adjust their apiary management for the season accordingly, thereby reducing the impact of the failed season.

1.4 Research Design and Thesis Outline

This exegesis thesis is structured according to Curtin University's "Guidelines for Thesis by Publication and Hybrid Thesis". Accordingly, as summarised in Table 1-1, Chapters 1 and 2 of this thesis cover the general introduction to the research, the literature review and review and discussion of the findings. Chapters 3 through 6 present the research, both as thesis chapters and peer-reviewed papers published from the research. The findings, limitations and recommendations for further work are presented in Chapter 7.

The general introduction contained in Chapter 1 covers background to the research, including the potential significance of the research, as well as an overview of the research program and thesis structure. The literature review presented in Chapter 2 provides more detail on the economic value of honey bees and introduces key terms that are used commonly in the thesis and published papers. This is followed by a review of the published material of factors affecting nectar and honey production, the use of remote sensing for phenological applications and specific efforts to develop tools to map flowers and predict honey production.

The publications and research presented in chapters 3 through 6 document the process of developing a model to predict *Corymbia calophylla* honey harvests. Chapter 3 presents the development of a simple algorithm to detect *Corymbia calophylla* flowers in images collected from lower cost, off-the-shelf UAV platforms, such as are readily available to beekeepers without extensive training and certification by the Civil Aviation and Safety Authority (CASA) in Australia. This publication also describes optimal UAV survey parameters for *Corymbia calophylla* flower mapping.

Spectral separation assessment of *Corymbia calophylla* flowers from their typical environment is presented in the publication in Chapter 4. This assessment was performed on field data collected with a hand-held spectroradiometer in *Corymbia calophylla* forests, with the spectral data adjusted to reflect the synthetic spectral separation of satellite-borne sensors commonly used for vegetation and ecological studies. Assessment of these sensors included estimates of the canopy cover and flower cover required for flowers to be detected in an average pixel size. This was tested against actual honey harvest data over several years for an apiary site in mature *Corymbia calophylla* forest.

Having established UAV and satellite-based methods for detection of *Corymbia calophylla* flowers, the research was extended in the publication in Chapter 5 to cover a wider range of *Corymbia calophylla* apiary sites (such as partially cleared forest and farmland with remnant *Corymbia calophylla* trees). In addition, following on from related studies discussed in the literature review in Chapter 2.6, efforts were made to include vegetation indices and weather variations into the assessment and an initial *Corymbia calophylla* honey prediction model was developed.

The quantitative prediction model is improved in the publication in Chapter 6, with the development of a larger, multidimensional dataset that contained over 100 input feature variables. Use of machine learning algorithms resulted in a predictive model able to provide a commercially useful prediction several months before the *Corymbia calophylla* honey typically starts to be collected, with only a few input features required.

Chapter 7 is a review and discussion of the key outcomes of each published paper, including the scope and limitations of these outcomes, and explores potential areas for further research (including expansion of the outcomes to other regions and even continents).

Table 1-1 - Summary of thesis structure

Chapter	Title	Summary
1	General Introduction	General background on the importance of the honey industry, the research objectives and significance.
2	Literature Review	Literature review of factors affecting honey production, remote sensing for phenological applications and studies specifically relating to remote sensing and honey production.
3	PUBLISHED PAPER Simple remote sensing detection of *Corymbia calophylla* flowers using common 3 –band imaging sensors (Campbell and Fearns, 2019) DOI: 10.1016/j.rsase.2018.04.009	Development of a simple algorithm to calculate percentage flower coverage of an apiary site's canopy using readily available Unmanned Aerial Vehicles (UAVs), such as could be readily deployed by a beekeeper
4	PUBLISHED PAPER Honey crop estimation from space: Detection of large flowering events in Western Australian forests (Campbell and Fearns, 2019) DOI: 10.5194/isprs-archives-XLII-1-79-2018	Spectral separation analysis of marri flowers against other environmental features that are typically found in *Corymbia calophylla* forests and use of this analysis to determine the most suitable satellite sensors for detection of marri flowers.
5	Prediction and detection of honey harvests from remote sensing and weather data	Assessment of relationships between weather data, remotely sensed vegetation indices and *Corymbia calophylla* honey harvests.
6	PUBLISHED PAPER Machine learning regression model for predicting honey harvests (Campbell et.al. 2020) DOI: 10.3390/agriculture10040118	Use of machine learning tools to develop a quantitative prediction model based on the key findings from published papers listed above.
7	Review and Discussion	Summary of findings from this study and perspectives on the further work.
8	Bibliography	A complete list of all publications referenced in both the thesis exegesis and the candidate's published papers.

2 Literature Review

2.1 Overview

The literature review in this chapter is supplemented by the respective introductory sections of the publications and research chapters of this thesis (Chapters 3 through 6). Therefore, this literature review focuses on the details that are not explicit in the contents of the publications. In particular, this chapter reviews the main factors that influence nectar and honey production, a summary of the underlying physics of remote sensing and common ecological applications of remote sensing and a summary of current work done on the use of remote sensing data to map and/or predict honey and nectar production.

Section 2.2 gives a brief background on the coevolution of plants and pollinators, including some unique examples from Australia.

Section 2.3 reviews the main ecological and climatic factors affecting phenology, for both agricultural crops in managed landscapes and perennial plant species in natural landscapes.

Section 2.4 provides an outline of the underlying physics of remote sensing, a summary of the main types of sensors and platforms and the main considerations regarding sensor selection.

Section 2.5 gives an overview of the history and current applications of remote sensing for phenological applications.

Section 2.6 provides a brief summary of existing studies on the use of remote sensing for nectar and honey production.

2.2 Coevolution of Plants and Pollinators

Land plants evolved approximately 450 million years ago (Morris et al. 2018), appearing during the middle of the Cambrian–Early Ordovician interval. Flowering plants, or angiosperms, are a relatively recent evolutionary group of plants, with the oldest fossils to date from the Middle–Late Jurassic period, approximately 160 million years ago (Liu and Wang 2016). Current estimates of the estimated number of species in the Kingdom Plantae are constantly being revised due to the discovery of new species and adoption of new phylogenetic tools, with an average of 2000 species resolved as new to science per year for the past decade (Christenhusz and Byng 2016). As of 2016, there were an estimated 403,911 described species of land plants, of which 369,434 were angiosperms (Lughadha et

al. 2016). This means that angiosperms dominate the plant kingdom, representing more than 90% of the diversity of land plant species.

The concept of coevolution between flowering plants and their animal pollinators dates back to the earliest theories of evolution, and indeed is discussed in Charles Darwin's seminal work 'On the Origin of Species' (Darwin 1859). One of Darwin's earliest specific coevolution theories was regarding the Madagascan long-spurred orchid (*Angraecum sesquipedale),* and the hawk moth (*Xanthopan morganii praedicta*), as this moth possessed the exceptionally long proboscis required to pollinate the exceptionally deep flowers of the orchid. This theory was proved by video observations late in the 20th century (Arditti et al. 2012) over a century after Darwin's prediction of the type of pollinator.

Studies of fossilised angiosperms has shown that as far back as the Early Cretaceous period, the majority of basal angiosperms were zoophilous (Hu et al. 2008). With a coevolutionary time frame well in excess of 100 million years, the pollination adaptations by both plants and animals is varied and widespread. Some plants have evolved flowers to appeal to as wide of a range of pollinators as possible ('generalists'), while others focus on a particular niche for specific pollinator interactions ('specialists'), as described by Fenster et al. (2004).

One widespread example of the coevolution towards specialised plant-pollinator interactions is the spectral signature of flowers. Several studies have shown that insects from the *Hymenoptera* order (which includes bees, wasps and ants) have poor to no colour discrimination for light with wavelengths greater than approximately 550 nm, or green light (Chittka et al. 1994; Chittka and Waser 1997). Comparisons between the spectral signatures of flowers and the vision characteristics have shown that often insect pollinated flowers have spectral signatures that closely match the spectral discrimination capabilities of hymenopteran pollinators, by Bischoff et al. (2013) for example. Indeed, it has been shown in Australia that on average, insect pollinated plants have flowers of lower wavelengths while bird pollinated plants have flowers with colours of higher wavelengths (Shrestha et al. 2013).

An extreme example of specialised pollination is the *Drakaea* genus of orchids, from southwest Australia (Hopper and Brown 2007). This genus of 10 species have evolved sexual mimicry of the females of several wasp species for pollinator attraction, including both the physical appearance and release of pheromones. When the male wasps land on the flower, he grabs the female 'wasp' and attempts to fly away to mate. A hinge in the flower swings the male upside down with his own momentum and onto the orchid's

reproductive structure. Such sexual deceptive tools are amongst the most sophisticated systems developed in the angiosperms for pollination, with the bulk of plant species engaging in this strategy found in the South West Australian Floristic Region (Hopper and Brown 2007).

2.3 Factors Affecting Flowering Phenology Times

With Angiosperms found on every continent (Cronquist et al. 2019), including two flowering plants on Antarctica (Bravo et al. 2001), these plants have evolved to survive in a wide range of environments. As a result, the phenology of flowering has developed to respond to a similarly wide range of environmental cues, particularly temperature and photoperiod for angiosperms in temperate environments (Tooke and Battey 2010).

Efforts to record flowering times over extended periods and determine the reasons for variations in flowering times have resulted in some datasets containing data for durations of over several centuries, for example the records started by the naturalist Robert Marsham in Norfolk in 1736, that were continued though to 1958 (Sparks and Lines 2008). Another approach is to compile historic records, such as done by Aono and Kazui (2008). Due to the historic cultural significance of the annual Japanese cherry blossom festival, these researchers were able to build a flower phenology dataset from 801 – 2005 by compiling data from herbarium specimens, photographs, diaries and even advertisements for the annual festivals.

An important consideration in recording flowering phenology is that the start of flower bud formation is also noted in addition to the emergence of the flower itself (Grainger 1939). Indeed, for some species floral initiation in the form of buds visibly starting to form commences up to a year before anthesis (for example *Nerine sarniensis* (Rees 1966) and *Eucalpytus marginata* in south-west Australia (Davison and Tay 1989)). Due to the variability of phenological timescales between species, and the variability of the definitions of flowering commencement (such as ‘visible anthers’ (Fitter et al. 1995) or ‘an open flower is a flower in which stamens or stigmas could be seen without the observer pushing petals aside’ (Last 2001)), there have been efforts to develop a standard universal system for phenological stages, including flowering phenology. One example of this is the BBCH scale for agricultural crops (Lancashire et al. 1991), named after the institutions that co-developed it (Biologische Bundesanstalt, Bundessortenamt, Chemische Industrie).

Studies have used these measures of flowering phenology to unravel the effect of the numerous factors that affect flowering times, including photoperiod, nutrients, ambient

temperature, drought, salinity, exogenously applied hormones and chemicals, and pathogenic microbes (Cho et al. 2017; Mouradov et al. 2002), to improve the understanding of how these factors interact with a view to improve agricultural crop production.

The complexities of the relationships between flowering times and environmental factors are summarised well by a study of flowering patterns of four species of co-occurring *Eucalyptus* trees in south-east Australia by Hudson et al. (2009). This study found that there is a non-linear relationship between flowering intensity and temperature, which varied between the four species. Two species flowered more intensely with increased minimum and maximum temperatures, one species flowered less intensely with increased minimum temperature and one species flowered less intensely with increased maximum temperature. Therefore, even for plants of the same genus that have evolved in the same climatic region, there is far from a simple relationship between flowering times and key climatic variables.

Due to the sensitivity of flowering phenology to these environmental conditions, studies have used data on the variation of flower phenology timing over several years to assess the impacts of climate change on phenology and, by extension, ecosystem services. This has included two species of plants in Florida (USA) that had significantly later flowering times and one species flowered significantly earlier in the season, due to colder winters and warmer summers from the early to late 20th century (Von Holle et al. 2010). Menzel et al. (2006) found that across Europe from 1971 to 2000, 31% of study locations recorded a significant advance in flowering times, whereas 3% of locations recorded a delay in flowering times. It has been suggested that these changes to flowering times may have severe long-term consequences to the ecosystems, with a phenological mismatch between the flowering times and the pollinator lifecycles leading to reduced pollination rates (Petanidou et al. 2014).

2.4 Remote Sensing Overview

Remote sensing is defined by Khoram et al. (2012) as “the acquisition and measurement of information about certain properties of phenomena, objects, or materials by a recording device not in physical contact with the features under surveillance”. While this general definition includes such areas as medical imaging with X-rays, the use of the term ‘remote sensing’ is generally understood to relate to targets on, above or below the Earth’s surface using active and/or passive electromagnetic sensors (Kyalo Kiema 2013).

The most common form of remote sensing, that is used many people every day, is the photograph. The first recorded example of remote sensing in the modern context is by Gaspar Felix Tournachon in 1858, who used a hot air balloon to take a photograph of a French village from an altitude of 80 m (Estes et al. 1977). Over the next few decades, photographic equipment was mounted on kites, pigeons (see Figure 2-1) and even rockets by 1912.

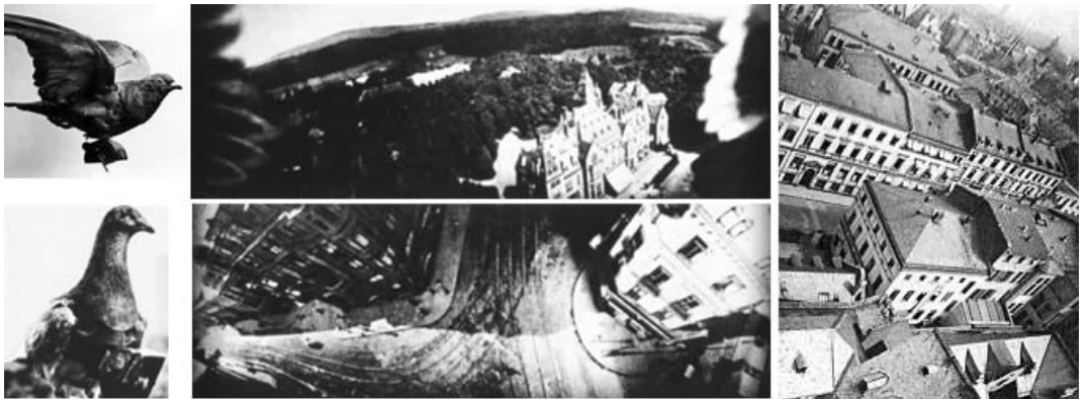


Figure 2-1: Example of remote sensing images obtained via a carrier pigeon (https://www.professionalaerialphotographers.com/)

While human eyes, and most cameras, can discern the visible part of the electromagnetic (EM) spectrum, the full range of this spectrum is much wider than what we can naturally perceive, as Figure 2-2 shows. With different chemicals and molecules having different responses across parts of the spectrum, recording measurements in the correct wavelengths gives the ability to discern between materials (Zielinski et al. 2005), even if they are indistinguishable to the human eye (see Figure 2-3).

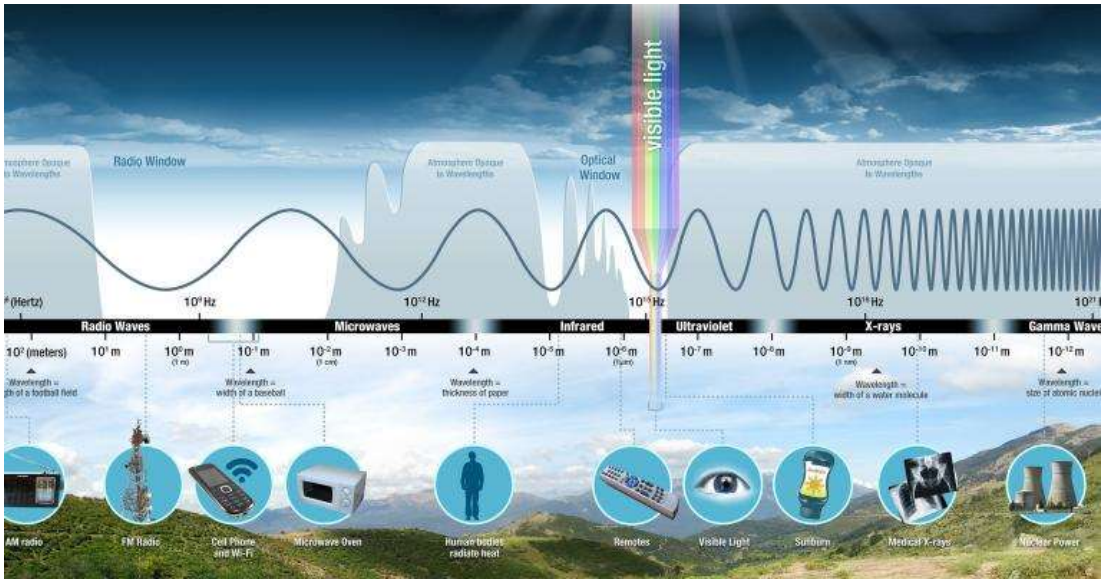


Figure 2-2: The electromagnetic spectrum (image courtesy of NASA)

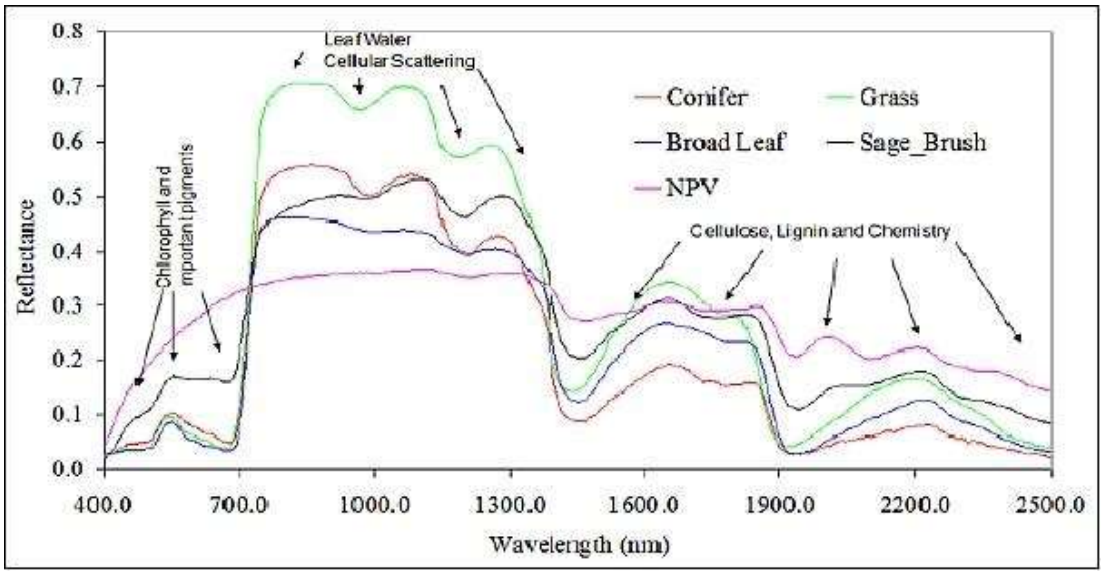


Figure 2-3: Typical spectral reflectance responses for a range of vegetation types. Note that 'visible light' (i.e. light that the average human eye can discern) is from wavelengths 380 nm to 740 nm (image courtesy of NASA)

Measurement of spectra requires specialised equipment, that typically use the diffraction properties of different wavelengths of the EM spectrum to focus different wavelengths on different parts of the sensor which record the intensity of the signal received for the different wavelengths. This can be from the entire Field of View (FOV) of the instrument (spectroscopy), or for each pixel of an image (image spectroscopy).

In order to record a sufficiently strong signal for accurate measurements, often instrument design is a compromise between the different types of resolution as defined by Liang (2012) below:

- Spatial resolution is a measure of the smallest object that can be resolved by the sensor
- Spectral resolution describes the number and width of spectral bands in a sensor system
- Temporal resolution is a measure of the repeat cycle or frequency with which a sensor revisits the same part of the Earth's surface
- Radiometric resolution refers to the dynamic range of the sensors, with a lower resolution having a lower dynamic range and therefore a lesser ability to measure more subtle changes in reflectance

Each of these resolutions has an impact on the overall capability, for example a sensor with a very high spatial resolution will receive less EM radiation per pixel, so may have to have lower spectral resolution to ensure sufficient signal is received for each pixel. These are also affected by the distance between the sensor and the remotely sensed object; as EM

radiation decreases by the square of the distance, a sensor 20 m from the target will receive four times as much EM radiation, or signal, as a sensor 40 m away. When comparing handheld, airborne and satellite borne sensors, this has a sizeable impact on the sensor capabilities due to the underlying physics. A comparison of the different resolutions for commonly used satellite borne sensors is given in Table 2-1. From this table, there is a clear inverse relationship between the spatial and temporal resolution. Note that there is a range given for the spatial and spectral resolutions for each sensor, with some spectral bands in each sensor having a lower spatial resolution and a higher spectral resolution and vice versa. This widens the applications for which each sensor can be used for.

Table 2-1 - Comparison of resolution specifications for some commonly used satellite borne imaging spectrometers

SENSOR NAME	SPATIAL RESOLUTION	SPECTRAL RESOLUTION	TEMPORAL RESOLUTION
Himawari8	500 – 2,000 m	30 – 960 nm	10 minutes
MODIS	250 – 1,000 m	10 – 500 nm	1 day
Sentinel 2A	10 – 60 m	15 – 175 nm	10 days

As satellite borne sensors measure the EM radiation reflected from the surface of the Earth and transmitted to the Top of Atmosphere (TOA), corrections must be applied. One of these corrections is the viewing angle (Kimes et al. 1984), which is particularly relevant for geo-stationary satellites that are able to view nearly a whole hemisphere of the earth for one image scan, such as Himawari8 (see Table 2-1) with a Field of View (FOV) of up to 78.3° either side of the equator. An initial correction can be for the attenuation of the signal by the square of the distance due to spherical spreading (Herman et al. 1995). The more complex corrections account for the effect of the atmospheric column, including the effect of a thicker atmospheric column at greater angles from nadir in the FOV. Repeatable and accurate corrections are particularly important when multi-temporal images are compared in time series for change detection (Hadjimitsis et al. 2010).

While the comparison of sensors in Table 2-1 is for satellite borne sensors, spectrometers can also be mounted on aircraft, Unmanned Aerial Vehicles (UAVs, or ‘drones’) or hand held platforms. While these platforms are rarely suitable for regular re-surveys of locations (hence have no effective temporal resolution), the sensors can have significantly higher spatial and spectral resolution than satellite borne sensors. In addition, the effects of viewing angles and the atmospheric column are reduced, or can even be effectively zero,

further improving the accuracy of the measurements (Iqbal et al. 2018). However, calibration of the EM radiation signal strength is still required for detailed spectral analysis (Honkavaara and Khoramshahi 2018).

Once the remote sensing spectral data has been appropriately corrected, the data can be analysed and interpreted. This can be a visual interpretation of features, where the accuracy of the result depends almost entirely on the skills of the interpreter (Richards and Jia 1999). More often however, computer algorithms are used for this process, with the interpreter making the final decision on the interpretation outcome based on a statistical analysis of the algorithm outputs and their professional expertise. These approaches are broadly separated into two groups; 'supervised classification' and 'unsupervised classification' (Congalton 2010).

Unsupervised classification, or 'clustering methods', attempt to group all data together into groups that are spectrally similar, i.e. there is a statistically significant difference between the average spectra of each group (Duda and Canty 2002). Typically the interpreter's input to this process is limited to specifying the number of spectral groups to classify into, although even this part of the process can be automated. Once the clustering process is completed, the interpreter typically reviews the different clusters produced and uses their professional expertise to determine the most suitable outcome (Richards 2013).

While unsupervised classification does produce a statistically rigorous separation of pixels into groups, unless there is a strong spectral separation between the target groups the algorithm may not be able to separate target groups effectively (Enderle and Weih 2005). As a result, the unsupervised classification approach is often used to investigate the separation characteristics of groups prior to undertaking the more time consuming, but often more accurate, supervised classification methods (Long and Srihann 2004).

Under the supervised classification approach, the interpreter selects data image areas of pixels from a known classification group, and this 'training' dataset is used by the algorithm to classify the remainder of the dataset into the prescribed classification groups (Abburu and Babu Golla 2015). These training datasets can come from a variety of sources, such as:

- Established spectral libraries (Herold et al. 2004)
- Field mapping of the image area (Key and Benson 2006)
- Visual interpretation of the image, or higher resolution imagery of a subset of the image area (Schowengerdt 2006)

Generally, an initial assessment is performed to ascertain how separable the target groups are, such as the Analysis of Variance (ANOVA) test (Vaiphasa et al. 2005), which is a measure of separation of the means of each target group combined with each group's standard deviation. If the ANOVA results indicate that the spectra of two groups are not statistically separate, that metric is often not used for more detailed analysis. This method can be used to determine which parts of the spectra are best suited for spectral separation of target groups (De Castro et al. 2014). More detailed calculations on spectral separability can also be applied to the optimal spectral regions, such as calculation of the Jeffries-Matusita (JM) Distance (Yongji et al. 2018) or Principal Component Analysis (PCA) (Munyati 2004).

With the optimum spectra identified, the data or image can then be classified into the different classes. This can be a simple Parallelepiped approach (Lu and Tang 2011), where upper and lower limits for each class in each spectra are specified. This method works well when there is little to no overlap between the spectra of the classes, but is not suitable for cases where there is spectral overlap (Perumal and Bhaskaran 2010), or where the image pixels contain a spectrum from a combination of classes. This "mixed spectrum" is common when the spatial resolution of the sensor is larger than target features (Quintano et al. 2012).

Where there is some overlap between classes, a maximum likelihood classification algorithm can be applied to interpret the data (Bolstad and Lillesand 1991). These algorithms have been long established as effective tools across different land cover types, conditions and sensor systems (Swain and Davis 1981). The key advantage of the maximum likelihood approach over parallelepiped methods is the inclusion of a likelihood/error metric in the output allowing users to determine how reliable the classification process has been and screen results accordingly (Strahler 1980).

To overcome the issue of measured spectra from individual pixels resulting from the combined radiance of multiple classes, spectral unmixing algorithms are commonly applied. These algorithms typically use spectral libraries to determine the proportion each class contributes to the nett radiance of a pixel (Du 2018). The output of this process is an expression of the linear combination of the class for each pixel, with an associated uncertainty metric (Plaza et al. 2004).

With the increase in readily available computing power over the past few decades, the ability of algorithms to handle large, highly multidimensional datasets has led to more

advanced classification algorithms being applied under the umbrella of 'Machine Learning' (Lary et al. 2016). While the mathematical models for these algorithms have often been in existence for several decades, there has been insufficient computing power generally available to make the application of these algorithms to remote sensing datasets widespread. These advances have improved the ability of algorithms to map target classes with very complex, and interrelated, characteristics (Maxwell et al. 2018). The commonly used algorithms include:

- Recursive Decision Trees (DT), which are essentially a set of if-then rules of if data is above or below a certain value (Pal and Mather 2003)
- Random Forest (RF), which uses multiple Decision Trees with weighted averages for results to find a global optimum result (Breiman 1984)
- Artificial Neural Networks (ANN), essentially mathematical analogues of an animal brain's axons and their many interconnections through synapses (Atkinson and Tatnall 1997)

The ability of these algorithms to handle large, complex, interrelated datasets (Elith et al. 2008) has allowed extraction of detailed biophysical and harvest quality properties in some cases, such as phosphorus levels in grassland (Gao et al. 2019), leaf chlorophyll content (Upreti et al. 2019) and smoke contamination in horticultural crops (Fuentes et al. 2019).

The volume of data produced to from high-resolution UAV imagery in particular lends itself to advanced algorithms such as ANNs, with this approach used for such applications as crop biomass calculations (Han et al. 2019), individual tree species mapping (Miyoshi et al. 2020) and rapid detection of the yellow-rust pathogen (Zhang et al. 2019).

2.5 Ecological and Agricultural Applications of Remote Sensing

The use of remote sensing for ecological and agricultural purposes was a key objective of NASA's Project Earth Resources Observation Satellites (EROS), with the Secretary of the Interior Stewart Udall announcing in 1966 that "the program will provide us with the opportunity to collect valuable resource data and use it to improve the quality of our environment." (Udall 1966). The Landsat 1 satellite launched in 1972 included a Multi-Spectral Scanner (MSS) with four bands – green, red and two infrared bands (MSS Standard Interface Document 1978).

The data from Landsat 1 was used to commence a program of global landcover classification within a few years of the satellite being launched (Williams and Coiner 1975),

with initial classification schemes based on up to nine major classification groups, such as Urban, Agricultural, Rangeland, Water, etc. with 37 subgroup classifications (Anderson 1976). Early efforts to obtain more detail on vegetation landcover classes included forest wood type, such as hardwood versus mixed wood or softwood (Beaubien 1979) and forest inventory data (Strahler 1981). These often had mixed results, owing to the ecological complexity of the classes and the limited spatial and spectral resolution of the data.

Early estimates for classification of landcover into agricultural crop types were generally more successful owing to the relative homogeneity of monoculture crop fields compared with natural forest ecosystems (for example Wiegand and Gautreaux (1977) and Malila et al. (1980)).

As the time series of remote sensing data built up, the ability to monitor change in landcover over time improved. In 2008, the GLS2005 unified global landcover dataset was created (Gutman et al. 2008) with a goal of global monitoring of landcover changes every 5 years. Historic data were used to create earlier versions for 1975, 1990 and 2000 and the program continued through to create GLS2010 (Gutman et al. 2013). Time-series remote sensing data has also been used to map landcover change on regional scales, such as the Amazon Rainforest (Souza et al. 2013) and the impact of palm oil plantations on deforestation in Indonesia (Margono et al. 2012).

Vegetation change detection often utilises Vegetation Indices (VIs), such as the Normalised Difference Vegetation Index (NDVI), to calculate relative changes in vegetation over time (Gandhi et al. 2015). NDVI was first developed by Rouse et al. (1974) to measure above ground biomass of the Great Plains region of the USA and was defined at this time as the Band Ratio Parameter (BRP) between Landsat-1 bands 5 (green) and 7 (NIR). This index is based on highlighting the spectral response of plant matter, with the chlorophyll in plant leaves absorbing a large proportion of visible light (band 5) and having a high reflectance of infrared light (band 7).

While NDVI is widely used, other indices have also been developed in efforts to counter the limitations of NDVI. These include the Soil Adjusted Vegetation Index (SAVI), which uses a constant factor to adjust for the effect of bare soil on NDVI (Huete 1988), the Enhanced Vegetation Index (EVI), which accounts for some atmospheric impacts as well as constant factors for soil effects (Huete et al. 1999).

To measure changes in NDVI over time, for assessment of seasonal and/or annual variations in vegetation properties, the series of individual NDVI measurements must be transformed

into a time-series dataset. This is often hindered by noise sources inherent to remote sensing datasets, such as varying atmospheric conditions (Carreiras et al. 2003), with aerosols having a particularly large impact on infrared reflectance (Goward et al. 1991).

Many algorithms have been tested to create accurate, continuous time-series remote sensing datasets. Algorithms have included polynomial smoothing (Van Dijk et al. 1987), Fourier wave adjustment (Sellers et al. 1994), asymmetric Gaussian function fitting (Jönsson and Eklundh 2002) and double logistic functions, which can be uniquely defined for each pixel by parameters that describe annual NDVI time series (Beck et al. 2006). These approaches have the general limitation of relying on data points earlier and later in the time series for their calculation; utilisation of these algorithms to calculate more accurate VIs for currently acquired data borders on extrapolation and hence these approaches to build time series datasets are generally more appropriate for recent to historical data analysis.

Several programs are available to create continuous time series data from multiple input files, such as TIMESAT (Jönsson and Eklundh 2004), and automatically detect changes in VIs related to key seasonal parameters, such as DBEST (Jamali et al. 2015). Both of these examples allow users to select the optimum time series creation algorithm for the project objective and input data considerations.

Ultimately, there is no ideal 'fit for purpose' solution to all applications and users must assess their objectives and the nature of the noise present selecting an approach to noise reduction and time series creation (Hird and McDermid 2009)

The insights obtained from differences in seasonal variations in VIs derived from remote sensing time-series datasets can improve classification of vegetation cover regimes (Yan et al. 2015) and ecosystem types (Clerici et al. 2012), with concomitant improvement in habitat and biodiversity mapping.

In deciduous forest environments, VIs have been used to measure timing of Start of Season (SOS), Growing Season Length (GSL) and End of Season (EOS) metrics, with these related to annual variations in weather (Testa et al. 2018). For temperate and tropical evergreen forests, these parameters can also be measured, although seasonal variations of VIs are smaller than deciduous forests to the point at which in some cases seasonal fluctuations are not reliably measurable (Moulin et al. 1997).

Long time series investigations can be used to assess the effect of large-scale climatic events on vegetation behaviour, such as El Nino and La Nina periods (Asner et al. 2000) and climate change (Du et al. 2014). This can be done in a quantitative manner through calculation of changes in annual SOS, GSL and EOS over time via VIs (Hogda et al. 2001). Due to the spatial coverage of remote sensing datasets, long term trends across different climate regions and latitudinal changes can be assessed concurrently (Butt et al. 2011).

While NDVI has proven to be a reliable measure of relative visible green biomass, it is a dimensionless ratio. Many efforts have been made to relate NDVI, and other dimensionless vegetation indices, to on-ground biophysical parameters. These are often measurements of leaf cover such as Leaf Area Index (LAI), which has been shown to have strong correlation with NDVI in grasslands (Fan et al. 2009) but often a more variable relationship with forested terrains (Wang et al. 2005) with strong intra-annual and inter-annual effects.

Due to the relative homogeneity of paddocks used for crop production, relationships between NDVI and crop biophysical parameter are generally more reliable. This includes LAI (Duchemin et al. 2006) and above ground biomass (Kross et al. 2015), which can be extrapolated to predict harvest yields based on peak NDVI over the growing season (Thorp et al. 2012).

Crop yield predictions have been made over large regions using a combination of weather and vegetation indices, with Seiler et al. (2000) for example forecasting crop yields in Argentina two months before harvest with an RMSE of between 5 and 7%. Wannebo and Rosenzweig (2003) analysed crop yields over the US cornbelt compared to the El Nino-Southern Oscillation (ENSO) fluctiations for 15 years and used this information to show that the ENSO has a relatively small effect on NDVI and corn yields.

With the increasing spatial and temporal resolution of remotely sensed datasets, farm and paddock-scale yield predictions are increasingly common (Al-Gaadi et al. 2016), with Variable Yield Technology (VYT) and other precision farming techniques used to target agricultural management practices for under-performing areas .

Many crop yield prediction studies have used a combination of NDVI with Vegetation Condition Index (VCI) and Temperature Condition Index (TCI), which are measures of NDVI and temperature scaled by multi-year absolute maximum and minimums (Kogan 1995). These indices assist with scaling remote sensing data to local conditions (reducing the relative nature of NDVI between locations) and have proven to provide useful information

on droughts, both wide spread, long term and localised, short term droughts (Jiao et al. 2016) and vegetation stress (Singh et al. 2003).

In order to incorporate local weather conditions, studies have also been conducted using remotely sensed VIs with remotely sensed local weather data, such as Prasad et al. (2006). This study combined NDVI data with temperature, rainfall and soil moisture data to estimate corn and soybean crop yields for Iowa (USA). In this study, a crop prediction model was created from a non-linear Quasi-Newton method which yielded R^2 values of 0.78 for corn crops and 0.86 for soybean. It was noted that other factors not incorporated into the model, such as pests, diseases and human activities are serious limitations to the model.

Efforts have been put towards development of generic freeware tools for crop prediction, such as the Crop Statistics Tool (CST), developed by Kerdiles et al. (2017). This program uses both multiple regression and historic data similarities to forecast a crop yield.

Due to the effects of non-linear relationships, multicolliniarity and other complex relationships between input factors and crop yields, artificial intelligence (AI) or Deep Learning (DL) algorithms have begun to be applied in recent years (Wang et al. 2018). While these algorithms typically outperform other techniques (You et al. 2017), significantly larger training datasets are typically required for this improved performance. As mentioned earlier in Section 2.4, the volume of data created from high-resolution UAV surveys typically lends itself to these applications.

Possibly one of the most comprehensive efforts to combine various remote sensing datasets into a crop yield prediction tool is the Graincast app, developed by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) in Australia. This app provides a fortnightly forecast of seasonal grain yields on national, state, local and farm paddock scales over the growing season (Hochman and Horan 2019). As described by Lawes et al. (2019), the following stages are used to generate the yield forecasts:

1. Remote sensing classification of crop type using Random Forest machine learning (Breiman 1984), with thousands of roadside images used in the training dataset.
2. Calculation of crop yield from the C-Crop model, based on MODIS-derived Fraction of Photosynthetically Active Radiation (FPAR) data and national weather grids from BOM (Donohue et al. 2018).
3. Delivery of the model to users via integration with APSIM, which provides farm management recommendations for the growing season (Holzworth et al. 2014)

2.6 Previous studies on flower detection and predicting honey flow

While mapping of vegetation species with multispectral data is often difficult due to the high degree of spectral correlation across vegetation foliage from their common chemical composition (Portigal et al. 1997b), zoophilous flowers have generally evolved to contrast against leaves for detection of flowers by pollinators (Dyer et al. 2012).

Studies have shown that in the spectral space there is often a greater spectral separation between leaves and flowers (Gross and Heumann 2014) and even between flowers of different species than between leaves of different species (Shrestha et al. 2013). High resolution airborne hyperspectral data were shown to have a high degree of accuracy in detecting flowering in native vegetation in Africa (Landmann et al. 2015).

In assessing the ability of multispectral data to map canola yields based on detection of their yellow flowers, Sulik and Long (2015) clearly demonstrated the high level of spectral separation of canola flowers versus leaves in the visual bands. The difference was such that a multispectral capable UAV with a standard RGB camera was able to map changes in flower abundance. This was developed further into a normalized difference yellowness index (NDYI) based on MODIS sensor characteristics, which correlated better with canola yields than NDVI across North Dakota for the study period (Sulik and Long 2016).

Several studies have investigated using NDVI-derived seasonal metrics to predict the timing and/or volume of honey harvests. For example, a European study of honey yield against temporal variations in the NDVI from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor (Blomstedt 2014) showed a consistent time delay between the Start of Season (SOS) 'green-up' indicator and peak honey flow time. The NDVI SOS measurement was shown to be a reasonably consistent predictor of peak honey flow timing ($R^2 = 0.76$).

In Australia, eucalypts across south-eastern Australia yielded a broad, qualitative correlation between changes in vegetation indices derived from MODIS data and honey yield data, with recommendations made for more detailed studies on local variations in floristic communities and the impact of climatic variables (Webber 2011). Similar work using the 'BeeBox' web-based portal (Winter et al. 2013) showed a stronger correlation between annual peak Enhanced Vegetation Index (EVI) from MODIS and *Eucalyptus tricarpa* (ironbark) winter honey flow over a 15-year period (Arundel et al. 2016). Again, this was a qualitative correlation and no quantitative model formed. These studies have both recommended incorporation of local weather data to improve honey yield correlations.

Vlad et al. (2008) developed a quantitative tool to predict honey yield from sunflower crops in Romania, based purely on the effect of temperature on the sunflower growing season. While good results were obtained, they were not transferrable between regions, highlighting the need for individual, localised models in this case.

Hastono et al. (2017) developed a method to predict honey harvest by *Apis cerrana* (Asian honeybee) from *Calliandra* flowers in Central Java, Indonesia, using rainfall data and the Tsukamoto Fuzzy Inference System. This simple input data resulted in an RMSE of less than 1% of the annual harvest yield.

Incorporation of weather data into predictive models is not always a straightforward process, with for example Myrtaceous trees in Australia found to have complex relationships between weather and phenology. A study over 30 years by Hudson et al. (2009) found that while temperature was consistently a key factor in flowering intensity, the effect was far from consistent, with trees in the same forest having opposite phenological responses to maximum and minimum temperatures.

Rocha and Dias (2018) found similar complexities when using Radial Basis Functions to predict honey yields in Andalusia, Spain, concluding that “meteorological variables are good predictors of honey production but they depend on the geographic region and the time frame considered.”

Hawkins et al. (2018) developed a qualitative, relativistic model for landscape/regional scale nectar availability in subtropical eastern Australia for assessing forage availability for nectarivores over an area of 314,400 Ha. This study identified some key factors that were related to nectar availability, notably the Gross Primary Productivity for the previous 12 months, average annual solar radiation and rainfall for the previous 6 months.

These studies have shown that while there is often a relationship between NDVI and/or weather with honey harvest volumes, the relationship is often localised, species specific and complex. With species-specific honey harvest data over a sufficient time frame to capture annual variations, a modelling approach capable of handling complex interrelated input factors may be capable of producing a predictive honey harvest model.

3 Published Paper: Simple remote sensing detection of *Corymbia calophylla* Flowers using common 3 –band imaging sensors

Campbell T, Fearns P (2018) Simple remote sensing detection of *Corymbia calophylla* Flowers using common 3 –band imaging sensors. Remote Sensing Applications: Society and Environment, 11, pp. 51- 63. https://doi.org/10.1016/j.rsase.2018.04.009

Remote Sensing Applications: Society and Environment 11 (2018) 51–63

Contents lists available at ScienceDirect

Remote Sensing Applications: Society and Environment

journal homepage: www.elsevier.com/locate/rsase

Simple remote sensing detection of *Corymbia calophylla* flowers using common 3 –band imaging sensors

Tristan Campbell*, Peter Fearns

Remote Sensing and Satellite Research Group, Curtin University of Technology, Australia

ARTICLE INFO

Keywords:
Parallelpiped
Corymbia calophylla
Multispectral
Flowers

ABSTRACT

With Apis mellifera (the European Honey Bee) having an average forage radius of less than one kilometre from their hive, selecting the best location for beehives is critical for commercial beekeepers to optimise their honey production. In this study, we have used standard three-band digital cameras to develop and assess a simple parallelpiped algorithm to detect the flowers of the Corymbia calophylla tree in Western Australia, the largest source of honey in the state.

The algorithm has been tested in a number of situations and, within the bounds of the study, works to a better than 90% classification accuracy for Digital Single Lens Reflex camera images of trees within 15 m distance, and often better than 95%.

To determine how this approach could be used by Unmanned Aerial Vehicle platforms to detect flowering Corymbia calophylla flowering, image resolution was progressively degraded until flowers could not be detected. It was found that the cluster size of flowers (i.e. whether flowers occur individually or in groups) played an important role in determining the accuracy of flower detection, but overall a minimum resolution of 10 pixels per flower is required for reliable detection of flower pixels. While there is some improvement above this resolution, the effect is minimal. The key to accurate measurement of percentage flower cover is the accuracy of the classification of the background. If the background classification error is known for an image or scene, the percentage flower cover can be calculated with as little as 2% flower cover.

Based on this study, UAV platforms with standard RGB cameras appear to be suitable for detection of Corymbia calophylla flowers if surveys are designed within the bounds described. Thus, UAV surveys may prove to be useful for beekeepers to optimise the location of their beehives amongst a choice of different locations.

1. Introduction

Classification of vegetation using remote sensing methods relies on the differences between the spectral responses of vegetation types being detectable within the constraints of the detection system. The ability to classify vegetation based on spectral differences is also often hampered by the spectral reflectance “signatures” of different species of vegetation overlapping in spectral classification space. Even with the application of detailed hyperspectral measurements with possibly hundreds of spectral bands, the spectral variability within a class, and the similarity between classes, has limited the use of generic ‘spectral libraries’ in being able to aid in species mapping across different regions (Price, 1994), or even within regions in some cases (Zomer et al., 2009).

The high degree of spectral correlation across vegetation species is due to their common chemical composition (Portigal et al., 1997). The overlapping of spectra between species is due to the similar chemical composition of leaf matter across species, and the variability of the composition within a species due to factors external to the plant, such as:

- Environmental factors, e.g. micro-climates, soil characteristics, precipitation, topography and others (Portigal et al., 1997)
- Stress factors, e.g. air pollution, heavy metals and drought (Westman and Price, 1987)
- Successional and phenological stages and seasonal impacts (Carvalho et al., 2013)

The spectral variability between plants due to the above factors is used in precision agriculture for inferring differences in crop health, Leaf Area Index (LAI), biomass and other biophysical variables (Thenkabail et al., 2000). In fact there can also be significant variability in spectral reflectance between individual leaves on a single tree (Cochrane, 2000).

However, flowers have evolved to contrast against leaves for

* Corresponding author.
E-mail address: tristan.campbell@postgrad.curtin.edu.au (T. Campbell).

https://doi.org/10.1016/j.rsase.2018.04.009
Received 26 September 2017; Received in revised form 14 April 2018; Accepted 17 April 2018
Available online 26 April 2018
2352-9385/ © 2018 Elsevier B.V. All rights reserved.

detection of flowers by pollinators. There is strong evidence for the co-evolution of flowers with the visual systems of Hymenoptera genus insects, where vegetation relies on these insects for pollination (Chittka et al., 1994) and it has been shown that some species that rely on birds for pollination have evolved different chromatic clues to be more visible to birds (Shrestha et al., 2013). As flowers have evolved to stand out visually against a background of leaves for pollinators, it is expected that they should have well defined spectral separation from leaves, and potentially between species depending on their target pollinators.

A study on the correlation of Hymenoptera vision and flower colours (Dyer et al., 2012) used an Ocean Optics S2000 spectrophotometer to measure the spectral reflectance of flowers from 300 nm to 700 nm with a resolution of between 0.3 and 10 nm (due to different resolutions of the instrument in different regions of the spectrum) in order to infer the potential parallel evolution of angiosperm flower colours and Hymenoptera visual cues. Their study found the largest portion of colour signals from flowers resides in the blue to green part of the spectrum, with pure UV or UV to blue ranges poorly represented.

Gross and Heumann (2014) collected 388 vegetation reflectance spectra from prairie fens in Michigan, USA, to compare the leaf and flower spectra across 22 species of plants. Using a StellarNet BlueWave spectroradiometer, they recorded spectra with a resolution of 1 nm across the 350 nm – 1000 nm range and showed that flowers can provide better spectral discrimination than leaves. An example of some of the study's flower and leaf spectral curves are shown in Fig. 1. For the three different species represented, the leaf spectra are all nearly identical, however the flower spectra are distinctly different from the leaves and also from flowers of the other species.

In a study of the effect of canola flowers on LAI calculations for canola plants (Sulik and Long, 2015), ASD FieldSpecPro data were used to calibrate multispectral Unmanned Aerial Vehicle (UAV) data. Using the high spectral resolution of the ASD tool (< 10 nm), it was proven that the yellow canola flowers did affect the spectral measurements and therefore reduced the accuracy of Normalised Difference Vegetation Index (NDVI) datasets for Leaf Area Index (LAI) calculations. A 'Yellowness Index' was derived to better estimate flower coverage and therefore estimate canola yield.

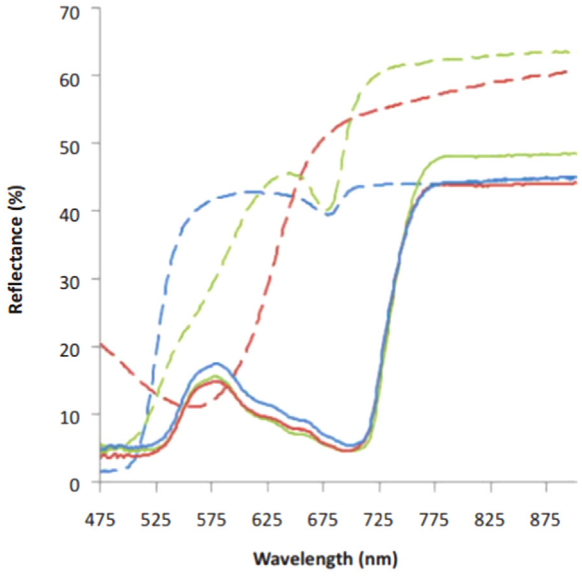


Fig. 1. Example of flower and leaf spectra from wetland species (Gross and Heumann, 2014). Leaves are solid lines and flowers are dashed lines. Green, red and blue spectra are Asclepias tuberosa, Cirsium muticum and Rudbeckia hirta respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

Fig. 2. Image of Corymbia calophylla in flower in Kings Park in March 2015 (Photo credit: Robert Campbell).

2. Visual band multispectral classification

2.1. Data collection

This study focused on the use of visual band multispectral data for identification of Corymbia calophylla flowers, as a precursor to more detailed research into the ability of remote sensing data to assist apiarists with predicting the timing and quality of honey-yield from this species. As well as producing an exceptionally large volume of nectar, the high density of flowers Corymbia calophylla produces makes the flowers clearly visible to the naked eye, even at a distance of many tens of meters, as Fig. 2 shows.

Initially a training dataset was collected, consisting of a series of digital images of three different Corymbia calophylla trees in flower in Kings Park, Western Australia. For each tree, a series of images were taken on 12/03/2015 with a Nikon D610 camera with natural light. As the objective of the study was to assess the ability to use standard drone RGB cameras for flower detection, automated exposure, aperture, ISO and F-stop settings were used to approximate how a camera would adjust settings to different light environments, flight heights and image composition. For this dataset, exposure ranged from 1/400 – 1/640 s and ISO from 400 to 1000, with an initial assessment of the colour spectrum for key image components made in the field to check whether there was any significant effect on the RGB components of key image features (flowers, leaves, etc). While there was a minor impact on feature brightness, the RGB ratios remained constant. An example of the range of images for each tree is presented in Fig. 3. A total of 11 images were used for classification training purposes.

3. Spectral clustering assessment

3.1. Parallelepiped training dataset

A simple approach, known as the 'Parallelepiped' approach (Lu and Tang, 2011), was used to assess the spectral clustering of objects in the images. This is a supervised classification method, whereby maximum and minimum values for each band of a multispectral image are defined for each class of objects. Pixels are then classified according to which minimum and maximum thresholds they fit within across all available bands of data.

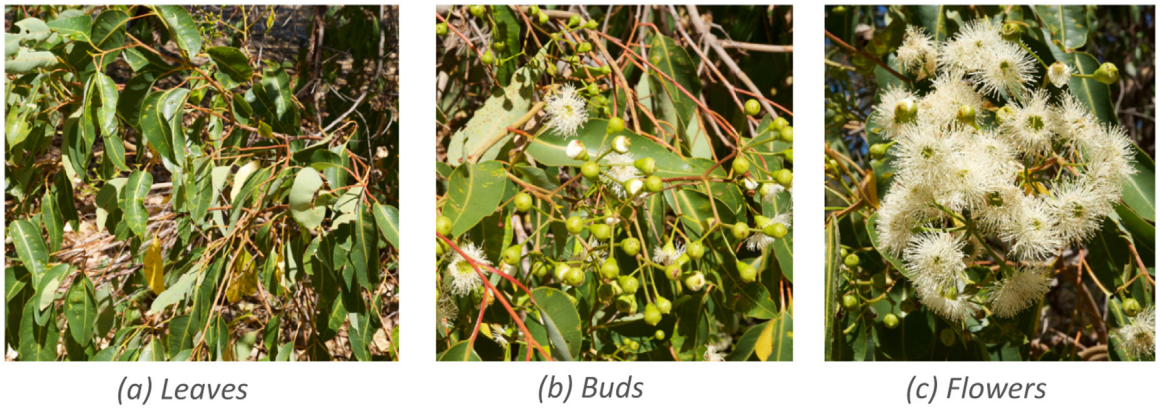


Fig. 3. Example of the types of images collected of each tree (Image credit: Robert Campbell).

Due to issues with pixels being unclassified if they are not exactly within the classification bounds, as well as overlap between maximum and minimum values for training pixels, this method is not widely adopted (Perumal and Bhaskaran, 2010). However, these limitations were not considered to have a significant impact on classification performance for this study as:

- Only a binary classification scheme is required (i.e. is the pixel a flower or not), therefore all pixels would be within the classification bands
- There was expected to be a good separation of flower pixels vs leaf pixels based on previous studies on the spectral separation of leaves and flowers (see Fig. 1 for example).
- For the purposes of detecting flowers in 3-band images, designation of pixels into multiple classes is not required, only a binary result to whether or not the pixel falls into the range of minimums and maximums for flowers as defined by the training set. Further classification of 'non-flower pixels' is not needed.

For the training dataset, image components were classified by visual interpretation of the digital images to define a range of vegetation components, including flowers, leaves, buds and branches, as well as background objects in different amounts of sunlight. Then the pixel intensity information for selected pixels from each component class were extracted from the digital files and analysed. An example of the results from one image are provided in Fig. 4(a – c). The figure shows the red, green and blue intensities for pixels in the image plotted as pairs across the five components. For the classification to be effective, the points for different components need to be in a cluster separate to other components.

A visual assessment of Fig. 4(a – c) shows that the flower pixels generally plot towards the upper right corner of the intensity range and are therefore generally separated from the majority of the other components in the image.

Fig. 4(d – f) also shows the results of the same data from Fig. 4(a – c), with the intensities from each band normalised by the total intensity of the pixel (i.e. the sum of the intensity of all three bands). Normalisation was applied to reduce the impact of different illumination on the same components across the image (Feilhauer, Asner, Martin, and Schmidtlein, 2010). As was expected from predominantly white flowers, the normalised intensity for each of the three bands for flower pixels was generally close to 33%. For the data in Fig. 4, representative of only one of the training images, a simple classification threshold of 0.29 for the normalised intensity of the blue band successfully separated the petals from all but 1 of the other training pixels (with 48 training points this results in an accuracy of 98%).

For the normalised results, the red versus green intensity data for the flowers were clustered very closely together and show how the flowers can potentially be separated from the majority of the other image components with three band minimum and maximum criteria compared with using the blue band only.

The process of classifying image components and extracting pixel intensity values was repeated for each of the 11 training images, resulting in a total of 494 training classification points. The cross-plots of red vs green vs blue data for both the raw and normalised intensities across all 11 images are presented in Fig. 5. While the flowers in full sunlight are well clustered in each of the different cross-plots, the flowers in shade display a wide range of both raw and normalised band intensity ratios. As the primary aim of this study is to look at the possibility of simple processing of airborne and spaceborne data to detect flowering Corymbia calophylla trees, only flowers in full sunlight were considered in the assessment of the accuracy of flower detection due to these datasets being acquired at near nadir orientation, the issues of flowers in shade are greatly reduced compared to taking images at ground level. To assess whether the flowers near the bottom of the tree varied in RGB space from flowers at the top, a series of pixels of flowers in full sunlight were extracted from Fig. 2 at different heights. An Analysis of Variance (ANOVA) analysis of the RBG values from flowers at the top and bottom of the tree showed that there was no statistical difference between the heights for each band ($\rho > 0.05$ in all cases).

The threshold values for flowers in full sun (summarised in Table 1) consistently span the range of approximately 0.30 – 0.33 for the normalised blue bands and approximately 0.33 – 0.35 for the normalised red and green bands.

In order to determine an optimum generic classification approach for all images, the median cutoff values in Table 1 were used, with a variation of 1 standard deviation (SD) in the minimum or maximum criteria (e.g. the normalised blue band minimum criteria was set at (0.30 – 0.02 = 0.28)). This approach resulted in the correct classification of 56 out of 60 training data points of flowers in full sun (93% accuracy), with 393 of the remaining 434 training dataset points correctly classified as non-flowers using the same criteria (90.6% accuracy). Across all 494 data points, 449 were classified correctly (90.9% accuracy).

Increasing the variation criteria to 2 standard deviations resulted in 60 of 60 training data flowers in full sun to be classified correctly but overall there was a decrease in accuracy of training data points to 445 points classified correctly of the total of 494 points (90.1% accuracy) due to the increased number of pixels from other objects incorrectly classified as flowers. As the primary objective of the classification process was to detect flowers, the 2 x SD variation criteria were applied going forward owing to the 100% accuracy in classifying flower pixels.

The same process as described for the normalised data was applied to calculating cutoff criteria for the raw intensity data from each band, with an accuracy of 94.5% achieved for a variation of 1 × SD and 92.1% for a variation of 2 x SD. A joint classification was also

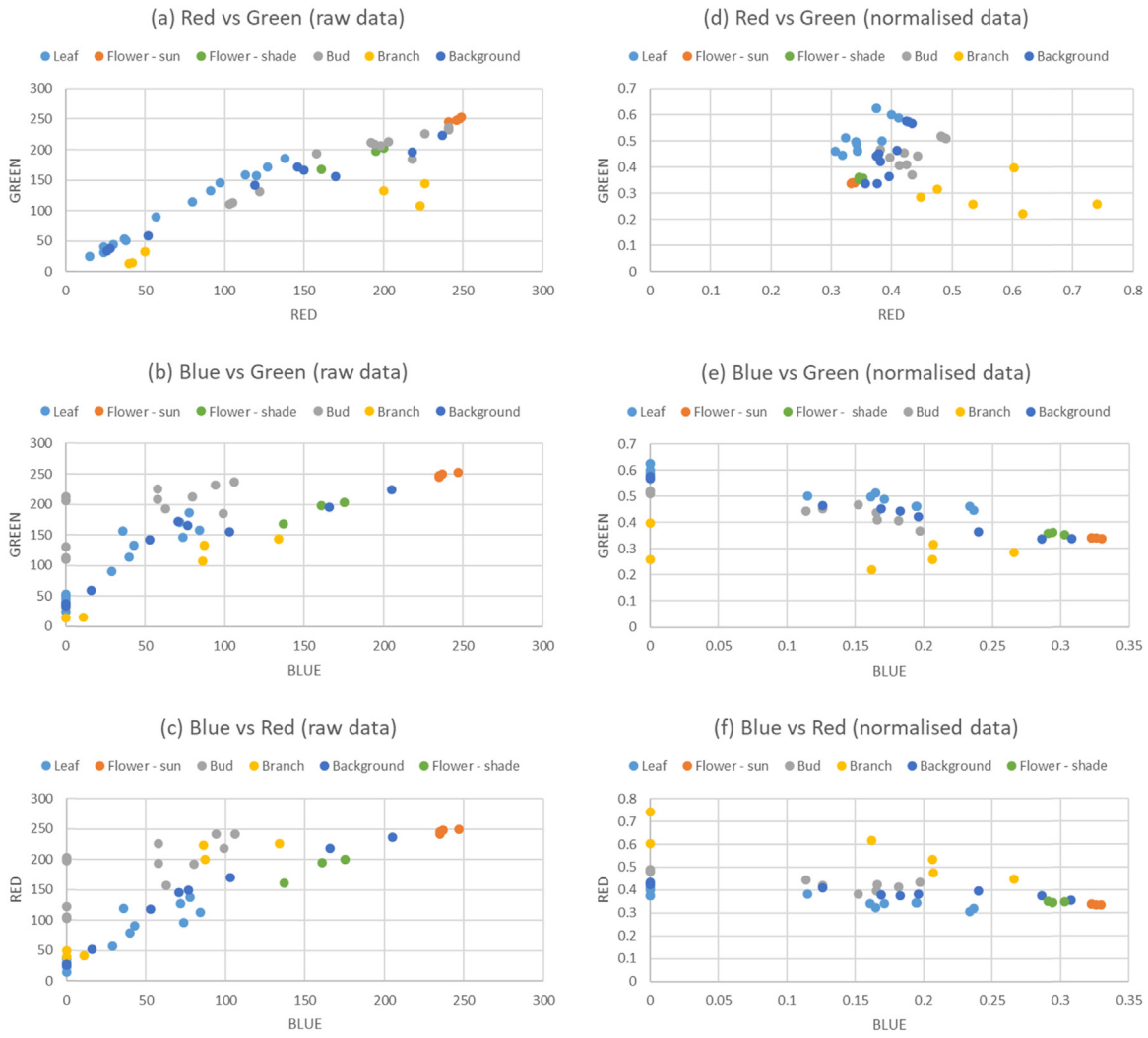


Fig. 4. Cross-plots of colour band intensities from DSC_0247. Figs. a - c represent raw data from the image and d - f are normalised against the total intensity for the pixel.

performed, using both the raw and normalised criteria data together where all 12 criteria must be met for the pixel to be classified as a flower (minimum and maximum criteria for red, green and blue bands in raw and normalised intensity data). The criteria bounds for both normalised and raw intensity data are presented graphically in Fig. 6, which is the same data from Fig. 5 zoomed into the flower data extent.

Table 2 details the classification accuracy of each image component, with the normalised, raw and joint criteria applied. The parallelepiped process classifies pixels as either flowers or non-flowers, so for the "incorrect" counts in Table 2 for each non-flower component, the count represent how many pixels were classified incorrectly as flowers. The table shows the improvement in the overall flower classification accuracy with the raw data compared to the normalised data discussed above (94.5% accurate vs 90.1% accurate respectively) and no difference between the accuracy of the raw and joint flower classification approaches. Based on this outcome, use of the raw classification criteria alone gives the most accurate result with the least amount of data processing.

In order to assess the sources of inaccuracy in more detail, the breakdown of the incorrect classification accuracy for each individual image component was assessed (see again Table 2). The results show that, for the raw data, the majority of inaccuracies are due to flower pixels being incorrectly classified as non-flower pixels (12 out of 27 incorrect pixels, or 44% of the inaccuracies). As discussed earlier, widening the criteria to include all flower pixels (sun and shade) resulted in an overall less accurate classification due to more non-flower pixels being classified as flowers so this result cannot be improved with the current dataset.

The majority of the remainder of the inaccuracies for the raw data occur due to objects that are in full sunlight (11 of the data points, or 41% of the inaccuracies). These generally occur where there is direct glare from the sun illumination into the camera lens, giving the pixel a colour close to white and therefore close to the colour of the target flowers. All non-flower components included at least one misclassified pixel in full sunlight. The remaining 4 misclassified data points are from bitumen, flowers in shade and brown background (leaf litter).

The high degree of accuracy in classification across training points in all images (94.5%) indicates good potential for detecting flowering Corymbia calophylla trees with simple, three-band visible data. There are potentially some straightforward methods to reduce the classification errors found from these images, such as:

1. Avoid paved surfaces within the survey area, or remove these pixels from the dataset
2. Collect test images to check for saturation of each band and adjust shutter and/or aperture settings if needed to reduce high-intensity pixels from non-flower sources
3. Acquire data close to nadir and direct overhead illumination as

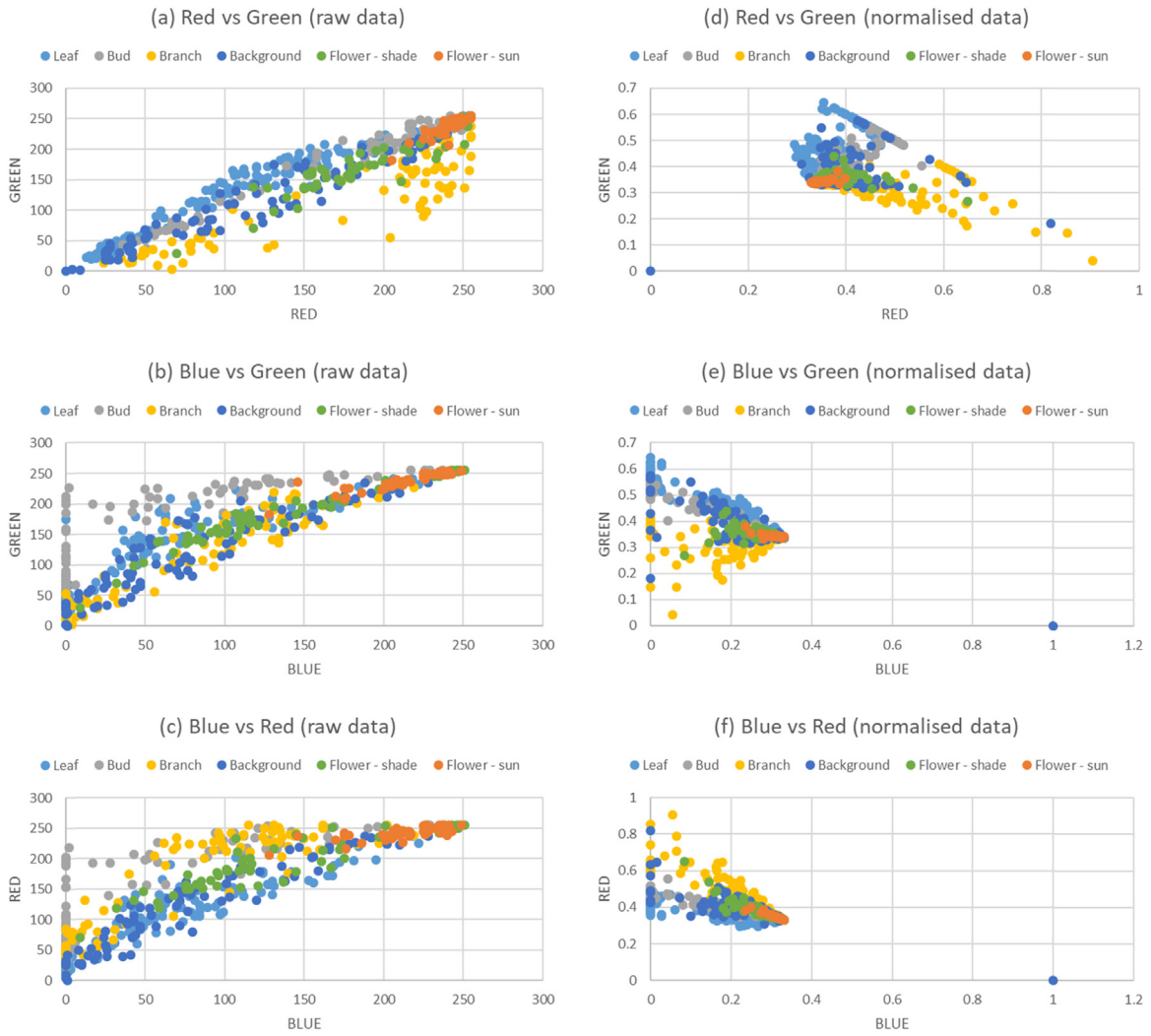


Fig. 5. Cross-plots of spectral band intensities from all images. Figs. a - c represent raw data from the image and d - f are normalised against the total intensity for the pixel.

possible, to reduce the proportion of flowers within each image that are in shade.

4. Assess viability of classifying pixels as background from data acquired outside of the flowering season and removing these pixels for subsequent datasets. This could be to remove larger areas of pixels (e.g. cleared land, roads, etc.) or to the level of extracting individual tree canopy coverage and to exclude all other pixels.

3.2. Training image classification

Following the generation of criteria for the parallelepiped algorithm based on the training data points, the criteria where used to classify all pixels for each of the images forming the training dataset. Both the normalised red, green and blue band criteria were applied to the images. An example of an original image and the normalised and raw

Table 1
Comparison of classification cutoff points and accuracy from all images.

Image	Samples	Flowers	Flowers in sun	Red min. criteria	Red max. criteria	Green min. criteria	Green max. criteria	Blue min. criteria	Blue max. criteria
247	48	7	4	0.33	0.34	0.33	0.35	0.32	0.33
260	45	8	5	0.33	0.35	0.33	0.35	0.31	0.34
261	50	10	5	0.33	0.34	0.33	0.35	0.31	0.34
269	40	8	5	0.32	0.34	0.33	0.35	0.31	0.34
281	43	10	7	0.34	0.38	0.33	0.36	0.27	0.31
290	43	6	3	0.33	0.35	0.33	0.34	0.31	0.33
294	49	12	7	0.34	0.35	0.33	0.35	0.30	0.33
297	42	12	7	0.34	0.36	0.34	0.35	0.29	0.32
301	42	14	9	0.34	0.36	0.33	0.35	0.29	0.32
308	42	7	4	0.34	0.36	0.33	0.35	0.29	0.32
316	`50	7	4	0.34	0.37	0.33	0.35	0.29	0.32
			Median	0.34	0.35	0.33	0.35	0.30	0.33
			SD	0.01	0.01	0.00	0.00	0.02	0.01
			% SD	2%	4%	1%	1%	5%	3%

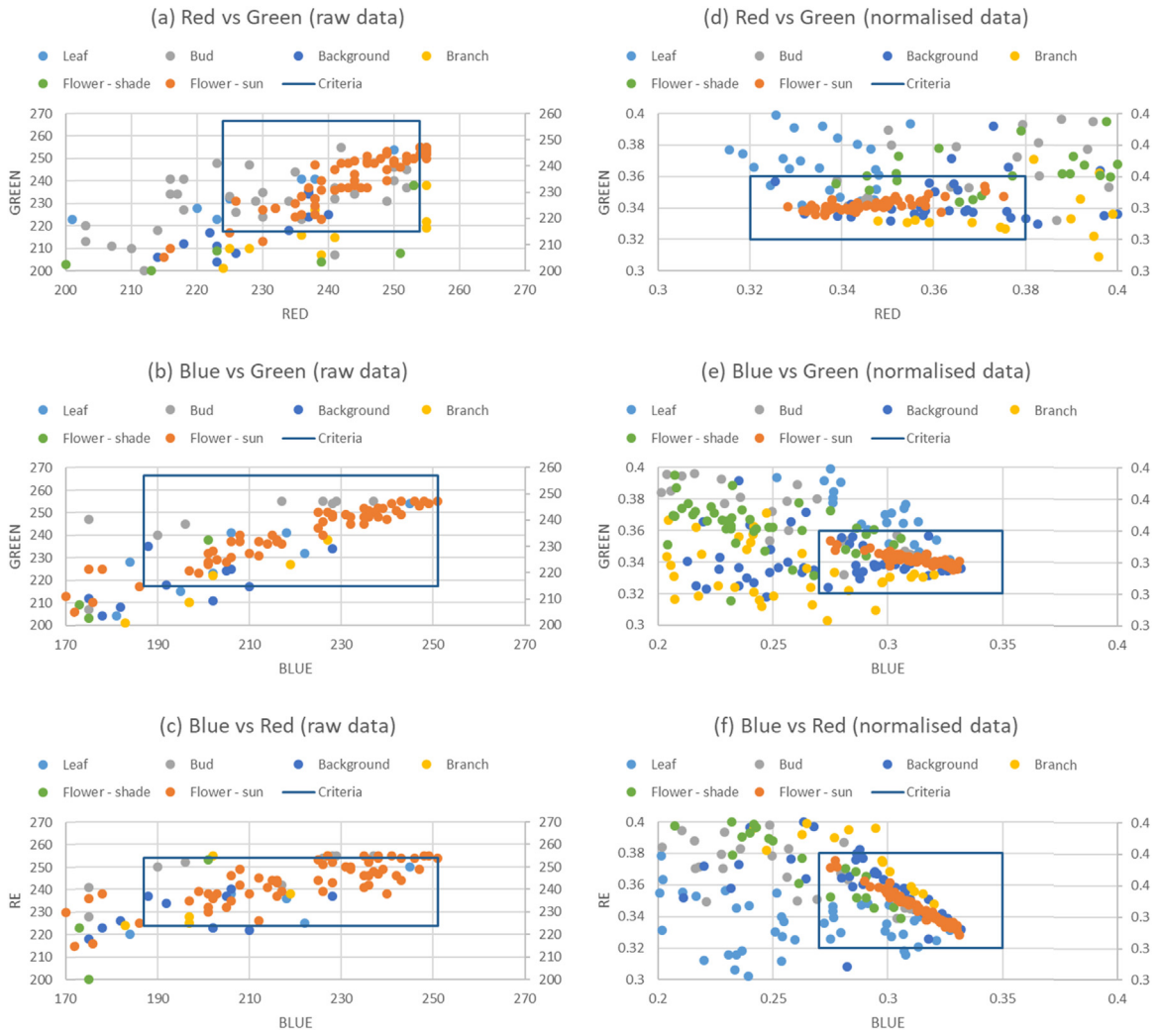


Fig. 6. Graphical representation of final classification criteria from the training dataset. Flower classification bounds are shown by the blue boxes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

classification results is shown in Fig. 7.

A visual inspection of this classification approach shows that using the normalised intensity classification by itself appears to capture the vast majority of the flowers (including those emerging from the buds), but there are numerous zones of misclassified pixels (Fig. 7b). These are primarily where the background vegetation is senesced or leaves or branches are directly reflecting bright sunlight towards the camera lens. This is due to the normalised criteria assessing only spectral quality, not the brightness of each pixel. In this scenario, pixels that are perfectly black, white or grey of similar RBG ratio will be classified as flowers. Using the raw intensity classification criteria filters out pixels that are of the correct spectral ratio but insufficient brightness and removes the majority of these misclassified zones (Fig. 7c), particularly from the grey bitumen in the upper right corner of the image.

Table 2
Assessment of sources of error in classification process. For the non-flower components, "incorrect" counts refer to pixels classified as "flower".

Component	Subcategory	Samples	Normalised incorrect	% incorrect	Raw incorrect	% incorrect	Joint incorrect
Background	Bitumen	10	10	20%	2	7%	2
Background	Bright	6	6	12%	2	7%	2
Background	Brown	33	5	10%	1	4%	1
Background	Dark	12	0	0%	0	0%	0
Background	Green	22	0	0%	0	0%	0
Branch	Full sun	53	7	14%	1	4%	1
Branch	Shade	29	0	0%	0	0%	0
Bud	Back/shade	35	0	0%	0	0%	0
Bud	Front/sun	45	7	14%	4	15%	4
Flower	Full sun	60	0	0%	12	44%	12
Flower	Shade	49	6	12%	1	4%	1
Leaf	Full sun	83	8	16%	4	15%	4
Leaf	Shade	57	0	0%	0	0%	0
	Total	494	49		27		27
			9.9%		5.5%		5.5%

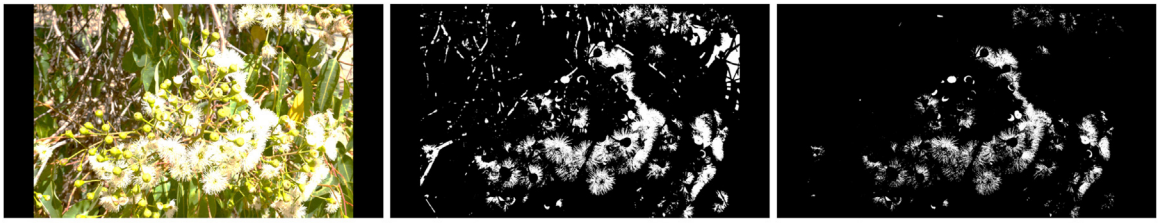


Fig. 7. Original image (a) and classified raster images for the original imageDSC_301, using normalised RGB intensity (b) and raw RGB intensity classification (c) from the spectral clustering analysis.

Across all 11 images, similar results were found as those shown in Fig. 7, with the normalised classification working well by correctly classifying the vast majority of flower pixels but leaving some obvious non-flower pixels selected as flowers. The raw classification method slightly reduced the number of correctly classified flower pixels but dramatically reduced the number of other pixels incorrectly classified as flowers.

Notwithstanding the fact that all training images were collected within a few hours of each other on a single day, and all collected in a limited spatial region, these classification results prove that a simple 3-band parallelepiped classification scheme, applied across different images of the style and of different trees of the same species, can accurately spectrally separate flowers from foliage and background.

3.3. Classification performance of non-flowering images

In order to further assess the potential issues with background material such as leaf litter and soil being incorrectly classified as flowers, images collected in a Corymbia calophylla forest in Mundaring, Western Australia, in September 2015 were classified using the same approach as outlined previously for Fig. 7. As the images were captured in spring, no Corymbia calophylla trees were in flower, and thus all data points classified as flowers are due to errors in background pixel classification.

The results for a selection of these non-flower images, shown in Fig. 8 - Example classification results from images with no Corymbia calophylla flowers Fig. 8, show that the raw classification criteria based on the Kings Park flower images worked well on images of leaves and gravel in Mundaring and moderately well on images of leaf litter and senesced vegetation. There is a marked decrease in classification accuracy if only the normalised criteria are used. As discussed above, this is due to the lack of a brightness filter in the normalised classification approach, where pixels having a spectral ratio close to 1:1:1 between the three bands are classified as flowers.

Overall the raw classification criteria worked close to or better than the results from the Kings Park training dataset (accuracies of 94.1–99.8%, against an average of 94.5% at Kings Park). Given that the normalised classification was as low as 40.9% accuracy against background material from a Corymbia calophylla forest, only the raw classification criteria are used from here on.

3.4. Classification performance against independent flowering images

To assess the performance of the parallelepiped flower classification method more robustly, a series of images of Corymbia calophylla in flower were collected in early 2016 (the following flowering season from the 2015 training dataset), using a Nikon D90 and an iPhone 6S camera and at different locations to the training datasets. These images are used to test how effectively the algorithm works under some data capture scenarios different to those for the initial training images.

The results from the parallelepiped classification, shown in Fig. 9, show that flowers are clearly detected in all cases. However, the classified images do show some areas of misclassified pixels, primarily from bright areas (operator's arm in the top image and reflections of buds and leaves in the lower image), confirming that pixels close to maximum intensity do cause errors in the classification process.

The overall accuracy of these classifications is higher than the training images for the joint classification process (see example in Fig. 7) and indicates that the parallelepiped classification approach with the minimum and maximum values determined from the training images are broadly applicable to the other data capture devices, seasons and locations (i.e. not just to images acquired of particular trees in Kings Park in 2015, using a Nikon D610 camera).

3.5. Assessment of impact of lower resolution

The classification methods assessed so far have all been with high spatial resolution images, with individual stamens from the Corymbia calophylla flowers clearly visible in the original high-resolution images. However, with 'high-resolution' aerial imagery collected from flight heights of approximately 100 – 150 m altitude the resulting pixel widths are slightly better than 5 cm (Barrell and Grant, 2015). By contrast, a typical flower size of Corymbia calophylla is 1.3 – 4.5 cm (Brooker and Kleinig, 2001). Clearly, aircraft cannot produce images with resolutions comparable to our training images used so far in the detection of Corymbia calophylla flowers.

To assess the limits of resolution in the parallelepiped method, a series of 6 images of Corymbia calophylla flowers were progressively degraded in spatial resolution to approximately 1 pixel per flower. Fig. 10 shows two pairs of original images and associated degraded images as an example of this process. Fig. 11 shows graphs representing the accuracy of flower detection with change in image resolution for each of the six images. Image spatial resolution increases from left to right in each plot, with the median flower cluster size for each image stated. The median cluster size for each image was calculated from the manually drawn polygons representing individual clusters of flowers. The area of each individual polygon in pixels was calculated and divided by the average number of pixels per flower (from the mean of the number of pixels per flower from 10 random flowers chosen from the image) and the median flowers per cluster calculated from this distribution.

Fig. 11 shows that with decreasing image resolution (from right to left on each plot) there is a decrease in flower detection accuracy, but there is also an increase in non-flower (background) classification accuracy. For cluster sizes from 1.0 to 1.9 the flower detection accuracy is close to 100% down to a resolution of approximately 10 pixels per flower, at which point the accuracy rapidly decreases to very low values (0–20%). For the larger flower cluster sizes of 4.1 and 9.1 the flower detection accuracy decreases consistently from high to low image resolution but remains at least 60% accuracy at an image resolution of 1 pixel per flower.

The trend in improvement in background accuracy with decreasing image resolution was not consistent for each graph. There was typically a sharp increase in accuracy at an image-specific resolution, but this point of improvement varied between 1 and 100 pixels per flower.

The range in accuracy at high image resolution (right hand side of

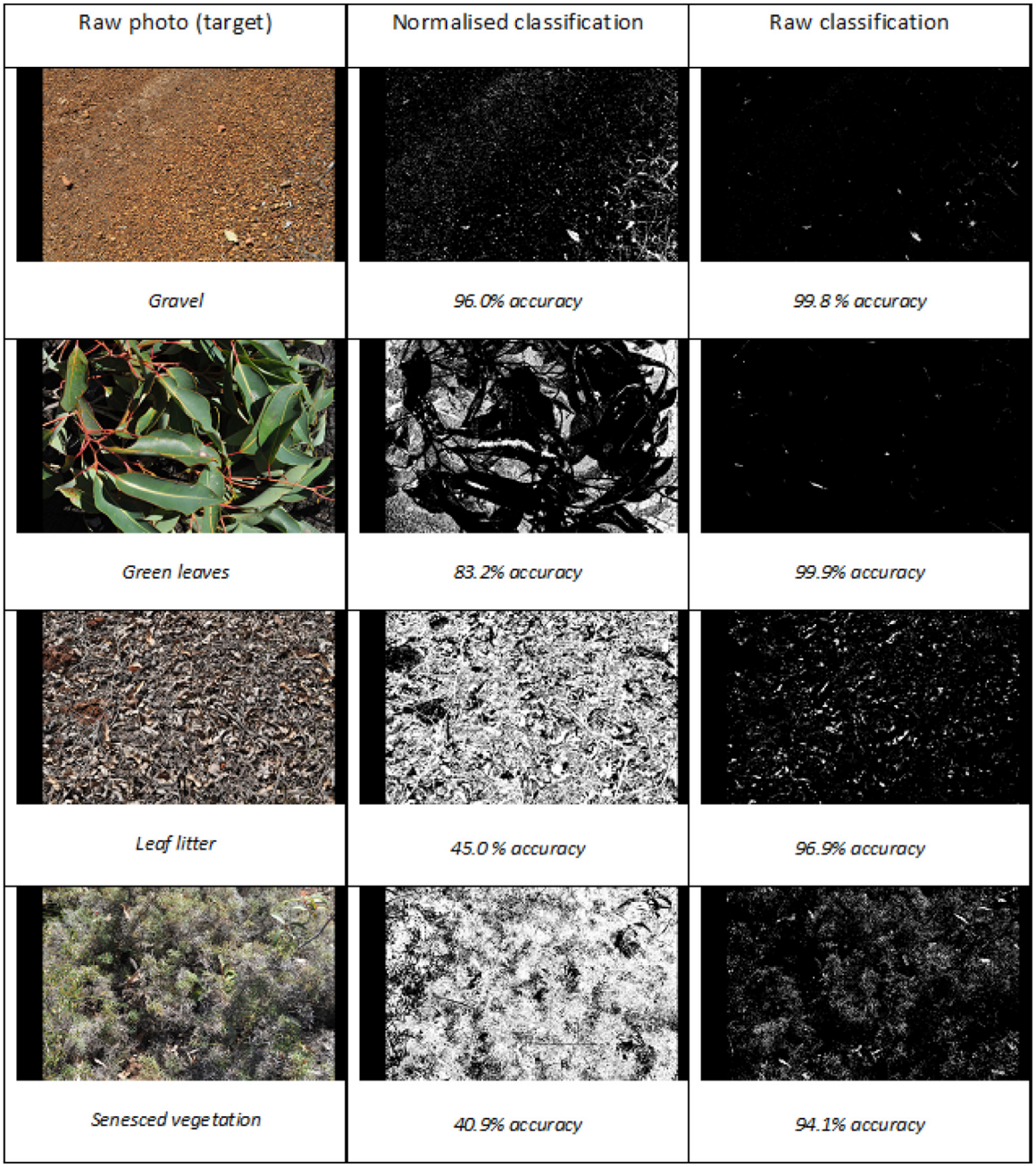


Fig. 8. Example classification results from images with no Corymbia calophylla flowers, with percentage accuracy noted.

each graph) between the six graphs for flower and non-flower classification accuracy is due to the composition of the higher resolution images; images with lower flower detection accuracies had a higher proportion of flowers in part shade and lower non-flower classification accuracies were primarily due to large areas of sand or bitumen within the image (refer back to Table 2).

Examples of small and larger cluster images are provided in Fig. 10. These examples are cropped areas of larger images to highlight the difference in classification outcomes. An examination of the progressively degraded images shows that the detection of flowers in larger clusters is due to the larger clusters effectively representing larger flowers in terms of target identification. While small clusters and individual flowers are not detected at low resolution, the larger clusters of flowers in the image are still detected due to the larger number of near-white pixels from the cluster.

The curves in Fig. 11 show that the resolution at which background classification improves varies, and can occur at either higher or lower than the resolution at which flower detection deteriorates. Accordingly the resolution required for the detection of flowers should be used as the primary decision tool for image resolution selection, with the background classification accuracy being of secondary importance. In practice the calculated flower occurrence in an image is determined by the combination of correctly detected flowers and incorrectly classified background pixels. The accuracy of the flower occurrence for an image is thus determined by the image resolution and by the proportion of flowers to background in the image.

To assess how the average flower cluster size, image resolution and proportion of flowers affects the ability of the algorithm to detect flowers against the background, the curves in Fig. 12 were constructed from images DSC_0881 (1.5 median flower cluster size) and DSC_0296 (9.1 median flower cluster size). The different curves in each graph show the calculated percent flower cover in the image versus different

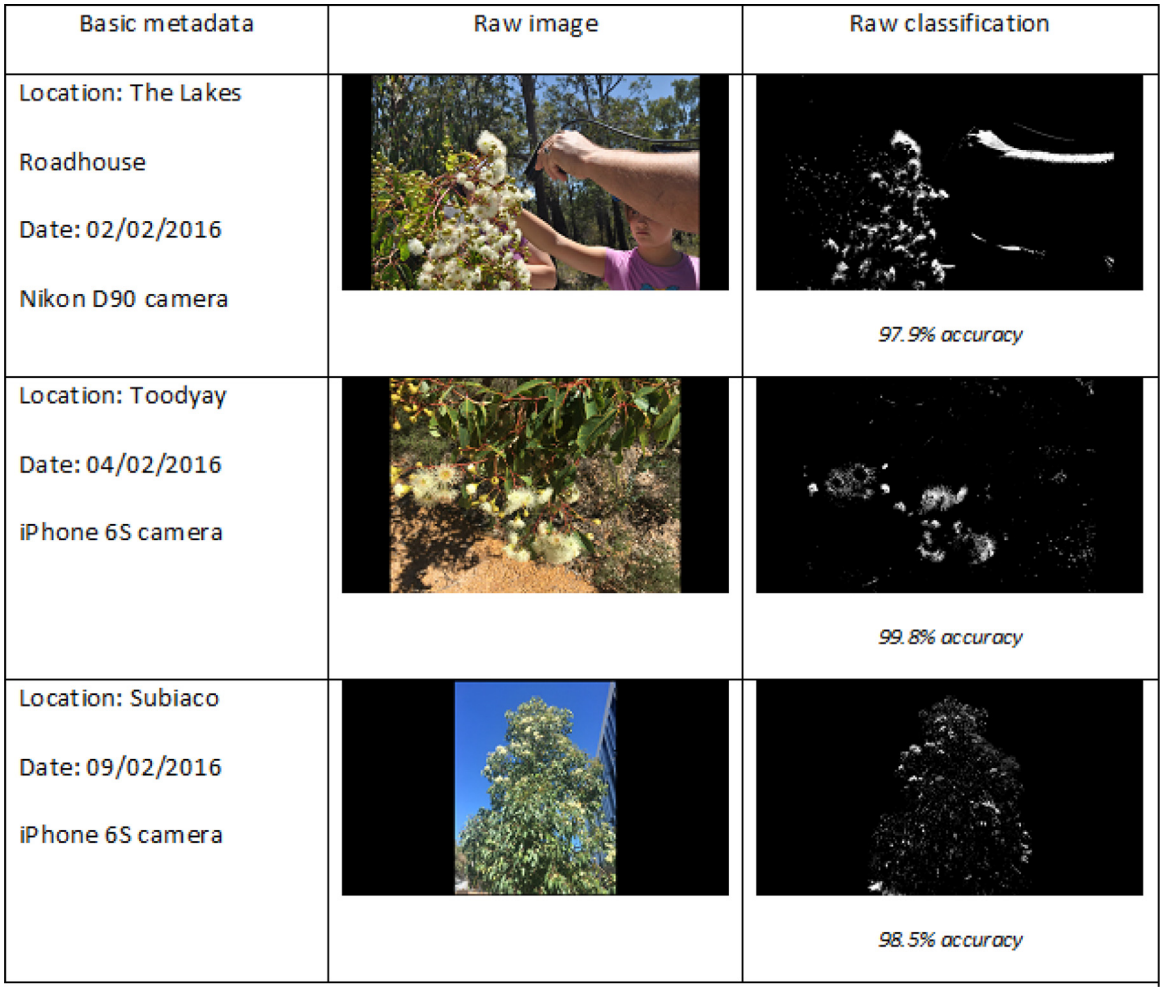


Fig. 9. Classification results from independent flowering images. Accuracy determined by manually creating a mask of all flower pixels for each image and calculating the percentage of non-flower pixels that were classified as flowers.

image resolutions for actual flower percentages ranging from 0.5% to 25%. The actual proportion of flowers in the images was determined manually.

The curves all rise sharply from 1 pixel per flower resolution up to approximately 10 pixels per flower and only increase gradually after this point. The sharp rise at low pixel per flower resolutions is less pronounced for the larger flower cluster size (compare left and right graphs). All curves effectively follow a line close to (% flower coverage) + (% background classification error). At low percentage flower cover, represented by the lower curves, the background error percentage dominates and at higher flower cover, represented by the upper curves, the actual flower percentage dominates. At 0.5% flower pixels per image, the classified percent flower cover is effectively the misclassified background pixels, or background error, for that resolution, close to 3.3% for the left hand and 9.9% for the right hand graph.

Three key findings can be drawn from this analysis:

1. The minimum resolution for effective use of this algorithm is 10 pixels per flower.
2. Increasing the resolution to greater than 10 pixels per flower does improve accuracy but to a minimal degree.
3. Background error is a significant factor, particularly for lower percentages of flower cover.

To equate the limits of detection to survey design for a drone, using the DJI Phantom 3 Professional drone with:

- Resolution 4000 × 3000 pixels.
- Field of view of 94 degrees.
- Assumed flower size of 15 mm diameter.

A resolution of 10 pixels per flower this equates to a flight height of 86.2 m above the canopy or 43.1 m height for a resolution of 20 pixels per flower. With a maximum flight height of 120 m above ground level for Very Small Remotely Piloted Aircraft in Australia before extensive certification is required by the Civil Aviation Safety Authority (CASA) in Australia (Remotely piloted aircraft systems, operation of excluded RPA (other than model aircraft), 2016) and a general maximum height of Corymbia calophylla trees of 40 m (Herbarium, 2015), drone flight heights for most operators are restricted primarily by UAV compliance requirements rather than resolution. Therefore all flights within CASA requirements for non-certified UAV operators should fall within the recommended minimum resolution for optimum detection of flowers.

3.6. Impact of other species flowering at the same time

In a forest setting, potentially other flowering plants may have an impact on the detection of Corymbia calophylla in flower. An interrogation of the floral database maintained by the Western Australian Government (Herbarium, 2015) shows that:

- 2311 other species occur within 1 km of surveyed occurrences of Corymbia calophylla.
- 108 other species occur at more than 25% of the Corymbia

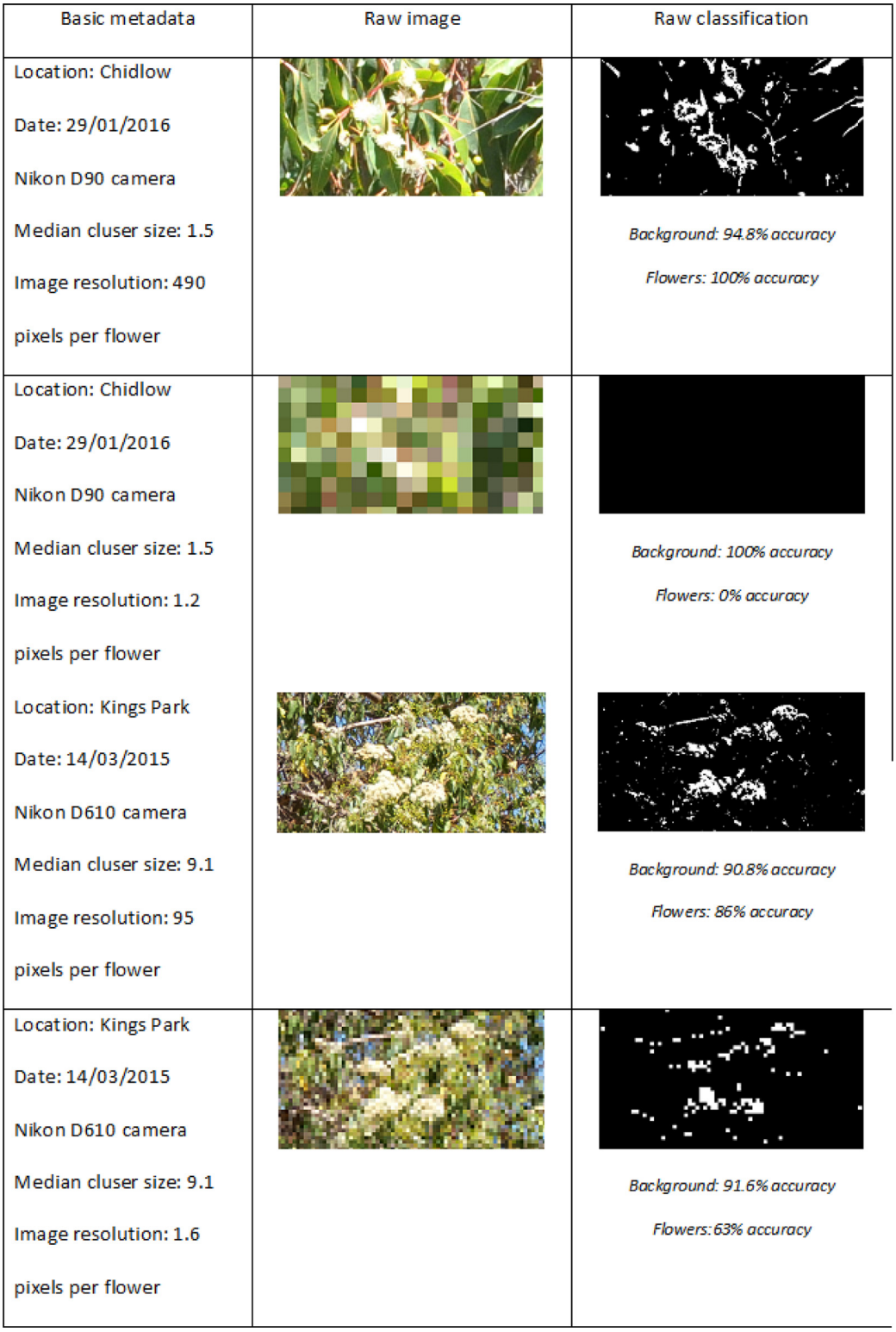


Fig. 10. Comparison of lower resolution classification for individual flowers vs larger clusters.

calophylla locations.
- Two-thirds of these species have non-white flowers.
- 97 of co-occurring species are shrub or herb in form, i.e. of limited spatial coverage).
- Two species flower at overlapping times with white flowers and are trees: Eucalyptus marginate (Jarrah) and Eucalyptus wandoo (Wandoo or White Gum).

These two species generally either flower earlier in the season (Jarrah) or on different soil types (Wandoo or White Gum), which would limit the impact on detection of Corymbia calophylla flowers in single survey datasets. In the application of routine or repeated surveys of the same region, a database of plant types and specific geospatial maps of the extent of Corymbia calophylla trees could be developed over time, using on ground surveys or visual interpretation of the high resolution imagery.

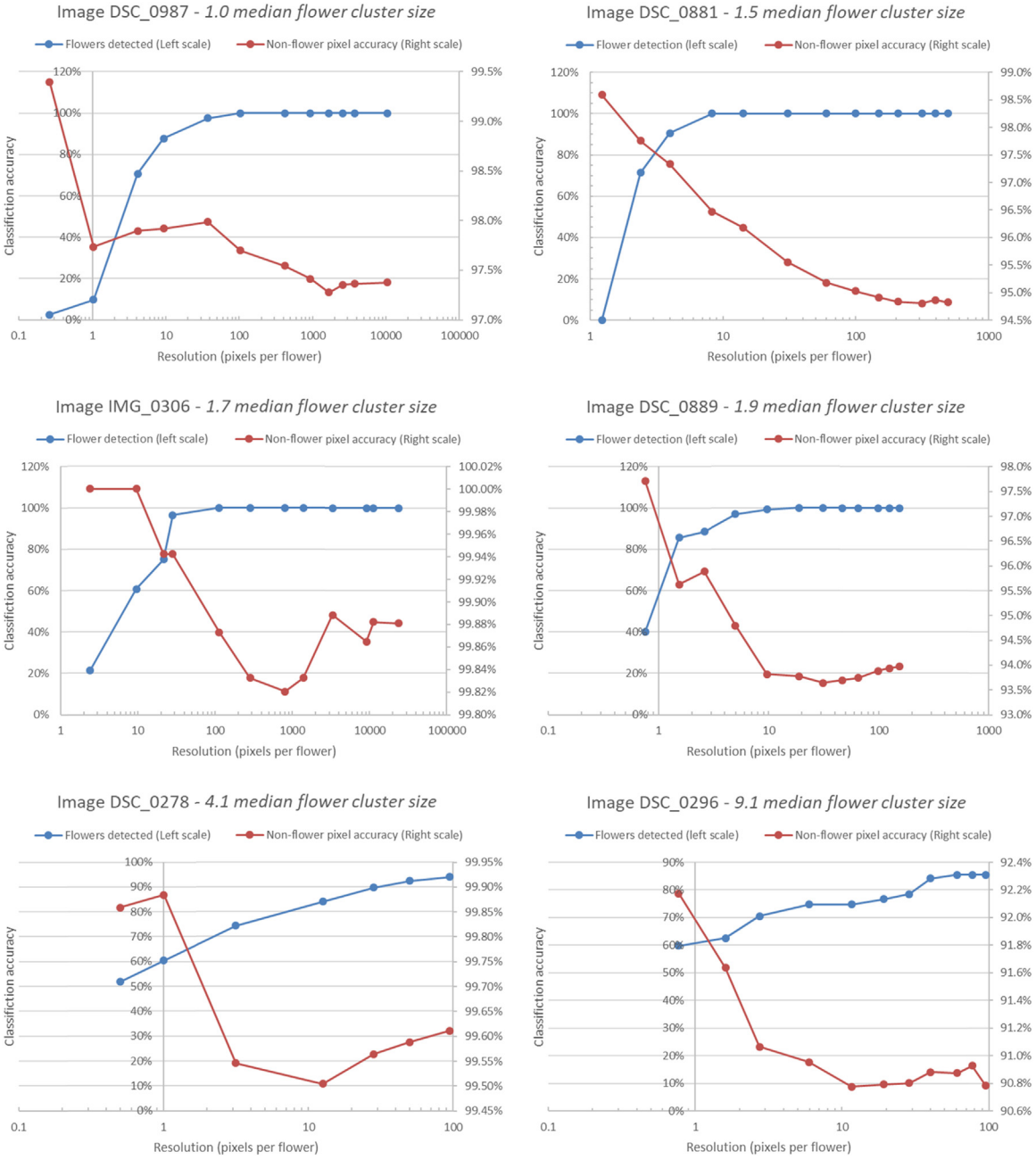


Fig. 11. Graphs of background classification accuracy and percentage of flowers detected against resolution (defined as average pixels per flower). The average flower cluster size is indicated for each graph.

Based on this initial assessment, even in the internationally recognised biodiversity hotspot of the Southwest Australian Floristic Region (Gole, 2006) potentially many co-occurring species can be disregarded from errors sources based on their size, form, flowering colour or flowering season. However, this needs to be considered in more detail in real-world aerial datasets.

4. Conclusions

The process of collecting high resolution images of Corymbia calophylla flowers and the surrounding environment and applying a simple parallelepiped classification algorithm has shown a clear spectral separation between the flowers and the majority of the materials that generally occur in the same area when the raw intensity of the red, green and blue bands of data are used.

Errors in the classification process come primarily from areas of bright reflections of direct sunlight onto the camera lens and from leaf litter and senesced vegetation. Potentially leaf litter can be excluded from the data by conducting a survey in winter months and applying a spatial filter to remove areas classified as leaf litter for datasets acquired over the flowering period for Corymbia calophylla (December – April).

We have demonstrated the potential of simple, low cost sensors to be used for collecting multi-spectral data for use in classification and detection of Corymbia calophylla flowers. Although the images we analysed were captured using hand held devices, we also assessed the impact on flower classification due to image spatial degradation, as would occur if data were collected from a UAV and have developed a representative survey design specification to this end.

We showed that classification accuracy varies with image spatial resolution, and that the size of flower clusters present on the tree also

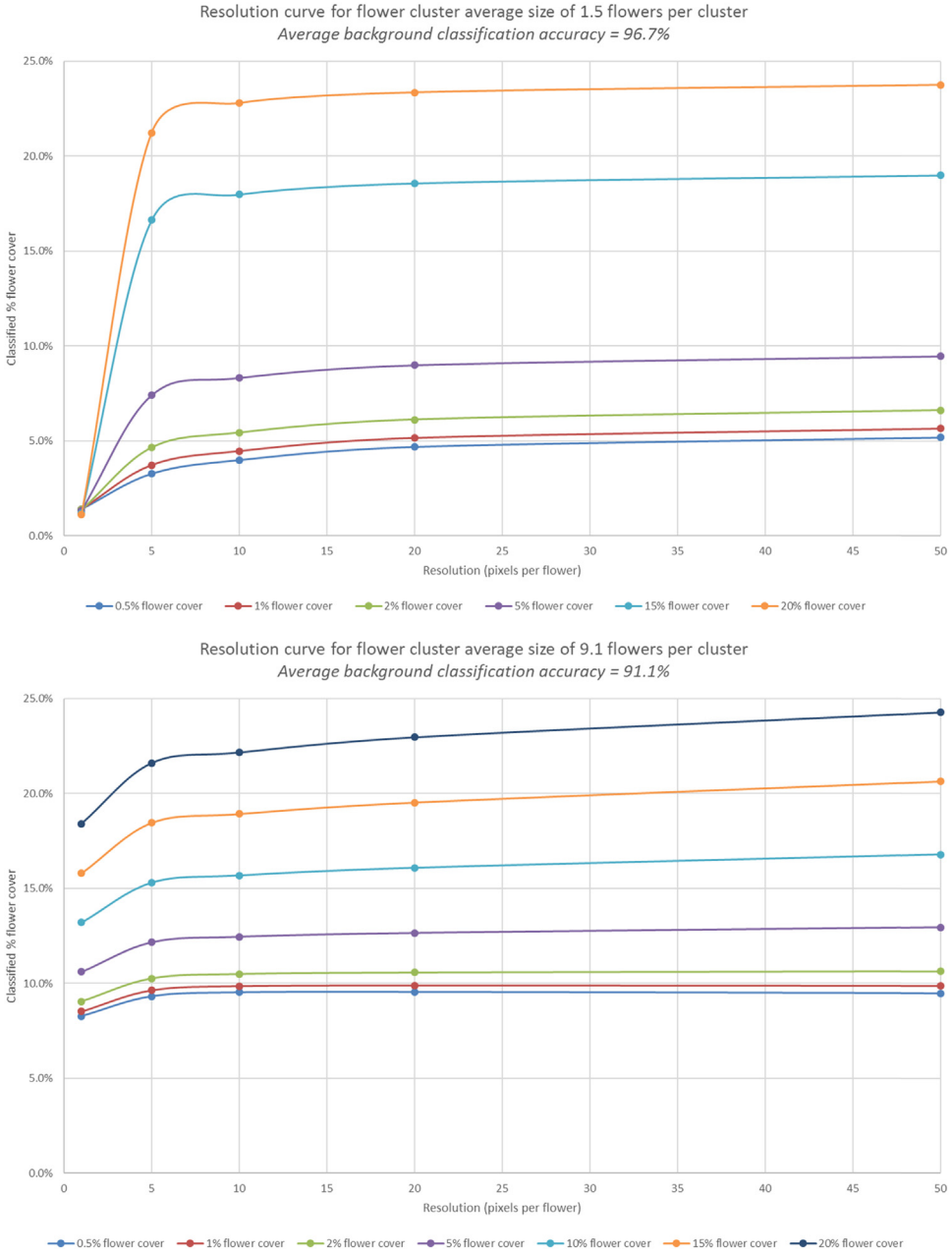


Fig. 12. Resolution vs percentage flower classed pixels for two different average flower cluster sizes.

affect the classification accuracy. These results provide support for the potential of simple 3 band cameras mounted on UAVs for undertaking surveys of flowering onset and flowering density of Corymbia calophylla trees. This potential application is of particular relevance to the management of bee hives. With bees typically foraging an average of less than 1 km from the hive (Hagler et al., 2011), UAVs could be used to optimise the location of beehives during a honey flow if multiple sites are available in the area.

Based on an assessment of the effect of image resolution on classification accuracy, flights at up to 100 m elevation should have sufficient resolution to detect flowers. The accuracy of this approach is primarily determined by the background classification error. If the background classification error is known for an image or scene, the percentage flower cover can be calculated with as little as 2% flower cover.

We have not considered the corrections in detail for the situation where other flowering plants may occur in close proximity to the Corymbia calophylla trees. In this case, if the other plants produced white or near-white flower, the simple parallelepiped classification could not be used to separate the different species. However, in the application of routine or repeated surveys of the same region, a database of plant types and specific geospatial maps of the extent of Corymbia calophylla trees could be developed over time, using on ground surveys or visual interpretation of the high resolution imagery.

References

Barrell, J., Grant, J., 2015. High-resolution, low-altitude aerial photography in physical geography: a case study characterizing eelgrass (Zostera marina L.) and blue mussel (Mytilus edulis L.) landscape mosaic structure. Prog. Phys. Geogr. 39 (4), 440–459. http://dx.doi.org/10.1177/0309133315578943.

Brooker, M.I.H., Kleinig, D.A., 2001. *Field guide to eucalypts.* (Vol. Volume 2, South-Western and Southern Australia). Bloomings Books, Melbourne, Australia.

Carvalho, S., Schlerf, M., van der Putten, W., Skidmore, A., 2013. Hyperspectral reflectance of leaves and flowers of an outbreak species discriminates season and successional stage of vegetation. Int. J. Appl. Earth Obs. Geoinf. 24, 32–41. http://dx.doi.org/10.1016/j.jag.2013.01.005.

Chittka, L., Shmida, A., Troje, N., Menzel, R., 1994. Ultraviolet as a component of flower

reflections, and the colour perception of Hymenoptera. Vision. Res. 34 (11), 1489–1508.
Cochrane, M.A., 2000. Using vegetation reflectance variability for species level classification of hyperspectral data. Int. J. Remote Sens. Appl. 21 (10), 2075–2087. http://dx.doi.org/10.1080/01431160050021303.
Dyer, A., Boyd-Gerny, S., McLoughlin, S., Rosa, M., Simonov, V., Wong, B., 2012. Parallel evolution of angiosperm colour signals: common evolutionary pressures linked to hynemopteran vision. Proc. R. Soc. 279, 3606–3615. http://dx.doi.org/10.1098/rspb.2012.0827.
Feilhauer, H., Asner, G., Martin, R., Schmidtlein, S., 2010. Brightness-normalized partial least squares regression for hyperspectral data. J. Quant. Spectrosc. Radiat. Transf. 111, 1947–1957. http://dx.doi.org/10.1016/j.jqsrt.2010.03.007.
Gole, G., 2006. The Southwest Australian Ecoregion – Jewel of the Australian Continent, Southwest Auatralia Ecoregion Initiative.
Gross, J., Heumann, B., 2014. Can flowers provide better spectral discrimination between herbaceous wetland species than leaves? Remote Sens. Lett. 5 (10), 892–901. http://dx.doi.org/10.1080/2150704X.2014.973077.
Hagler, J., Mueller, S., Teuber, L., Machtley, S., Van Deynze, A., 2011. Foraging range of honey bees, Apis mellifera, in alfalfa seed production fields. J. Insect Sci. 11 (144).
Herbarium, W.A., 2015. Florabase. Retrieved from ⟨https://florabase.dpaw.wa.gov.au/⟩.
Lu, Q., Tang, M., 2011. Detection of Hidden Bruise on Kiwi fruit Using Hyperspectral Imaging and Parallelepiped Classification. Paper presented at the International Conference on Environmental Science and Engineering, Indonesia.
Perumal, K., Bhaskaran, R., 2010. Supervised classification performance of multispectral images. J. Comput. 2 (2), 124–129.
Portigal, F., Holasek, R., Mooradian, G., Owensby, P., Dickson, M., Fene, M., Driggett, D., 1997. Vegetation classification using red-edge first derivative and green peak statistical moment indices with the Advanced Airborne Hyperspectral Imaging System (AAHIS). Paper presented at the Third International Airborne Remote Sensing Conference and Exhibition, Copenhagen, Denmark.
Price, J., 1994. How unique are spectral signatures? Remote Sens. Environ. 49, 181–185.
Remotely piloted aircraft systems – operation of excluded RPA (other than model aircraft), 2016. Advisory Circular 101-10, Civi Aviation Authority, Australian Government.
Shrestha, M., Dyer, A., Boyd-Gerny, S., Wong, B., Burd, M., 2013. Shades of red: bird-pollinated flowers target the specific colour discrimination abilities of avian vision. New Phytol. 198, 301–310. http://dx.doi.org/10.1111/nph.12135.
Sulik, J., Long, D., 2015. Spectral indices for yellow canola flowers. Int. J. Remote Sens. 36 (10), 2751–2765. http://dx.doi.org/10.1080/01431161.2015.1047994.
Thenkabail, P., Smith, R., De Pauw, E., 2000. Hyperspectral vegetation indicies and their relationships with agricultural crop characteristics. Remote Sens. Environ. 71, 158–182.
Westman, W., Price, C., 1987. Remote detection of air pollution stress to vegetation - Laboratory-level studies. Paper presented at the International Geoscience and Remote Sensing Symposium, Michigan, USA.
Zomer, R., Trabucco, A., Ustin, S., 2009. Building spectral libraries for wetlands land cover classification and hyperspectral remote sensing. J. Environ. Manag. 90, 2170–2177. http://dx.doi.org/10.1016/j.jenvman.2007.06.028.

4 Published Paper: Honey crop estimation from space: Detection of large flowering events in Western Australian forests

Campbell T, Fearns P (2018) Honey crop estimation from space: Detection of large flowering events in Western Australian forests. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-1, pp. 79–86. https://doi.org/10.5194/isprs-archives-XLII-1-79-2018

HONEY CROP ESTIMATION FROM SPACE: DETECTION OF LARGE FLOWERING EVENTS IN WESTERN AUSTRALIAN FORESTS

T. Campbell [1*], P. Fearns [1]

[1] Remote Sensing and Satellite Research Group, Curtin University, Kent St, Bentley, 6102, Australia – tristan.campbell@postgrad.curtin.edu.au, p.fearns@exchange.curtin.edu.au

KEY WORDS: Corymbia calophylla, phenology, spectroradiometer, ANOVA, JM Distance, MODIS, honey yield

ABSTRACT:

Recent studies have shown that in the spectral space there is often a better spectral separation between leaves and flowers and even between flowers of different species than between leaves of different species. In this study we assess the ability of satellite remotely sensed data to detect the flowering of Red Gum trees (*Corymbia calophylla*) in Western Australia, the state's largest annual honey crop. Spectroradiometer measurements of flowers, leaves and groundcover from Red Gum forests were subjected to ANOVA analysis, which showed that flowers are spectrally different to their environment for 92% of the wavelengths between 350 nm and 1800 nm. A more detailed assessment, using the JM Distance calculation, showed that the spectra can be reliably separated using 10% of the wavelengths, with peak separation between 518 nm and 557 nm. To assess the ability of satellite-borne sensors to detect the presence of flowers, the spectroradiometer data were convolved with satellite instruments' response curves to create synthetic remotely sensed datasets on which JM Distance analysis was performed. MODIS blue bands achieved a median JM Distance of greater than 1.9 and therefore should be able to detect the presence of flowers from the environment. Further assessment showed that the shortest wavelength bands for MODIS, VIIRS and Sentinel 3 all occur where the flower spectra have lower reflectance than their natural background. A sensitivity analysis of percentage flower cover for a pixel showed that the highest sensitivity was obtained by dividing the band closest to 520 nm by the shortest wavelength band for data from these three sources. The MODIS band 10/band 8 metric was tested for its ability to detect flowers in real-world data using 15 years of qualitative honey harvest data from one apiary site as a proxy for flower density. This test was successful as, while there was some overlap between good, moderate and poor years, the poor years could be separated from the other years with nearly 80% accuracy.

1. INTRODUCTION

Remote sensing classification relies on the spectral response of different objects of interest having key spectral regions of difference, and those differences being measurable within the detection limits of the system. While often used to map vegetation types, the accuracy of remote sensing can be limited in this application due to the overlap of spectra between different species (Price, 1994). Indeed, the variation in spectra between individual leaves on a single tree can be greater than the difference in spectra between species in some cases (Cochrane, 2000).

Recent studies have shown that in the spectral space there is often a better spectral separation between leaves and flowers (Gross and Heumann, 2014) and even between flowers of different species than between leaves of different species (Shrestha et al., 2013).

Even with the limited spectral dimensionality of a standard 3-band Digital Single Lens Reflex (DSLR) camera, it can be possible to clearly differentiate between flowers and leaves of some species (Campbell and Fearns, 2018), however the appropriate spatial resolution required for effective spectral separation is not typically available except via the use of Unmanned Aerial Vehicle (UAV) platforms. As a consequence, low spectral resolution sensing is not considered suitable for large spatial scale or high temporal frequency mapping.

In this paper, we explore the potential for the use of imagery with a higher spectral resolution but lower spatial resolution than 3-band DSLR data for the detection of flowering *Corymbia calophylla* (Red Gum) trees in Western Australia. Honeybees foraging on this species produce the state's largest annual honey crop (Painter, 2010) and can produce honey with some of the highest antibacterial activity of any honey variety in Australia (Irish et al., 2011). As the species occurs across a large portion of the South-West Floristic Province (see Figure 1), the ability to remotely detect where Red Gum trees are flowering may help apiarists to better manage their seasonal apiary movements and thereby increase production of higher-value honey.

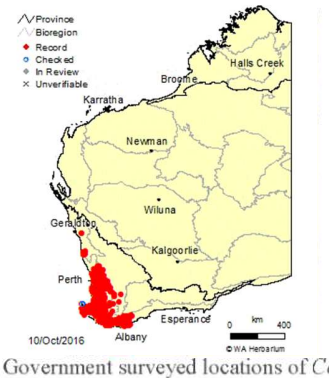


Figure 1. Government surveyed locations of *Corymbia calophylla* (Herbarium, 2015)

* Corresponding author

Figure 2. Locations of three registered apiary and spectral measurement sites, indicated by "pin" locations (Google, 2016)

2. DATA COLLECTION

Spectral reflectance data were acquired at three sites in State Forest areas between Spring 2015 and Autumn 2016 using an ASD FieldSpecPro 3 spectroradiometer. The location of the three sites, that are registered apiary locations with the Department of Biodiversity, Conservation and Attractions (DBCA), are shown in Figure 2.

A total of 325 reflectance spectra of leaves, groundcover and flowers were acquired on clear, sunny days in September 2015 (spring) and February 2016 (summer – peak flowering period) between 10 am and 2 pm. White reference measurements (Spectralon) were collected at least every 10 minutes (as per ASD manual) and 10 records taken and averaged per measurement. A minimum of 5 measurements were collected per target, with the spectroradiometer optic fibre held by hand between 5 – 10 cm from vegetation targets and 1 m height above ground targets. Typically 2 or 3 different targets were measured on individual trees (e.g. multiple different leaf or flower clusters). Measurements of groundcover spectra were also taken, which included leaf litter, low vegetation (< 0.5 m tall), gravel and sealed roads.

2,151 spectral bands were acquired at between 350 – 2,500 nm wavelengths. Spectral data at wavelengths greater than 1,800 nm were deleted to remove atmospheric water absorption features. The spectral resolution was also reduced from the 1 nm bandwidth exported by the ASD software to the actual spectral resolution of the sensors. That is, 3 nm from 350 nm to 1,000 nm and 10 nm from 1,000 nm to 2,500 nm. Spectra were also corrected for steps in sensor brightness calibration (Dallon, 2003).

While some studies have normalised individual reflectance spectra to reduce the impact of differences in brightness from target orientation and illumination (Feilhauer et al., 2010), this was not done for this study as work on multi-band DSLR data has shown a better discrimination of flowers in raw reflectance data (Campbell and Fearns, 2018) due to the overall pixel intensity being able to separate pixels close to white rather than grey or near black pixels with the same normalised ratio of red, green and blue reflectance. Accordingly, additional care was taken to only record spectra of targets in full, direct sunlight.

3. SPECTRAL SEPARATION OF CORYMBIA CALOPHYLLA FLOWERS

Previous studies have used a range of methods to assess correlation between spectral data and vegetation species and/or assemblages. Often a staged approach is used to progressively discriminate hyperspectral bands using progressively more complex algorithms and this has often been successful. The primary reason for this approach is that by using the simpler but less quantitative algorithm first significantly reduces the dimensionality of the dataset, reducing the volume of data required for the more complex and time-consuming, but more quantitative, algorithms (Vaiphasa, Ongsomwang et al. 2005).

A review of eight different studies that included band selection from spectroscopy data showed that the most commonly used approach is an analysis of variance (ANOVA) analysis to reduce dimensionality, followed by calculation of the Jeffries-Matusita (JM) Distance for the bands with the highest mean separation from the ANOVA results. This approach has been used for applications such as plant species discrimination from leaf reflectance for broadleaf trees, papyrus, mangroves and grasses, as well as fungal effects on soybeans (see Table 1 for a summary of the methods used). As the expected spectral separation between flowers and leaves was expected to be greater than many of these applications, particularly spectral separation between leaves of different species (Gross and Heumann, 2014), a more complex approach to quantify subtle spectral differences was not warranted.

REFERENCE	SPECTRAL RANGE (nm)	SPECTRA	ANOVA	CART	JM	t-TEST	MESMA	PLSR	PCA	OTHER
(Vaiphasa et al., 2005)	350 – 2151	480	✓		✓					
(Adam and Mutanga, 2009)	350 – 2500	459	✓	✓	✓					
(Pu, 2009)	350 – 2500	394	✓							ANN[1] LDA[2]
(Kisevic et al., 2011)	350 – 1050	Unknown				✓	✓			
(Ullah et al., 2012)	2500 – 14000	455	✓		✓					
(Magiera et al., 2013)	325 – 1075	240						✓	✓	
(Bao et al., 2017)	350 – 1050	153	✓		✓					
(Al-Ahmadi et al., 2018)	480 – 2400	160	✓		✓					

Table 1 - Summary of successful spectral separation methods from literature review

[1] Artificial Neural Network
[2] Linear Discriminant Analysis

For this staged approach to assess spectral separation initially, the spectra were divided into three groups: Flowers (F), Leaves (L) and Ground (G). The range and mean of these groups are shown in Figure 3.

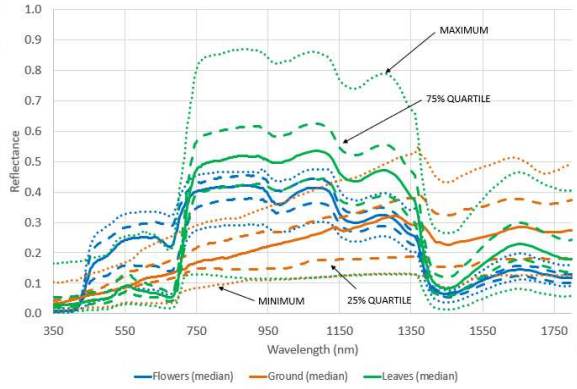


Figure 3. Range and median of spectra for the three groups

The first stage of separation analysis, an ANOVA assessment for dimensionality reduction, was performed on the three groups of spectra using a custom Python script, testing the null hypothesis that there was no statistically significant difference between the means of each pair of groups (results in Figure 3). Based on a ρ-value of less than 0.05 to disprove this hypothesis, reflectances were significantly statistically different for:

- 94% of the Flower vs Ground wavelengths
- 98% of the Flower vs Leaf wavelengths
- 92% of the Flower vs Ground AND Leaf wavelengths

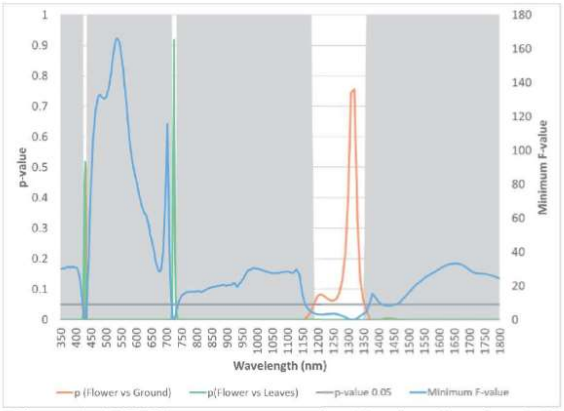


Figure 4. ANOVA assessment results. Numbers less than 0.05 represent a statistically significant separation. Areas shaded in grey show where the ρ-value is less than 0.05 between flowers and both ground and leaves. Blue line is the F-value.

To further reduce spectral dimensionality and assess the ability of different groups of bands to reliably differentiate between flower and other objects, the Jeffries-Matusita (JM) Distance was calculated for a series of bands (Schmidt and Skidmore, 2003). This started with all bands with a ρ-value greater than 0.05 and then using groups of progressively higher mean separation (ANOVA F-value – the blue line in Figure 4). After being calculated for all bands with the JM Distance calculation on all data with ρ-value < 0.05, the JM Distance was then calculated on the highest 50% of the F-values, highest 25% F-values, etc until two bands were remaining (0.5% F-values).

The progressive increase in mean separation was done to evaluate the effectiveness of higher or lower data dimensionality versus higher or lower data quality (i.e. are fewer bands of higher separation better than more bands (and therefore data points) of lower separation). As Figure 4 shows, the mean separation is generally higher in the visible bands, particularly in the region of 500 nm to 550 nm.

Graphs of JM distance for each class pair (flower-ground and flower-leaves) are shown in Figure 5. There is a general increase in JM Distance (and therefore spectral separation) with restriction of the spectral bands to those associated with increasing F-value from the ANOVA analysis.

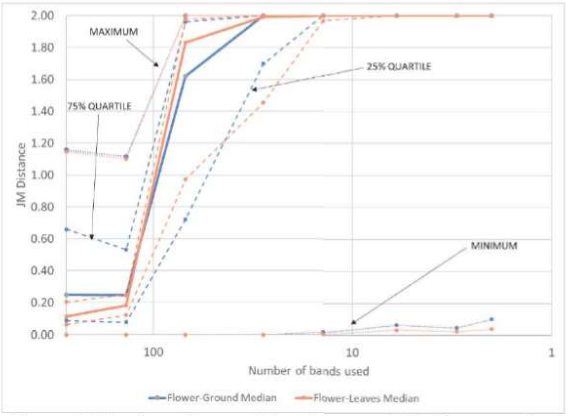


Figure 5. Number of spectral bands versus JM Distance values

Using the median values as a reliable estimate of separability and a filter of median JM Distance > 1.9 for accurate classification (Vaiphasa, Ongsomwang et al. 2005), the top 10% of spectral bands from the ANOVA classification should be able to achieve reliable classification results between flowers and other objects (28 bands of wavelengths 476 – 566 nm). Using the top 14 of the spectral bands, or less, means that over 75% of the data points are clearly separated (wavelengths 518 – 557 nm).

Note that this finding only applies to the detection of flowers versus leaves or ground in Red Gum forests for spectral data of equivalent spectral and spatial resolution as this dataset.

4. MULTISPECTRAL SEPARATION ASSESSMENT

While a proven spectral discrimination ability from hyperspectral data is a useful finding, the ability to detect flowering plants typically relies on the temporal variability of the area of interest to track phenological changes (Blomstedt, 2014). Given the cost of acquiring hyperspectral images, acquiring repeat hyperspectral datasets at regular intervals over a season is unlikely to be a practical solution.

As a result, the raw spectra acquired with the FieldSpecPro were convolved with the spectral response functions of a range of different multispectral satellite-borne sensors to produce synthetic pixels of the hypothetical measured reflectance of flowers, leaves and ground classes. The convolution was applied to each of the field-measured spectra, creating 325 measurements for each band of each satellite sensor. Graphs of the spectral bands for each sensor overlain on the median spectra for leaves, flowers and ground are provided in Figure 6. Note that, as per the previous section of this paper, the peak

separation is between 518 – 557 nm from the JM Distances calculated from the original spectroradiometer data.

The JM Distances of the synthetic satellite reflectances were calculated, using progressively greater mean separation to select the bands (similar to the assessment of the original spectroradiometer JM distances from the ANOVA results).

The results from this process are shown in Table 2. With the highest separation from the ASD spectroradiometer data being achieved across a bandwidth of 39 nm, this result correlates with Figure 4 and Figure 5, as while the majority of the wavelengths between 350 nm to 1800 nm are statistically different based on ANOVA analysis, only a small portion of the wavelengths are able to reliably separate the target classes. Satellites with too broad a bandwidth are not able to adequately resolve this key spectral zone, as are satellites with higher spectral resolution but not attuned to the 518 – 557 nm wavelengths.

Table 2 shows that only the MODIS sensor has a median JM distance greater than 1.9 and thus has sufficient spectral resolution to reliably distinguish between the flower, leaves and ground classes in the spectral space. Two bands from MODIS achieved this degree of separation; Band 4 (538 – 568 nm) and Band 11 (519 – 540 nm). Both of these bands are within the portion of the spectrum that achieved a median JM Distance of 2 from the ASD spectroradiometer data (476 – 566 nm).

Satellite	Spectral resolution range (min and max bandwidth)	Highest median JM Distance
AVHRR	60 – 1,000 nm	0.17
Sentinel 2	15 – 180 nm	0.37
Landsat 7	60 – 2,100 nm	0.47
CBERS	60 – 2,100 nm	0.49
Landsat 8	20 – 1,100 nm	0.81
ASTER	40 – 700 nm	1.11
Sentinel 3	2.5 – 40 nm	1.56
VIIRS	15 – 600 nm	1.82
MODIS	10 – 1,000 nm	2.00

Table 2. Median JM Distance results for common multispectral satellite sensors

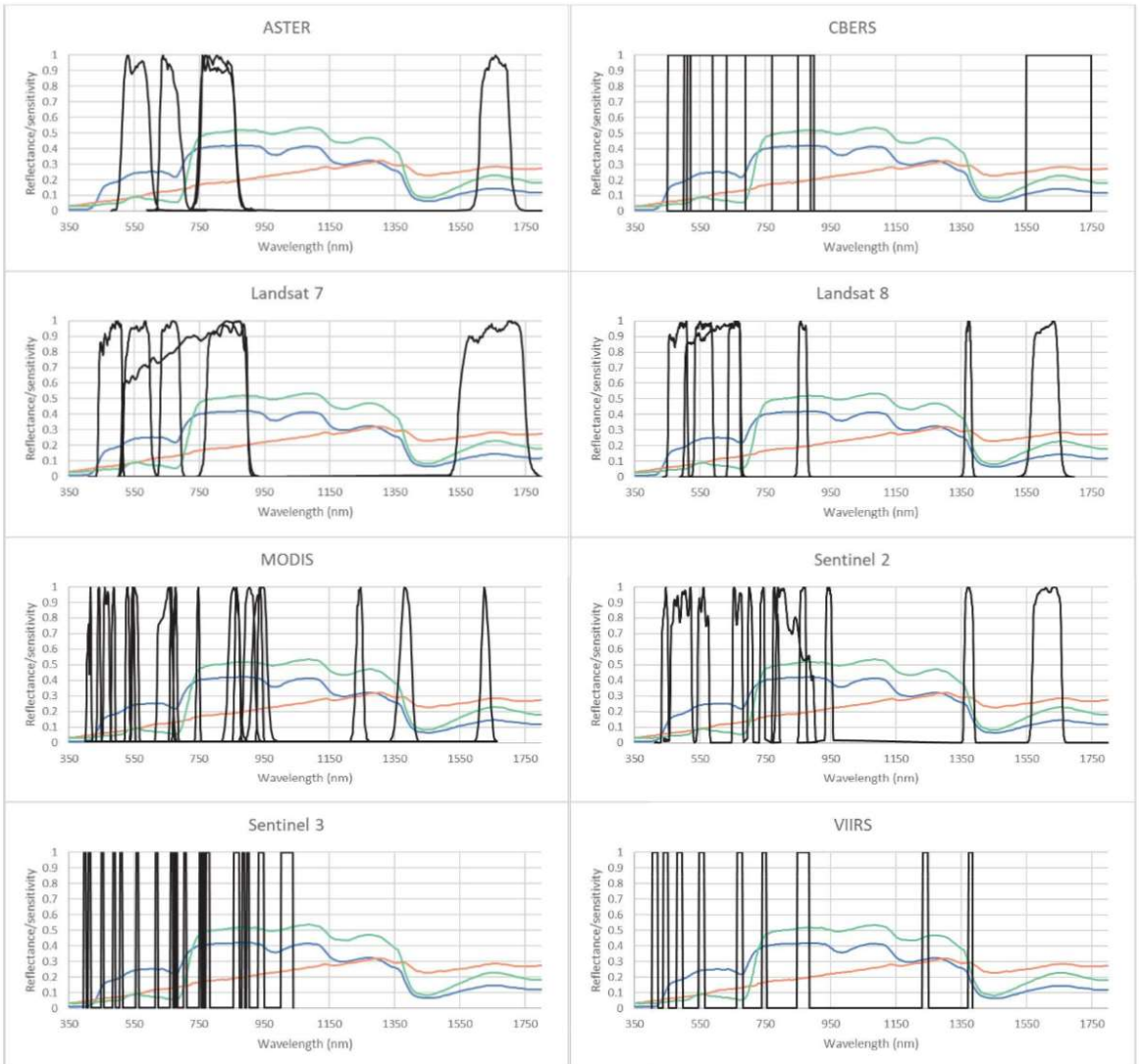


Figure 6. Spectral bands for satellite sensors overlain on median reflectance spectra from ASD spectroradiometer measurements.

5. MINIMUM FLOWER COVERAGE FOR DETECTION

While the JM Distance analysis has indicated that the MODIS sensor is capable of spectrally separating Red Gum flowers from background reflectance, the very high spatial resolution of the data in this analysis (field of view typically less than 50 cm^2) means that the flowers made up almost the entire field of view of the spectroradiometer. With the optimum MODIS bands having a spatial resolution of 500 m (Band 4) or 1,000 m (Band 10), it is considered improbable that a single MODIS pixel would consist entirely of Red Gum flowers.

In order to estimate the percentage flower cover required for flowers to be detected, the percentage of flower coverage was calculated to increase the reflectance by 1 standard deviation (SD) for backgrounds ranging from 0 – 100% leaf cover. This was done for the MODIS bands with the highest JM Distance (bands 3, 4, 10, and 12), as well as for several derived spectral products. These derived products included calculated NDVI and EVI products, as well as other combinations of MODIS visible spectral bands to determine the effectiveness of a 'visual band intensity'metric. This is in a similar vein to the work completed by Sulik and Long (2016), who developed a Normalised Difference Yellowness Index (NDYI) to better predict canola yields based on the coverage of yellow canola flowers as it was found that NDVI decreased as flower cover decreased.

The results from this process are presented graphically in Figure 7 for the MODIS bands and Figure 8 for the MODIS derived indices. The minimum, maximum and mean flower coverages required for 1 SD variation for all background scenarios are provided in Table 3.

For a minimum detection limit cutoff of 1 standard deviation (i.e. the intensity of the pixel would be at least 1 SD different to the background median value), the percent flower coverage limit is between 20 and 30% for the visible MODIS spectral bands tested here, that all had a median JM distance of greater than 1.9 between flowers and other classes. The best performing band was Band 4, with a 24.9% mean flower coverage.

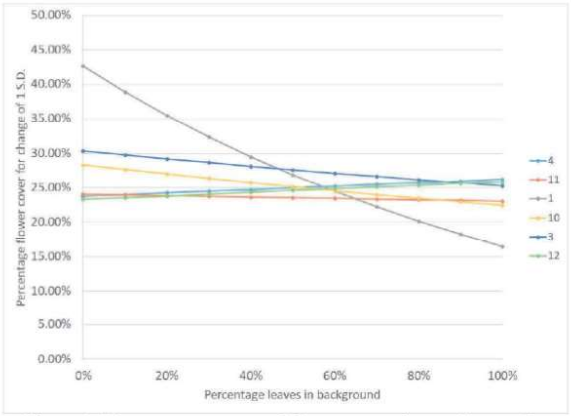


Figure 7. Percentage cover of flowers require to change the reflectance by 1 standard deviation for different background ratios of ground and leaves for MODIS bands

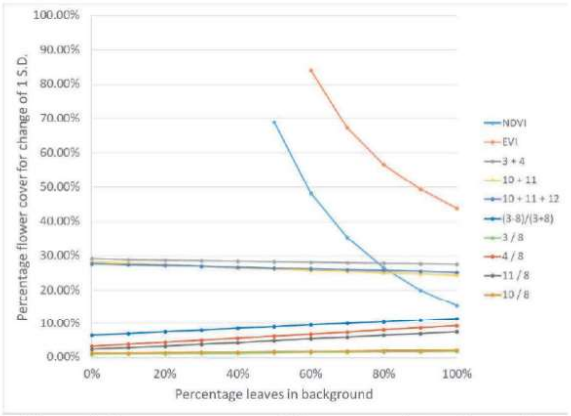


Figure 8. Percentage cover of flowers require to change the reflectance by 1 standard deviation for different background ratios of ground and leaves for MODIS derived indices

Band/Indices	1	3	4	10	11	12	NDVI	EVI	3+4	10+11	10+11 +12	(3-8) (3+8)	3/8	4/8	11/8	10/8
Minimum	16.4%	22.5%	23.7%	22.5%	23.5%	23.3%	15.3%	44.0%	27.5%	24.4%	25.3%	6.5%	1.0%	3.3%	2.4%	1.2%
Mean	27.9%	27.4%	24.9%	25.2%	27.5%	24.6%	35.8%	60.3%	28.3%	26.3%	26.5%	9.0%	1.4%	6.3%	4.9%	1.7%
Maximum	42.6%	30.4%	26.2%	28.3%	24.5%	25.9%	69.0%	84.0%	29.1%	28.2%	27.8%	11.5%	1.9%	9.3%	7.5%	2.1%

Table 3. Minimum, maximum and mean flower coverage required for 1 SD change in reflectance for MODIS synthetic pixels

The vegetation indices required a minimum of 50 - 60% leaf cover for the background for flowers to create a 1 SD variation (NDVI required as much as 69% flower cover at 50% background leaf cover and EVI required 84% flower cover at 60% leaf background cover). These minimum coverage percentages decrease with increasing flower coverage, which is not surprising due to the larger decrease in chlorophyll content with increasing flower coverage for higher leaf coverage. This correlates with the research by Sulik and Long (2015) who found the same result in canola, with NDVI decreasing as the yellow flowers became a dominant influence on the spectra.

The best results are from indices created by dividing the visible band data with the highest JM Distance (bands 3 or 4 in the 500 m spatial resolution data and bands 10 or 11 in the 1,000 m data) by Band 8 (ultraviolet). As flowers have predominantly higher reflectance in the visible bands and lower reflectance in the UV and near-UV bands (see detail of spectra in Figure 9), this process highlights both differences.

As a result, these calculated indices reach the 1 SD criteria at a mean of 1.4% flower coverage for Band 3 (500 m resolution data) and 1.7% for Band 10 (1,000 m resolution data). The response was also consistent regardless of leaf versus ground present in the background for the synthetic pixel, with Band 3

requiring no more than 1.9% flower cover and Band 10 no more than 2.1%. This is a significant improvement over using individual visible band data, visible band combinations or vegetation indices for flower detection.

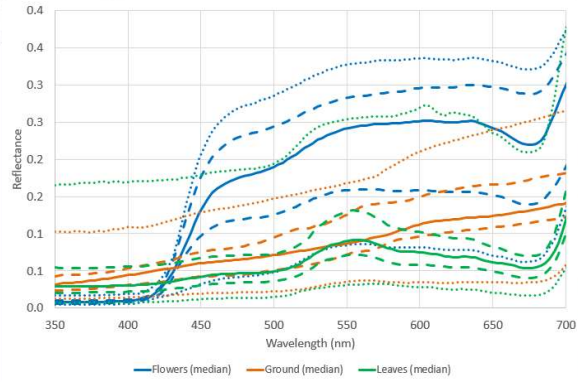


Figure 9. Flower, leaf and ground spectra from 350 - 700 nm

A similar process of assessing flower coverage detection was performed on synthetic datasets for other satellite-borne sensors (Figure 10 and Table 4). The results for individual sensor bands correlated with the results of the JM distance analysis, showing that the bands most sensitive to flower detection for all sensors are less sensitive than the MODIS bands with higher JM distances. This correlates with the JM distance analysis of the synthetic data summarised in Table 2.

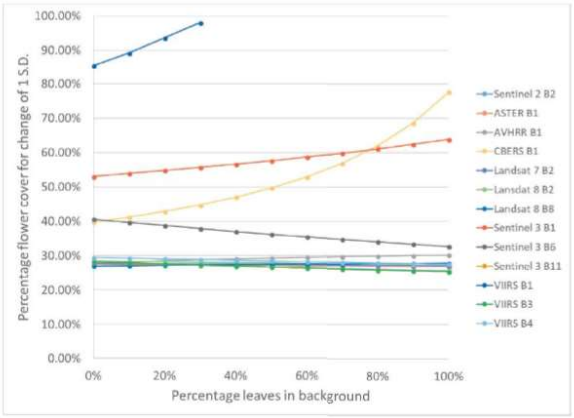


Figure 10. Percentage cover of flowers require to change the reflectance by 1 standard deviation for different background ratios of ground and leaves for other sensors and satellites

Where satellites had UV and near-UV bands (Band 1 for both Sentinel 3 and VIIRS), the same indices were calculated as were done for the MODIS data (visible band divided by the UV or near-UV band). The results of this, shown in Figure 11 and Table 4, show greater sensitivity to flower detection at low flower coverages than visual band data alone, producing detection limits more sensitive to flower coverage than when the same process was applied to MODIS data. The highest sensitivity was from Sentinel 3 (Band 6 / Band 1), with a mean flower coverage of just 1.1% required to make a difference of 1 SD to the synthetic pixel reflectance.

The improvement relative to MODIS is likely due to the specific bandwidth of the UV to near-UV bands, for example Band 8 in MODIS ranges from 405 nm to 420 nm, while Sentinel 3 is from 350 nm to 405 nm. These slightly different bands mean that the Sentinel 3 band is located entirely in the low reflectance spectral region for flowers (< 410 nm) whereas the MODIS band includes a portion of the spectra where the flower reflectance is starting to increase (see Figure 9). As a result, the UV band reflectance of flowers from Sentinel 3 and VIIRS is generally lower than for MODIS, thus having a larger impact on the band ratio.

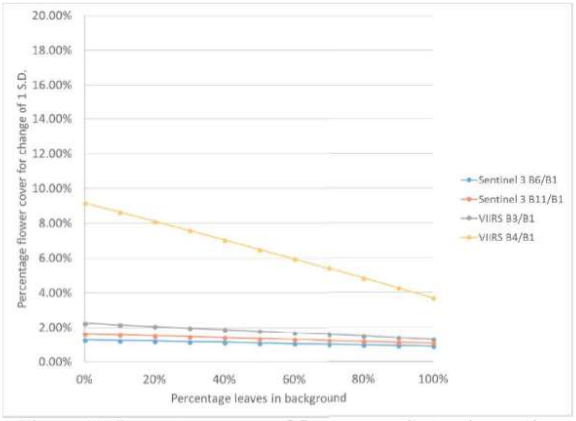


Figure 11. Percentage cover of flowers require to change the reflectance by 1 standard deviation for different ratios of ground and leaves for Sentinel 3 and VIIRS derived indices

6. MODIS FLOWER DETECTION TEST USING HONEY HARVEST DATA

To assess the ability of satellite data to detect flowers from real-world rather than synthetic data, the flower coverage as inferred from MODIS data was compared to qualitative honey harvest data from 2003 – 2017 for an apiary site near Mundaring (near Perth, Western Australia). This site was also one of the locations where the spectroradiometer data were collected (Figure 2). The vegetation consists of Red Gum forest for more than 4 km in each direction from the apiary site.

Honey harvest data were in the form of the times when hives were placed and removed from the site (corresponding to the main flowering event for the season) and a rating of the year by the apiarist as poor/failed, moderate or good. While not a detailed quantitative dataset, it does give a reliable indication of honey flow. The harvest data are summarised in Table 5, with the week of the month that the hives were placed on and removed from the site shown by the width of the bar for the year.

It was suggested by the apiarist that poor to failed harvests were due to either hot, dry or wet summers. This claim is substantiated by a comparison of the harvest data in Table 5 with mean maximum temperature and rainfall over January and February from the Bureau of Meteorology, which are shown in Figure 12. The data were from the closest meteorological station to the apiary site, which were Mundaring Weir for rainfall (6.7 km north) and Bickley for temperature (12.5 km west).

The good to moderate years are clearly bounded by years of less than 40mm but more than 0 mm of rainfall and a mean maximum temperature of less than 32 degrees Celsius. There is only one failed year within these bounds (0.6 mm rainfall, mean maximum temperature 31.4 degrees Celsius).

Band/Indices	Sentinel 2 (B2)	ASTER (B1)	AVHRR (B1)	CBERS (B1)	Landsat 7 (B2)	Landsat 8 (B2)	Landsat 8 (B8)	Sentinel 3 (B1)	Sentinel 3 (B6)	Sentinel 3 (B11)	VIIRS (B1)	VIIRS (B3)	VIIRS (B4)	Sentinel 3 (B6/B1)	Sentinel 3 (B11/B1)	VIIRS (B3/B1)	VIIRS (B4/B1)
Minimum	25.7%	26.8%	27.9%	39.8%	27.1%	27.4%	27.0%	53.0%	32.7%	25.4%	85.5%	25.4%	27.6%	0.9%	1.1%	1.3%	3.7%
Mean	26.8%	27.2%	29.2%	53.0%	27.5%	27.9%	27.5%	57.9%	36.4%	26.7%	91.6%	26.8%	28.5%	1.1%	1.3%	1.8%	6.5%
Maximum	28.0%	27.6%	30.3%	77.6%	27.8%	28.5%	27.9%	63.8%	40.6%	28.2%	98.0%	28.3%	29.5%	1.3%	1.6%	2.3%	9.2%

Table 4. Minimum, maximum and mean flower coverage required for 1 SD change in reflectance for non-MODIS synthetic pixels

	JANUARY				FEBRUARY				MARCH			
Year	1	2	3	4	1	2	3	4	1	2	3	4
2003					MODERATE	MODERATE	MODERATE	MODERATE				
2004	FAILED FLOW											
2005			MODERATE	MODERATE	MODERATE	MODERATE	MODERATE	MODERATE	MODERATE	MODERATE		
2006	FAILED FLOW											
2007			GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD		
2008	FAILED FLOW											
2009				MODERATE	MODERATE	MODERATE	MODERATE	MODERATE	MODERATE	MODERATE		
2010					POOR	POOR	POOR	POOR	POOR			
2011	FAILED FLOW											
2012	NO HARVEST FROM SITE - HIVES KEPT SOUTH OF PERTH											
2013				GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD		
2014	FAILED FLOW											
2015			GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD		
2016	FAILED FLOW											
2017	FAILED FLOW											

Table 5. Honey harvest data from Mundaring, Western Australia

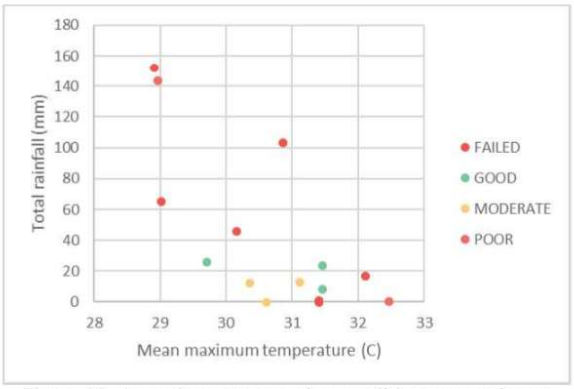


Figure 12. Annual summer weather conditions versus honey harvest quality

MODIS data were chosen for the correlation with honey harvest data as it has a longer temporal range of Sentinel 3 and VIIRS (since 2002 for MODIS, since 2016 for Sentinel 3 and since 2011 for VIIRS) and better temporal resolution than VIIRS (1 - 2 days for MODIS versus 16 days for VIIRS).

Daily MODIS data were processed for a 3x3 pixel region at the apiary site (9 pixels, or a 3 x 3 km area). The median value of the most sensitive metric of consistent pixel size (Band 10/Band 8) was calculated for February each year. For 9 pixels and 15 years of data, this created 135 datapoints for comparison. Figure 13 shows the annual MODIS average band ratio from 2003 to 2017 with the honey harvest data represented by colours indicating the poor/failed, moderate and good years.

Using the ANOVA algorithm on the null hypothesis of there is no difference in the means between the good, moderate and poor years, it was found that there was a statistically significant difference between good vs poor and moderate vs poor years ($\rho < 0.0002$ in both cases). However, there is not a statistically significant difference between the means of good vs moderate years ($\rho = 0.0657$). The data ranges are shown in Figure 14, where the difference in the medians of the groups is clear.

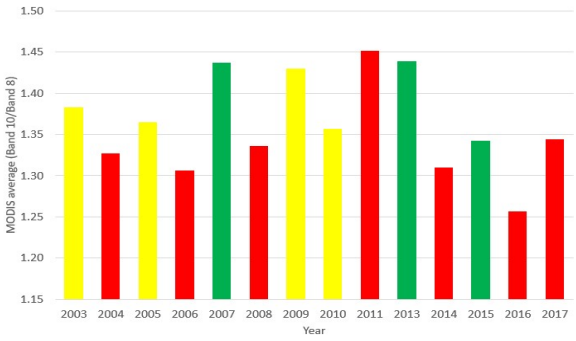


Figure 13. Median February MODIS and honey harvest data

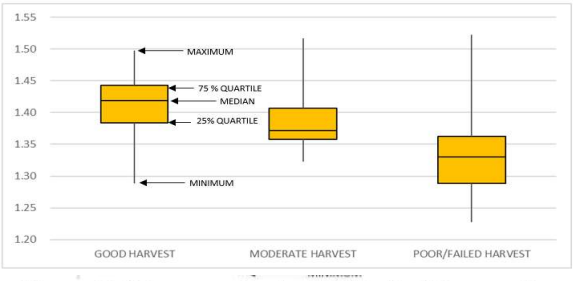


Figure 14. February median band 10/8 ratio data ranges for good to poor years

The ability of the MODIS data to generate a classification tool was also tested, by considering minimum cutoff values for good and moderate vs poor years and calculating how accurately these cutoffs separated the data into the correct classifications. The results show that a cutoff for the band 10/8 ratio of between 1.34 and 1.35 gives the best result, with a classification accuracy of 78%. A lower cutoff than this results in classification accuracy of the poor years decreasing. Above this cutoff, the accuracy of prediction of the good and moderate years decreases. While limited in spatial extent, this analysis does show that the quality of honey flow can be assessed using MODIS data and classified on a site-specific basis.

7. CONCLUSIONS

Based on an ANOVA analysis of field spectroradiometer data we can conclude that Red Gum flower reflectance spectra are significantly statistically different from the spectra of the surrounding environment for most wavelengths between 350 nm and 1800 nm. The JM Distance assessment suggests flowers are separable in hyperspectral data for 10% of wavelengths. The highest degree of separation is between 518 – 557 nm.

However, this difference does not translate well to bands of most satellites. Using the JM Distance assessment on synthetic satellite data generated from the field spectroradiometer data, only the MODIS data achieved a JM Distance of greater than 1.9 (using bands 3, 4, 10 or 11).

Further analysis showed that this spectral separation could be improved by dividing these bands by Band 8, which is on the blue/ultraviolet boundary and where flowers have a lower reflectance than their surrounding environment. A simple sensitivity assessment showed that as little as 1.0% flower coverage in a pixel may increase the band 10/8 ratio value by more than 1 SD of the background.

The Sentinel 3 and VIIRS synthetic satellite data also perform well with this approach, as their shortest wavelength bands are at slightly shorter wavelengths than the MODIS Band 8 and therefore increase the effect of dividing by this band.

The MODIS Band 10/Band 8 metric was tested for its ability to detect flowers in real-world data using 15 years of qualitative honey harvest data as a proxy for the presence of flowering Red Gum trees. This test was successful as, while there was some overlap between good, moderate and poor years, the poor years could be separated from the other years to nearly 80% accuracy.

ACKNOWLEDGEMENTS

This research is/was supported by an Australian Government Research Training Program (RTP) Scholarship.

REFERENCES

Adam, E., Mutanga, O., 2009. Spectral discrimination of papyrus vegetation (Cyperus papyrus L.) in swamp wetlands using field spectrometry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64, pp. 612-620.

Al-Ahmadi, A., Subedi, A., Wang, G., Choudhary, R., Fakhoury, A., 2018. Detection of charcoal rot (Macrophomina phaseolina) toxin effects in soybean (Glycine max) seedlings using hyperspectral spectroscopy. *Computers and Electronics in Agriculture*, 150, pp. 188 - 195.

Bao, S., Cao, C., Chen, W., Tian, H., 2017. Spectral features and separability of alpine wetland grass species. *Spectroscopy Letters*, 50, pp. 245 - 256.

Blomstedt, W., 2014. Mapping The Phenology of European Honey Bee Nectar Flows. School of Geosciences. University of Edinburgh, Scotland, United Kingdom.

Campbell, T., Fearns, P., 2018. Simple remote sensing detection of Corymbia calophylla Flowers using common 3 –band imaging sensors. *Remote Sensing Applications: Society and Environment*, 11, pp. 51-63.

Cochrane, M.A., 2000. Using vegetation reflectance variability for species level classification of hyperspectral data. *International Journal of Remote Sensing Applications*, 21, pp. 2075-2087.

Dallon, D., 2003. Comparison of the Analytical Spectral Devices FieldSpec Pro JR and the Apogee/StellarNet Model SPEC-PAR/NIR Spectroradiometers. Utah State University, Utah, USA.

Feilhauer, H., Asner, G., Martin, R., Schmidtlein, S., 2010. Brightness-normalized Partial Least Squares Regression for hyperspectral data. *Journal of Quantitative Spectroscopy & Radiative Transfer*, 111, pp. 1947-1957.

Gross, J., Heumann, B., 2014. Can flowers provide better spectral discrimination between herbaceous wetland species than leaves? *Remote Sensing Letters*, 5, pp. 892-901.

Irish, J., Blair, S., Carter, D., 2011. The Antibacterial Activity of Honey Derived from Australian Flora. *PLoS ONE*, 6.

Kisevic, M., Smailbegovic, A., Gray, K., Andricevic, R., Craft, J., Petrov, V., Brajcic, D., Dragicevic, I., 2011. Spectral reflectance profile of Caulerpa racemosa var. cylindracea and Caulerpa taxifola in the Adriatic Sea. *Acta Adriatica*, 52, pp. 21-28.

Magiera, A., Feilhauer, H., Otte, A., Waldhardt, R., Simmering, D., 2013. Relating canopy reflectance to the vegetation composition of mountainous grasslands in the Greater Caucasus. *Agriculture, Ecosystems and Environment*, 177, pp. 101-112.

Painter, S., 2010. Jarrah honey crisis as yield wiped out, The West Australian, Perth, Australia.

Price, J., 1994. How unique are spectral signatures? *Remote Sensing of Environment*, 49, pp. 181-185.

Pu, R., 2009. Broadleaf species recognition with in situ hyperspectral data. *International Journal of Remote Sensing*, 30, pp. 2759-2779.

Schmidt, K.S., Skidmore, A.K., 2003. Spectral discrimination of vegetation types in a coastal wetland. *Remote Sensing of Environment*, 85, pp. 92-108.

Shrestha, M., Dyer, A., Boyd-Gerny, S., Wong, B., Burd, M., 2013. Shades of red: Bird-pollinated flowers target the specific colour discrimination abilities of avian vision. *New Phytologist*, 198, pp. 301-310.

Sulik, J., Long, D., 2015. Spectral indices for yellow canola flowers. *International Journal of Remote Sensing*, 36, pp. 2751-2765.

Sulik, J., Long, D., 2016. Spectral considerations for modeling yield of canola. *Remote Sensing of Environment*, 184, pp. 161 - 174.

Ullah, S., Schlerf, M., Skidmore, A., Hecker, C., 2012. Identifying plant species using mid-wave infrared (2.5–6 μm) and thermal infrared (8–14 μm) emissivity spectra. *Remote Sensing of Environment*, 118, pp. 95 - 102.

Vaiphasa, C., Ongsomwang, S., Vaiphasa, T., Skidmore, A.K., 2005. Tropical mangrove species discrimination using hyperspectral data: A laboratory study. *Estaurine Coastal and Shelf Science*, 65, pp. 371-379.

5 Prediction and detection of honey harvests from remote sensing and weather data

5.1 Abstract

There have been many efforts to use remotely sensed data to map or predict the production of honey. Several studies recommend that weather data are incorporated into the model. Here we assess the ability of satellite and weather data to predict the volume of honey produced from *Corymbia calophylla* (Myrtaceae) in southwest Australia.

Utilising honey harvest data over 8 years, it was found that January NDVI could predict a 'good' harvest (more than 40 kg of honey harvested per hive) to 79% accuracy. Poor harvests (less than 20 kg of honey) and moderate harvests (between 20 and 40 kg of honey) were not distinguishable.

Assessing weather for January and February showed that the weather data from January was highly influential. Good harvests generally occurred after a cool, dry January, moderate harvests after a warmer, wetter January and poor harvests associated with warmer, drier January. Using a decision-tree approach, the combination of January weather and NDVI classified good harvests to 90% accuracy. Classification into the three quality levels achieved 69% accuracy from the overlapping data for poor and moderate years.

This study used monthly weather data. Addition of daily weather data and apiary health variables may improve the predictive accuracy.

5.2 Introduction

In nectar productive regions, such as the biodiverse region of the southwest of Australia, beekeeping has expanded in recent years, with the number of registered beekeepers growing from 660 in 2010 to over 3,000 in 2019 (Thomson 2019) and significant international interest in local varietals. For example honey produced from Western Australian endemic species has some of the highest antimicrobial properties known for honey (Irish et al. 2011). This has led to the 'farm gate' price for monofloral marri honey (produced by European honey bees (*Apis mellifera*) from the *Corymbia calophylla* trees) reaching prices of A$30/kg in 2018, compared to a decade earlier at of A$ 3/kg (Hastie 2018).

While annual honey yields from *Corymbia calophylla* forests can be exceptionally high, being over 70 kg per hive in some cases, it can also 'fail' in some seasons, with declines in

marri honey produced due to depleted flowering events (Painter 2010). In addition to the financial loss to the beekeeper, these failed years often result in bees starving and hive loss. Thus developing tools that can accurately predict flowering events in major nectar producing native species has major commercial benefits.

Developing predictive modelling of honey flow is a major area of interest with high commercial value to apiarists worldwide. Since the advent of remote sensing, there have been efforts to use remotely sensed data to map or predict the timing and volume of honey production for different regions for predictive hive placement.

For example, a European study of honey yield against temporal variations in the Normalized Difference Vegetation Index (NDVI) from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor (Blomstedt 2014) showed a consistent time delay between the Start of Season (SOS) 'green-up' indicator and peak honey flow time. The NDVI SOS measurement was shown to be a reasonably consistent predictor of peak honey flow timing ($R^2 = 0.76$). However, no correlation was found between the SOS and the start of honey flow as this is more dependent on the appearance of flowers on isolated plants, whereas the peak honey flow is a result of more widespread flowering and therefore more closely tied to the MODIS sensor scale.

In Australia, eucalypts across south-eastern Australia yielded a broad, qualitative correlation between changes in vegetation indices derived from MODIS data and honey yield data, with recommendations made for more detailed studies on local variations in floristic communities and the impact of climatic variables (Webber 2011). Similar work using the 'BeeBox' web-based portal (Winter et al. 2013) showed a stronger correlation between annual peak Enhanced Vegetation Index (EVI) from MODIS and *Eucalyptus tricarpa* (ironbark) winter honey flow over a 15-year period (Arundel et al. 2016). Again, this was a qualitative correlation and no quantitative model formed.

Spectral indices for predicting canola yield have shown that NDVI and EVI actually decrease over canola (*Brassica napus*) fields at the peak of flowering, despite the increase in floral cover and biomass (Sulik and Long 2015). As a result, a 'Normalized Difference Yellowness Index' (or NDYI) was developed, which provided a better yield predictor for canola than NDVI (Sulik and Long 2016).

Higher resolution airborne hyperspectral data were shown to have a high degree of accuracy in detecting flowering in native vegetation in Africa (Landmann et al. 2015). Using

60 cm pixel hyperspectral data from 400 – 990 nm, this study was 77 – 97% accurate in detecting peak flowering depending on the plant species involved.

Spectral reflectance work by the authors has established the potential for direct detection of flowers in the peak nectar producing tree *Corymbia calophylla* (marri) using drone (Campbell and Fearns 2018b) and satellite platforms (Campbell and Fearns 2018a). The success of the approach was influenced by the combination of the strong spectral separation of flowers of the target species from leaves and groundcover of these forests (see Figure 5-1), particularly in the visible and ultra-violet (UV) spectra, combined with a lack of other large canopy dominant species that flower at the same time and the large volume of flowers of marri during the flowering season (see Figure 5-2). Using the high degree of spectral separation of marri flowers, an algorithm (Marri Flowering Index (MFI)) was developed using MODIS data (Band 10 for the visible spectrum and Band 8 for the near-UV) that separated poor honey yield years from moderate to good years with an 80% accuracy for a trial location.

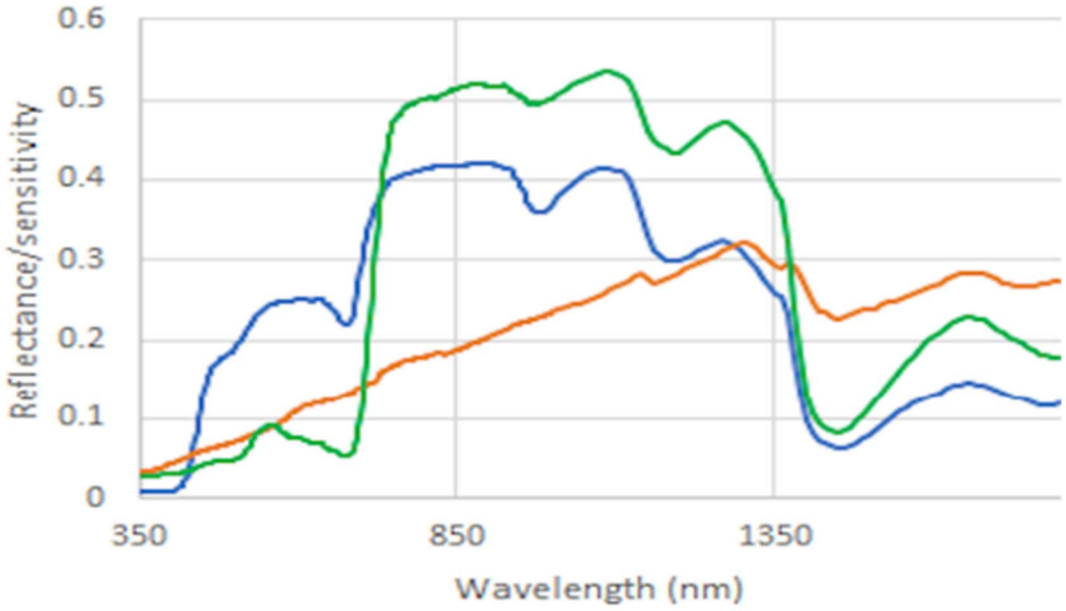


Figure 5-1: Median reflectance spectra collected from Corymbia calophylla *forests using a handheld spectroradiometer. Blue represents* Corymbia calophylla *flowers, green represents leaves from* Corymbia calophylla *and other species present, orange represents ground spectra. For a more complete discussion of the data, refer to Campbell and Fearns (2018a). The full spectral data are available as a published dataset (Campbell 2019).*

Figure 5-2: Corymbia calophylla *(marri) forest in bloom.*

The objective of this study was to test the hypothesis that satellite-borne remotely sensed data supplemented by the inclusion of local weather conditions has the ability to accurately predict marri flowering events across a range of vegetation communities and climatic regions for the southwest of Australia, a source of high quality mono-floral varietal honeys (Irish et al. 2011). Marri trees typically flower over late summer (February is generally the prime honey producing month (Herbarium 1998)) and an interrogation of the local Western Australian botanical database (Herbarium 1998) shows that there are rarely other canopy species flowering at this time (Campbell and Fearns 2018a). Developing such a predictive model of honey production provides an invaluable tool for apiarists to plan hive placement to maximise yield and apiary profitability.

5.3 Materials and methods

5.3.1 Honey harvest and other site data

Honey harvest data were collected from across the south-west of Australia (see Figure 5-3), with over 350 km between the northern and southern sites. Most sites are clustered centrally, within 65 km of the Western Australian state capital city of Perth. Data were from two apiarists; one 'commercial' apiarist with approximately 700 hives and one 'hobby' apiarist with approximately 50 hives. Both apiarists use the Italian subspecies of *Apis mellifera*, with queen bees bred from the 'Better Bee' queen breeding program (based on Western Australia's Rottnest Island). The 'northern apiary sites' (i.e. those more than 50 km

north of Perth) are typically used for end of season harvests by the commercial apiarist as the marri trees typically finish flowering later in the season further north (Leyland 2015).

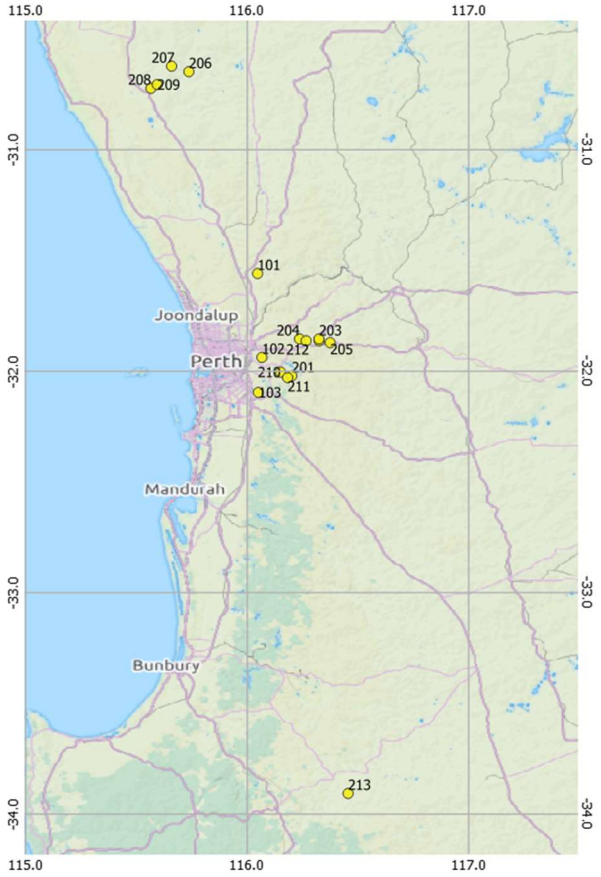


Figure 5-3: Apiary site locations indicated by yellow markers. Site labels correspond to those listed in Table 1. Coordinates are in WGS84.

Data included the number of hives at the apiary site for the season, dates the hives were moved onto and off the apiary site and the total weight of marri honey harvested from the apiary site for the season. This resulted in data from 58 harvests between 2010 – 2018 across 16 apiary sites, with not all apiary sites being utilised each year. Other apiary variables, such as queen bee quality and hive condition or strength (for example measures of adult bee population and brood, as per Delaplane et al. (2013)) and when the marri trees started to flower, were not recorded. All apiary sites were at least three kilometres from other known apiary sites, and therefore outside the forage range for honey bees when nectar is moderately to readily available (Hagler et al. 2011).

Harvest data were reduced to the average weight of honey harvested per hive per year per apiary site. The summary of the honey harvest data presented in Table 5-1 highlights the difficulty in obtaining continuous, or at least annual, data for all sites. Not all apiary sites are utilised every year, however the absence of data should not be interpreted as implying a

low annual yield for the site. For example, sometimes a site is not utilised because better yield is expected elsewhere. The practice of apiarists locating hives at the sites expecting higher yields each year limits the number of data points representing failed honey flow due to no flowers, thus limiting the potential for verification of the "null result" in the satellite-derived data. There are four "null results" in the honey harvest data used, all of which occur from 2010 - 2012. It is important to note that in many years, the harvest data ranged from poor to good depending on the location, indicating a high spatial variability in nectar production for a given year across different locations.

Table 5-1: Summary of average marri honey harvest per hive by apiary site and year, with the number of hives at the apiary site in brackets after the weight. Red = poor harvest (< 20 kg per hive), yellow = moderate harvest (20 - 40 kg per hive) and green = good harvest (> 40 kg per hive)

Site	2010	2011	2012	2013	2014	2015	2016	2017	2018
101	*Not used*	*Not used*	*Not used*	*Not used*	*Not used*	39.3 (2)	20.3 (4)	12.5 (4)	35.0 (4)
102	*Not used*	*Not used*	*Not used*	*Not used*	*Not used*	36.3 (2)	0.0 (4)	4.0 (4)	30.0 (4)
103	*Not used*	*Not used*	*Not used*	*Not used*	*Not used*	*Not used*	6.0 (3)	10.0 (4)	28.1 (5)
201	*Not used*	*Not used*	*Not used*	49.6 (112)	*Not used*	52.2 (112)	*Not used*	*Not used*	*Not used*
202	*Not used*	*Not used*	*Not used*	65.6 (112)	*Not used*	45.5 (112)	*Not used*	*Not used*	*Not used*
204	0.0 (112)	16.1 (112)	*Not used*	71.0 (112)	*Not used*	*Not used*	*Not used*	26.8 (112)	40.2 (112)
205	16.1 (112)	0.0 (112)	*Not used*	*Not used*	*Not used*	48.2 (112)	*Not used*	13.4 (112)	38.8 (112)
206	34.8 (112)	*Not used*	0.0 (112)	18.8 (112)	*Not used*	29.5 (112)	*Not used*	*Not used*	16.1 (112)
207	37.5 (112)	*Not used*	0.0 (112)	18.8 (112)	*Not used*	29.5 (112)	*Not used*	*Not used*	16.1 (112)
208	18.0 (224)	*Not used*	8.0 (224)	16.1 (112)	16.1 (112)	12.1 (112)	*Not used*	*Not used*	16.1 (112)
209	25.4 (112)	*Not used*	*Not used*	18.8 (112)	*Not used*	20.1 (112)	*Not used*	*Not used*	16.1 (112)
210	34.8 (112)	*Not used*	*Not used*	56.3 (112)	*Not used*	67.0 (112)	*Not used*	*Not used*	*Not used*
211	16.1 (112)	*Not used*	*Not used*	49.6 (112)	*Not used*	52.2 (112)	*Not used*	*Not used*	*Not used*
212	*Not used*	8.9 (112)	*Not used*	42.9 (112)	*Not used*	32.1 (112)	*Not used*	11.6 (112)	26.8 (112)
213	*Not used*	*Not used*	17.1 (112)	*Not used*	*Not used*	*Not used*	*Not used*	44.2 (112)	*Not used*
Annual average	*22.8*	*8.3*	6.3	*40.8*	*16.1*	*38.6*	*8.8*	17.5	*26.3*

Previous work on correlating MODIS data with marri honey harvests by Campbell and Fearns (2018b) utilised an apiary site surrounded in all directions by several kilometres of relatively uniform, mature marri forest. It was expected that this would show the greatest change in satellite measured reflectance due to flowering intensity from flowers resulting in a larger proportion of the pixel's area with the floral signature compared with mixed land-use areas. However not all apiary sites have such consistent forest coverage.

To assess the impact of canopy cover on the vegetation indices and the MFI metric across a broader range on canopy cover densities than the previous study (i.e. including younger regrowth forest and partially cleared land rather than only mature forest), vegetation structure products produced by AusCover (Scarth 2009) were used to represent the mature

canopy cover for each site. The AusCover vegetation structure data were derived from remotely sensed Landsat data, resulting in a ground resolution of approximately 30 m. While the uncertainty for the vegetation cover has been proven to be as high as 20% (Guerschman et al. 2012), the AusCover data were considered sufficiently accurate to distinguish complete forest cover from partially cleared land or land with some remnant trees.

With flowering marri trees being from 10 – 40 m high depending on age (Brooker and Kleinig 2001), the proportions of plant cover for the 10 – 30 m and > 30 m intervals were summed (the AusCover dataset provides fraction of vegetation cover from between 0 – 5 m height, 5 – 10 m height, 10 – 30 m height and greater than 30 m height). As the typical forage range for honey bees with moderate nectar availability is 1 – 2 km (Hagler et al. 2011), the AusCover mean mature tree canopy cover data from this process were averaged to the nearest 1 km MODIS pixel extents to assist with assessing the relative impact of changes in vegetation phenology per pixel. Examples of this process are shown in Figure 5-4.

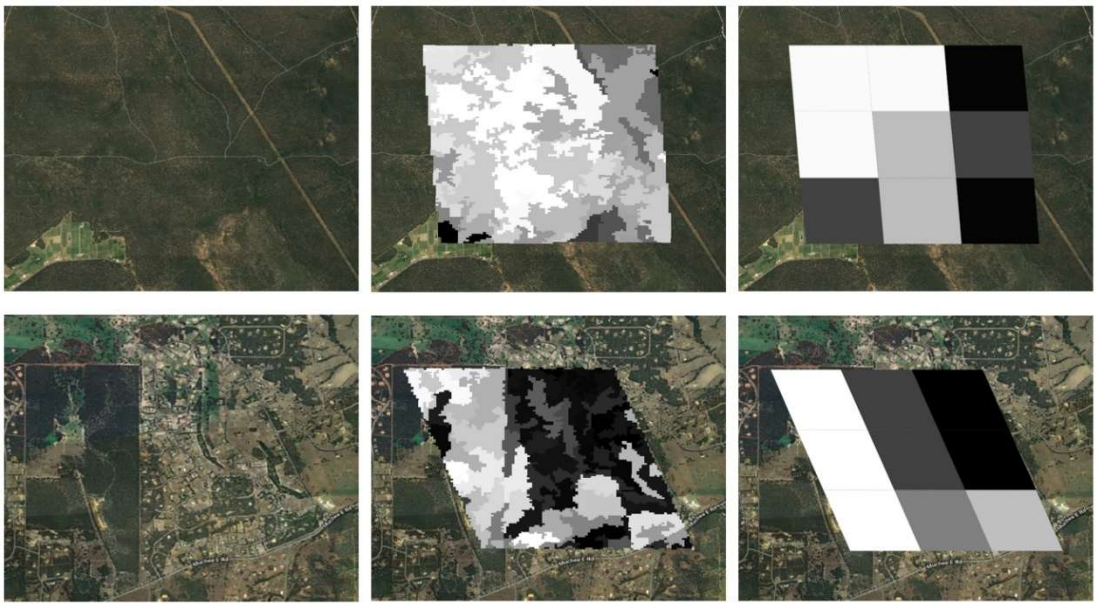


Figure 5-4: Comparison of aerial imagery (left panels) with AusCover vegetation height data at ~ 30 m spatial resolution (middle panels) and AusCover mean per (1km) MODIS pixel (right panels) for forest (top row) and mixed land use (bottom row) apiary locations.

5.3.2 MODIS derived indices and honey yields

Satellite-derived data from the MODIS sensor were used in this study in preference to other sensors with higher spatial resolution (e.g. Landsat or Sentinel-2) for the following reasons:

- The narrower spectral bands for MODIS compared with Landsat and their specific wavelengths result in a significantly higher spectral separation for marri flowers (Campbell and Fearns 2018a).
- The daily temporal resolution of MODIS, compared with the 16-day temporal resolution of Landsat, is more suited to accurately measuring maximum values over a 1-month period.
- While Sentinel 2 has a better temporal resolution than Landsat (5 days versus 16 days), this data is only available from after March 2017. This gives two years of Marri harvest data to analyse.

Previous studies correlating MODIS NDVI and EVI to honey harvests in the southeast Australian tree *Eucalyptus tricarpa* (ironbark) showed a degree of qualitative correlation between vegetation indices peaks in winter and the peak quarterly honey production for the same site (Arundel et al. 2016). However, no quantitative assessment was performed. In addition, marri trees are a late summer to early autumn flowering species (the annual low period for NDVI in Australia). It was expected that seasonal variations in vegetation indices would be larger than spectral variations due to flower coverage, as demonstrated by Campbell and Fearns (2018a), who estimated that NDVI and EVI require minimum canopy cover of 50% and 60% respectively for changes in flower coverage to be detectable. Even at 100% canopy cover, it was estimated in this previous study that these two indices still require 15% and 44% flower coverage of the pixel respectively for detection of flowers to be viable.

Over the apiary sites used, the average canopy cover was only 22%. Accordingly, three sites with the highest canopy coverage and years of honey harvest data were initially selected to test the ability of NDVI and EVI to detect marri flowers. These three sites had canopy coverages ranging from 38% to 53%. The remaining 13 sites with lower canopy coverage were then assessed.

Maximum monthly NDVI and EVI data for these sites were calculated from the MODIS 1 km pixel size database (as an indication of the highest level of canopy chlorophyll for the month (Gitelson and Merzlyak 1997)).These maximum values were extracted from the same MODIS pixels used to calculate the Marri Flowering Index (MFI), with the MFI calculated by dividing the value of MODIS Band 10 by Band 8, and as described in Campbell and Fearns (2018a).

5.3.3 Weather data

To achieve a greater understanding of the causes of the variations in honey harvest weight, average maximum temperature and monthly rainfall data for January and February (the month immediately preceding the main flowering period and the month of the main flowering period (Brooker and Kleinig 2001)) were obtained from the Australian Government's Bureau of Meteorology (BOM) website (http://www.bom.gov.au/climate/data/) for the weather stations closest to the apiary sites. Note that the distance between the source of weather data and harvest data was up to 40 km in some cases as apiary sites can be situated in remote, unpopulated locations.

These weather data were plotted both against the honey harvest weight data for each site and year, as well as plotted against each other, with points coloured by harvest quality to locate multiple weather factors that work together to affect the honey harvest quality.

5.4 Results

5.4.1 MODIS derived indices and honey yields

With the increase in flower coverage over the flowering period resulting in less photosynthetically active vegetation at the top of the canopy, it would be expected that if the vegetation indices were directly detecting the presence of flowers then the vegetation index would decrease due to vegetation being occluded by flowers (as per spectral separation analysis by Sulik and Long (2015) and Campbell and Fearns (2018a)). However, as the February NDVI data in Figure 5-5 show, there is a weak positive correlation between February maximum NDVI and marri honey harvest weight (R = 0.73).

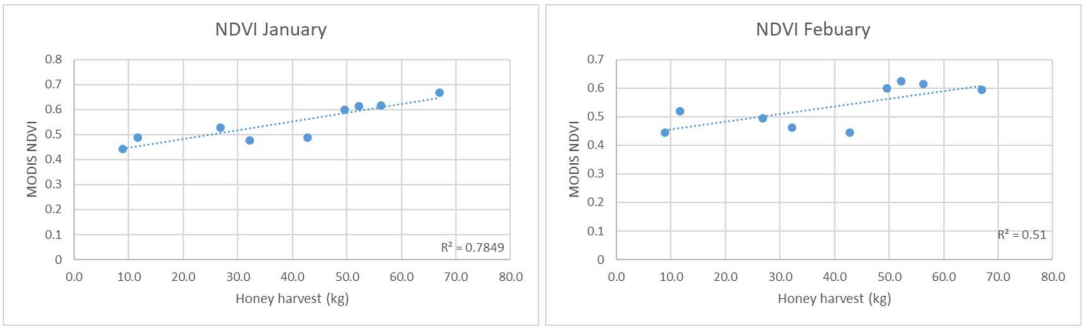


Figure 5-5: MODIS NDVI versus annual marri honey yield

The correlation between NDVI and honey harvest is stronger in January (R = 0.88) with increased chlorophyll content, as measured by the NDVI data, linked to more active green floral bud formation in the pre-flowering stage with resultant higher flowering, nectar

production and concomitant increase in honey harvest. Marri buds are held in terminal corymbs and can be visible as a lighter green 'canopy' over the tree compared with marri leaves.

The January data shows that data points can be divided into two parts based on marri honey harvests above or below 45 kg per hive for the season. Below this harvest level there is a very poor correlation between honey harvest and NDVI (R = 0.40). Above this harvest (the four largest honey harvests), there is a stronger linear correlation between honey harvest and the January maximum NDVI (R = 0.99) (Figure 5-6). It is possible that for lower nectar production years, the variation in chlorophyll content is not sufficient to be measurably different by MODIS NDVI. For years with exceptionally high nectar production, the chlorophyll rises above the detection threshold and this linear relationship is measurable.

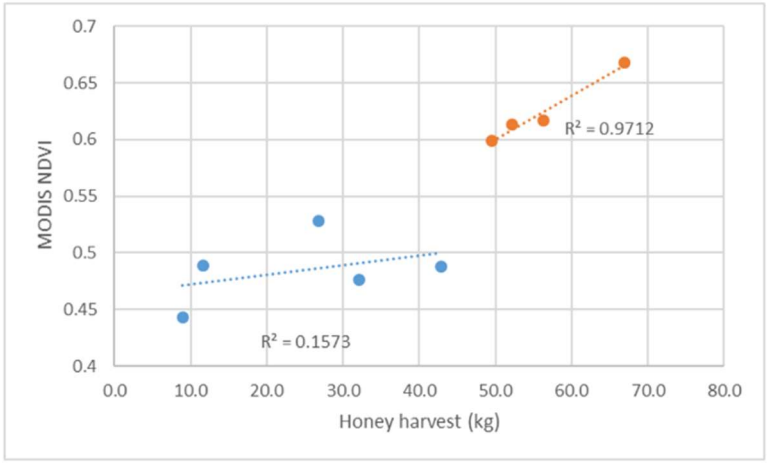


Figure 5-6: Two zones of different relationships between January NDVI and annual marri honey yield

The data analysis was then extended to include all honey harvest data points, regardless of degree of canopy cover. The result, shown in Figure 5-7, indicates a weak positive correlation between January NDVI and honey harvest (R = 0.46). The box plot of the data in Figure 5-7 does show that 'good' harvest years (honey harvest weight > 40 kg) have generally higher January NDVI. This was confirmed by an ANOVA analysis, which showed that there is no significant difference between the January MODIS NDVI data for poor (< 20 kg) and moderate harvest amounts (20 – 40 kg) whereas good harvests were separable from poor and moderate harvests (ρ<0.05 for good vs moderate and good vs poor years but not moderate vs poor years). Note that these ranges for poor, moderate and good harvest years are based on the average weight of honey harvested per 'box', or full-depth

super (the upper boxes in a hive where honey is deposited), of honey (average weight of 20 kg per 'box' (Leyland 2015)).

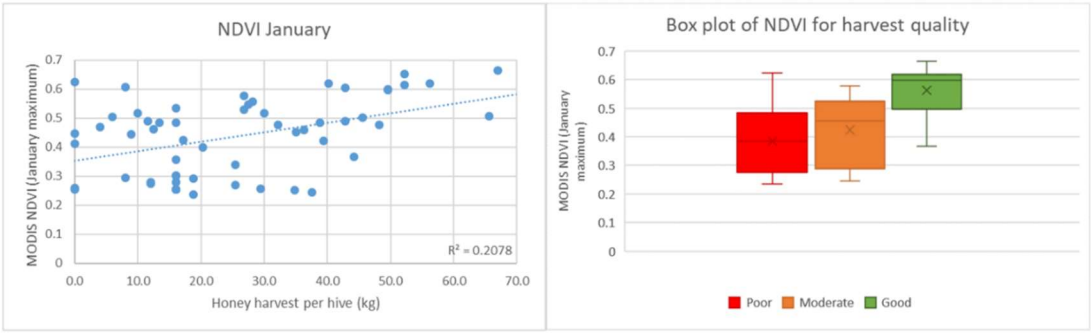


Figure 5-7: Maximum January MODIS NDVI for all honey harvest data points from Table 5-1.

To assess the effect of site or year specific effects, the data in Figure 5-7 were grouped into apiary sites and years (Figure 5-8 and Table 5-2). These groups show that there is significantly better relationship between harvest volume and NDVI by year than by site (median R^2 0.75 by year vs 0.13 by site), and that the positive relationships dominate by year as the harvest volume increases (there are no negative correlations above 45 kg per hive harvest weight). However, the majority of the relationship between NDVI and honey harvest by year is likely to be a result of increasing NDVI with increasing canopy cover rather than from the honey harvest itself (see Figure 5-9).

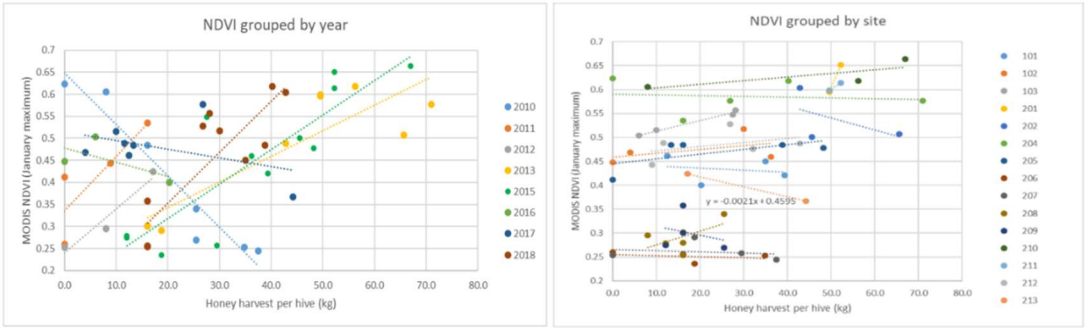


Figure 5-8: Maximum January NDVI vs honey harvest per hive, grouped into years (left frame) and apiary sites (right frame) with linear regression trendlines shown

Table 5-2: NDVI by year and site linear regression summary

YEAR	R^2	GRADIENT
2010	0.93	-0.012
2011	0.71	0.012
2012	0.94	0.010

SITE	R^2	GRADIENT
101	0.04	-0.0004
102	0.23	0.0008
103	0.98	0.0022

2013	0.78	0.006
2015	0.18	-0.002
2016	0.42	-0.003
2017	0.75	0.008
2018	0.75	0.012
MEDIAN	***0.75***	***0.007***

201	N/A	0.0208
202	0.03	-0.0025
204	0.03	-0.0002
205	0.37	0.0010
206	0.10	-0.0002
207	0.03	-0.0002
208	0.32	0.0028
209	0.07	0.0019
210	0.61	0.0008
211	N/A	0.0056
212	0.16	0.0008
213	N/A	-0.0021
MEDIAN	***0.13***	***0.0008***

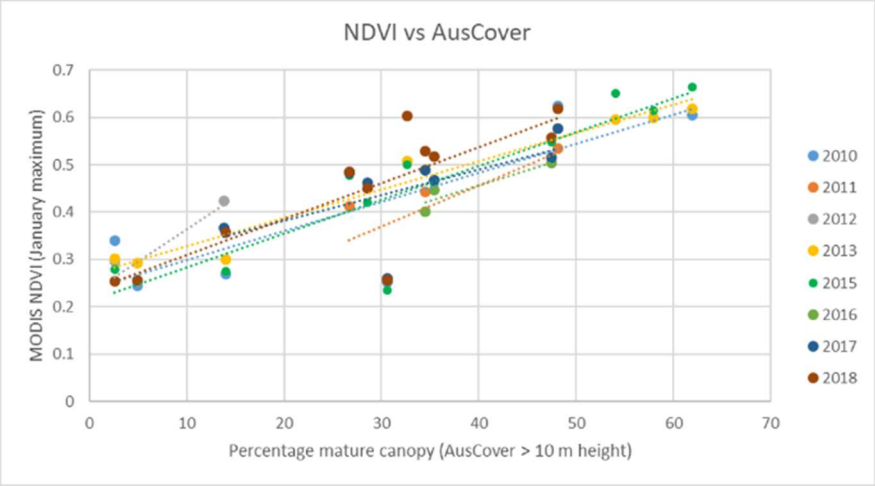


Figure 5-9: January maximum NDVI versus mature canopy cover (AusCover > 10 m height)

As MFI and honey harvest weight were reported by Campbell and Fearns (2018a) as being linked, the MFI was compared to the honey harvest data used for NDVI and EVI evaluation. A plot of MFI vs honey harvest data (Figure 5-10a) shows no correlation across the full range of harvest weights.

Given the range of canopy coverage as derived from AusCover data (2.6% – 61.9%), the MFI was multiplied by the canopy coverage in an effort to enhance the effect that canopy cover has on the index (Figure 5-10b). This scaled MFI (sMFI) showed a similar result to the

maximum NDVI in January, with low correlation for poor to moderate years (<40 kg per hive) and generally higher sMFI for good years (>40 kg per hive). This is more apparent in the box plots in Figure 5-11, with the sMFI for good years being noticeably higher. However, an ANOVA analysis of the sMFI for the three different groups of harvest weight only yielded a value of $\rho < 0.05$ for moderate vs good years (the NDVI ANOVA analysis gave this result for both poor vs good and moderate vs good years).

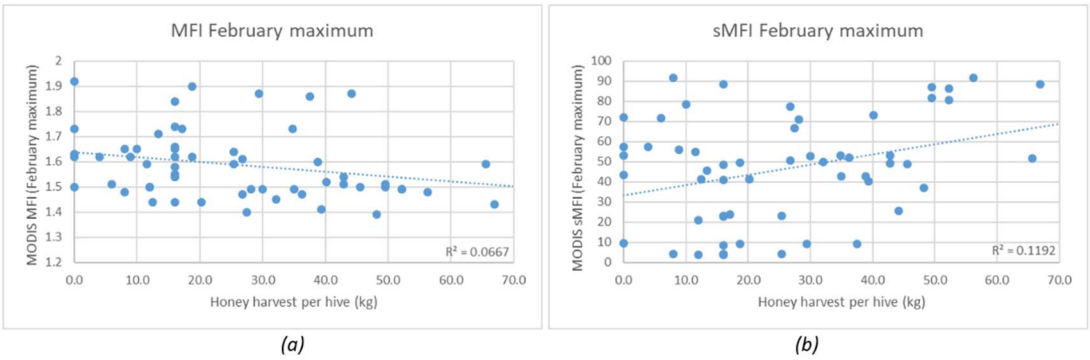


Figure 5-10: Marri Flowering Index (MFI) and scaled Marri Flowering Index (sMFI) vs honey harvest weight

Figure 5-11: Box plots of MFI and sMFI for different honey harvest weights

The assessment of linear regression for sMFI when grouped by year and site (Figure 5-12 and Table 5-3) again showed poor correlation between sMFI and honey harvest weight. Indeed, the median gradient for sMFI grouped by site in Figure 5-12 and Table 5-3 is negative, indicating that there is a negative correlation between sMFI and honey harvest weight. It may be that MFI can be effective in some cases where there is a high proportion of canopy cover (as found by Campbell and Fearns (2018a)). However the signal may too close to the noise floor for the MODIS data to be used reliably across areas of differing canopy covers, especially those areas with a lower proportion of canopy cover.

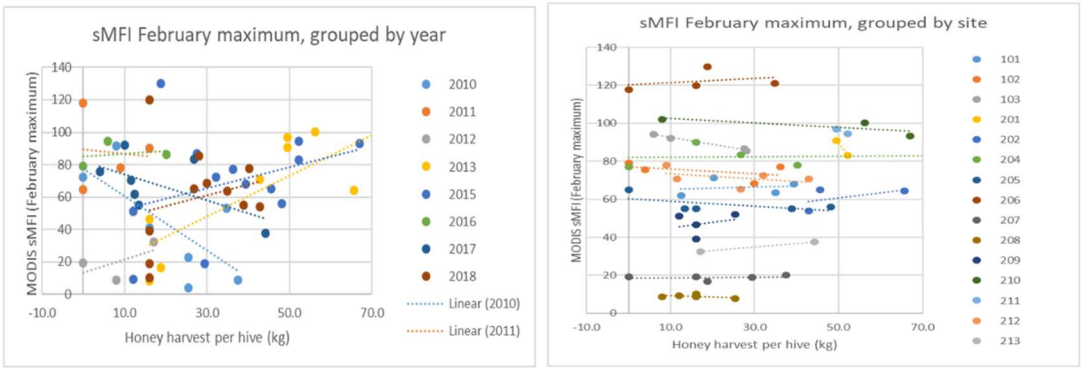


Figure 5-12: Maximum February sMFI grouped into years and apiary sites with linear regression trendlines

Table 5-3: sMFI by year and site linear regression summary

YEAR	R^2	GRADIENT
2010	0.51	1.69
2011	0.01	-0.27
2012	0.06	0.18
2013	0.62	1.26
2015	0.12	0.65
2016	0.06	0.18
2017	0.35	-0.79
2018	0.05	0.66
MEDIAN	*0.09*	*0.42*

SITE	R^2	GRADIENT
101	0.03	0.056
102	0.23	-0.124
103	0.98	-0.368
201	N/A	-2.907
202	0.28	0.267
204	0.01	0.016
205	0.38	-0.113
206	0.09	0.113
207	0.04	0.018
208	0.25	-0.074
209	0.09	0.301
210	0.63	-0.119
211	N/A	-0.878
212	0.18	-0.134
213	N/A	-0.185
MEDIAN	*0.21*	*-0.113*

5.4.2 Relationships between honey yields and weather data

Honey harvest weight versus weather data are shown in Figure 5-13, with broad trends indicated but there is still overlap between each weather metric and the harvest quality

categories (poor < 20 kg, moderate 20 – 40 kg and good > 40 kg). This is borne out by the ANOVA analysis of the data, which is summarised in

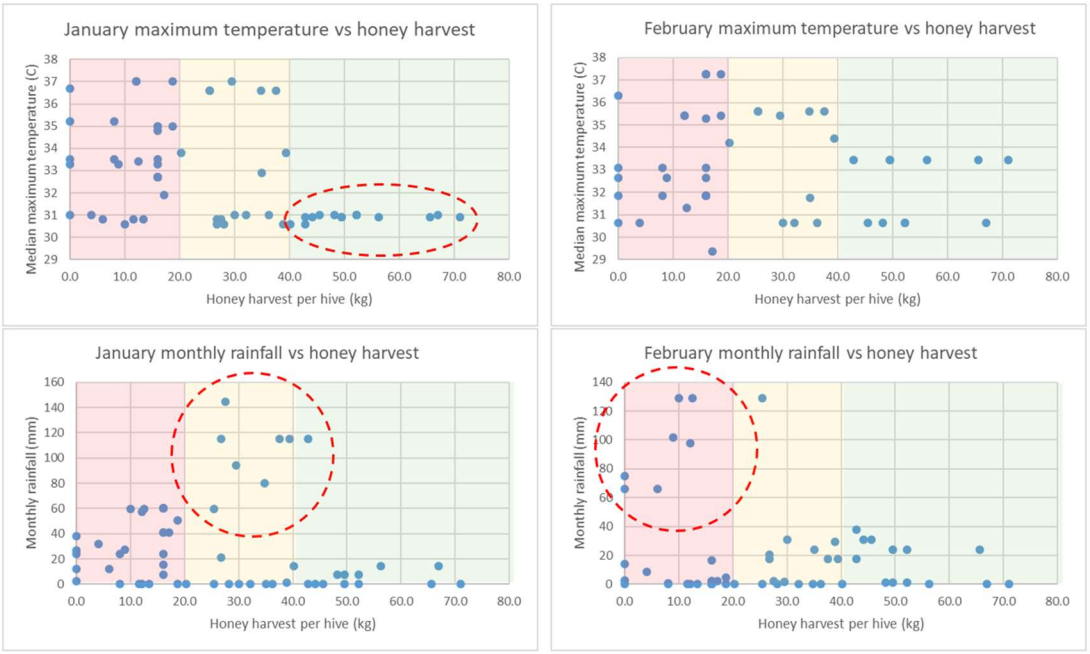


Figure 5-13: Monthly weather data vs honey harvest weight. Each data point is for a harvest site. Background is coloured by poor, moderate and good harvest classifications (red, yellow and green respectively)

Table 5-4. The only weather data that has a statistically significant difference is the January median maximum temperature (i.e. the median of the daily maximum temperatures in January). For this dataset, good harvest years do not exceed 31°C degrees whereas poor to moderate harvest years range from 30.5°C to 37.0°C. While there are other clusters in the weather data (circled in red in Figure 5-13: higher January rainfall for moderate harvests and higher February rainfall for poor harvests), these only apply to some data points for a given harvest category and therefore are not the only factor influencing harvest quality.

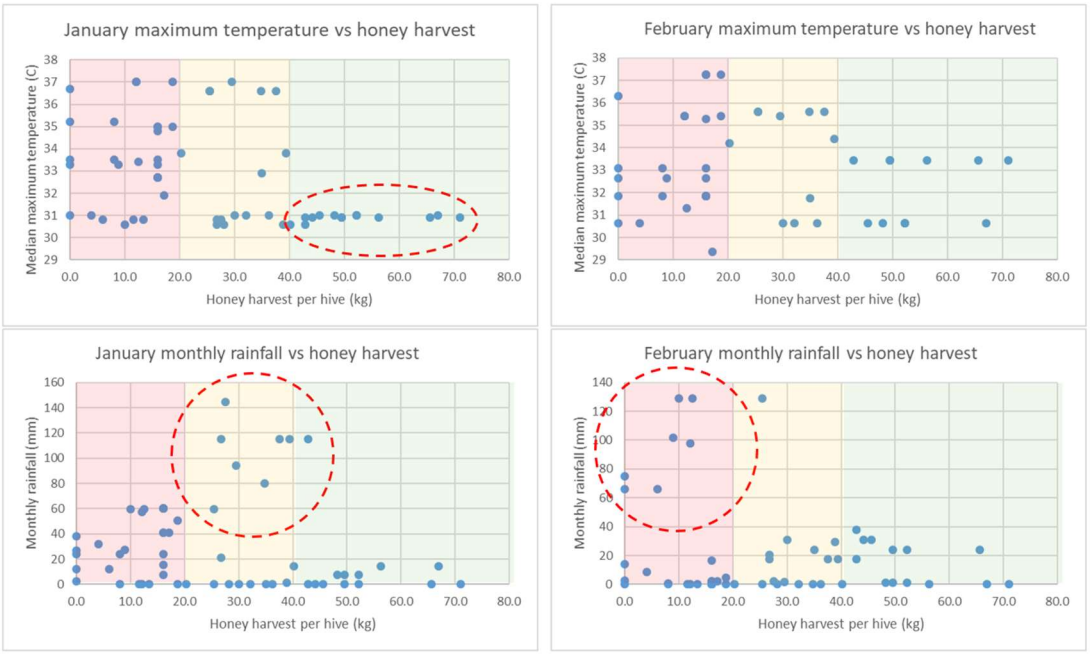


Figure 5-13: Monthly weather data vs honey harvest weight. Each data point is for a harvest site. Background is coloured by poor, moderate and good harvest classifications (red, yellow and green respectively)

Table 5-4: Monthly weather vs honey harvest ANOVA ρ-value results. Note that significant results (where $\rho < 0.05$) are highlighted in green.

COMPARISON	JANUARY TEMPERATURE	JANUARY RAINFALL	FEBRUARY TEMPERATURE	FEBRUARY RAINFALL
POOR VS MOD	$\rho > 0.05$	$\rho > 0.05$	$\rho > 0.05$	$\rho > 0.05$
POOR VS GOOD	$\rho < 0.05$	$\rho > 0.05$	$\rho > 0.05$	$\rho > 0.05$
MOD VS GOOD	$\rho < 0.05$	$\rho > 0.05$	$\rho > 0.05$	$\rho > 0.05$

To assess the effect of multiple weather and vegetation index variables on harvest quality, a series of cross-plots were produced, with the harvest quality grouped by colour. The results, shown in Figure 5-14, show the potential for combinations of measurements to improve on Boolean classification results compared to single variables alone (see combinations indicated in Figure 5-14). Interestingly, there are clusters and trends for the good and moderate harvest quality categories, but the poor harvest quality category is spread across the full range of results for all metrics except January rainfall. This indicates that there are a number of conditions that can cause harvest of poor quality, but only a few conditions that may support moderate or good quality harvests.

5.5 Discussion

5.5.1 MODIS derived indices and honey yields

The correlations between NDVI and honey yield show that the increased vegetation activity associated with increased nectar production has a larger effect than flower coverage alone (i.e. direct detection of increased flower coverage with the NDVI data is not feasible within this dataset). Comparison of the ANOVA analysis of NDVI and scaled Marri Flowering Index (sMFI) compared to honey harvests showed that January NDVI is a more reliable indicator of honey harvest weight than the sMFI.

To assess the accuracy of an NDVI cutoff point in classifying a year as good vs moderate or poor harvest quality, the percentages of years classified correctly for NDVI ranging between 0.4 and 0.6 were calculated. A cutoff at an NDVI of 0.5 was found to be the most accurate, with 79% of the 'good' years being above this point and 77% of the other years being below. Thus it is possible to predict honey harvest from satellite remotely sensed data, although the prediction is not reliable across all canopy covers nor harvest qualities.

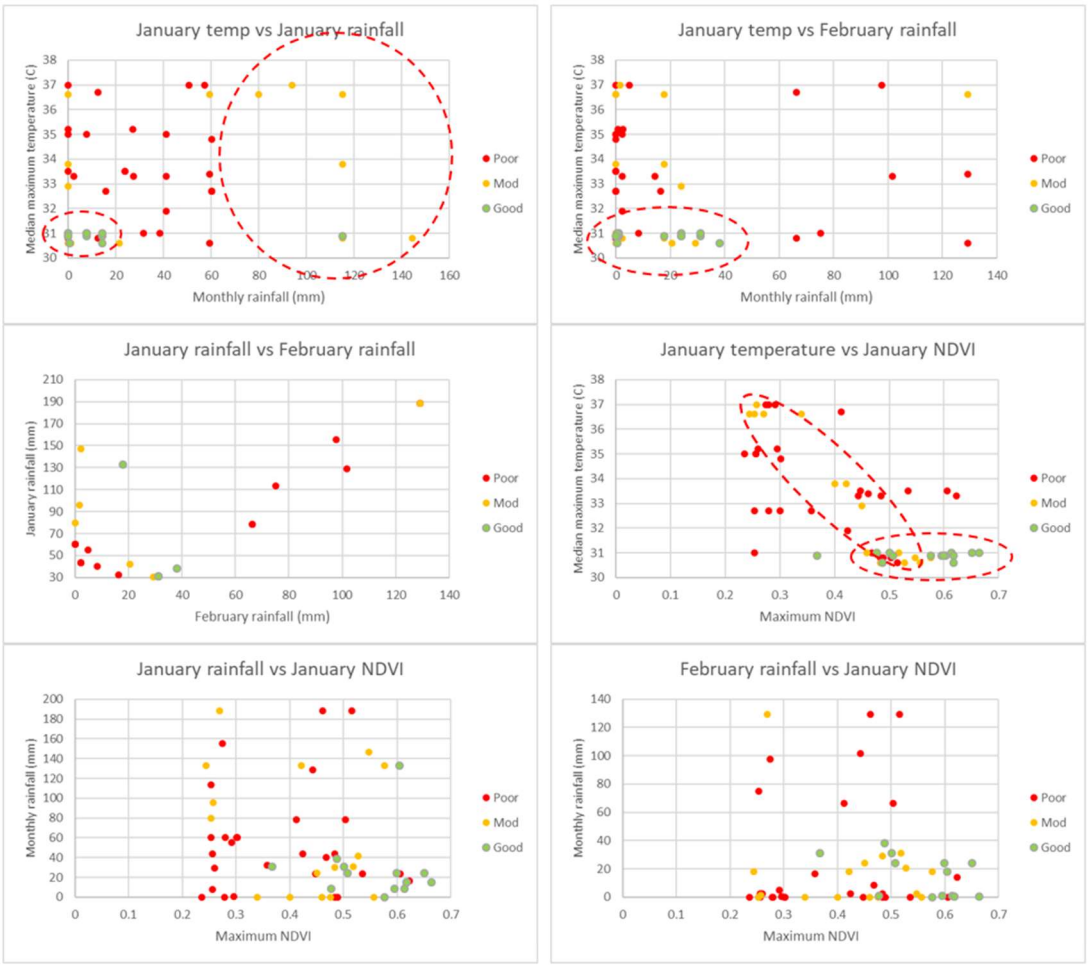


Figure 5-14: Cross-plots of weather and vegetation index data vs harvest quality. Red dashed lines highlight areas where good or moderate harvest datapoints cluster together well.

5.5.2 Relationships between honey yields and weather

The effectiveness of using multiple weather metrics to assign a honey yield quality was quantitatively tested by calculating the classification accuracy of different combinations of metrics, with the highest preceding this being a 78% accuracy in classifying good quality harvests using January maximum NDVI (note that moderate and poor quality harvest years were unable to be separated using NDVI). As summarised in Table 5-5, greater predictive accuracy is obtained using a combination of NDVI, median maximum temperature and monthly rainfall for January to classify a harvest as either good quality or of lower quality (accuracy of 90%). Of the 58 datapoints, two moderate quality years are classified as good years and four good years are classed as moderate years. Adjusting any of the three criteria to improve the number of good years that are correctly classified also increases the number of poor and moderate harvest years within the criteria, reducing the overall accuracy.

After classification of good quality harvests, the next most accurate prediction is classification of harvests better than poor quality (i.e. moderate or good quality years), with

at an accuracy of 72%. This uses a combination of the good quality harvest classification criteria as well as a set range of January temperature to January NDVI ratios for moderate years (see the moderate harvest quality range indicated by the red oval in Figure 5-14). As seen in Figure 5-13 and Figure 5-14, there is significant overlap between poor and moderate quality harvest years in the NDVI and weather metrics that is a limiting factor in the classification accuracy.

Moderate harvest quality years and poor quality harvest years were classified to 66% and 69% accuracies respectively, using a combination of the good harvest quality criteria and higher quality indices than poor harvest. As discussed above, there is significant overlap with the metrics used for poor harvest quality years and moderate harvest quality years, that limits the effectiveness of a Boolean classification approach. It may be that more detailed analysis of daily temperature, rainfall and/or NDVI data over the pre-flowering period may achieve better results than the monthly figures.

Attempts to create a combined classification approach for all three harvest quality levels yielded at best a classification accuracy of 69%, mostly due to the difficulty of separating moderate and poor harvest years.

The outcome of the inclusion of the weather data into the predictive model shows that the honey harvest weight prediction model based on satellite remotely sensed data improved the model's predictive capability. As only monthly weather data were used, it may be that a more detailed analysis of daily weather data may achieve greater accuracies for these poor and moderate harvest quality years. Greater predictability may also be achieved if apiary variables were incorporated in future models (such as queen bee quality and an index of hive strength as proposed by Delaplane et al. (2013)).

Table 5-5: Multivariate classification accuracy summary

TARGET	METRIC(S)	ACCURACY	TARGET	METRIC(S)	ACCURACY
GOOD HARVESTS	JAN NDVI > 0.5	79%	MODERATE HARVESTS	JAN NDVI > (-0.042*JAN TEMP+1.75)	45%
	JAN NDVI > 0.5 *AND IF* JAN TEMP < 31.5	83%		JAN NDVI < (-0.042*JAN TEMP+1.75) *AND IF* JAN NDVI > (-0.042*JAN TEMP+1.90) *AND IF* JAN NDVI < 0.5 *AND IF*	66%
	JAN NDVI > 0.5 *AND IF* JAN TEMP < 31.5 *AND IF* JAN RAIN < 40	90%			

				JAN TEMP > 31.5 *AND IF* JAN RAIN > 40	
TARGET	**METRIC(S)**	**ACCURACY**	**TARGET**	**METRIC(S)**	**ACCURACY**
> POOR HARVEST	JAN NDVI > 0.5 *AND IF* JAN TEMP < 31.5 *AND IF* JAN RAIN < 40 *AND IF* JAN NDVI < (-0.042*JAN TEMP+1.75) *AND IF* JAN NDVI > (-0.042*JAN TEMP+1.90)	72%	POOR HARVEST	JAN NDVI > 0.5 *OR IF* JAN TEMP < 31.5 *OR IF* JAN RAIN < 40	69%

5.6 Conclusions

Building on previous work in Europe and south-east Australia correlating satellite-derived vegetation indices to honey yields (summarised in the introduction sector of this paper), an assessment of NDVI versus honey harvest weight found a strong correlation between higher January MODIS NDVI and good quality marri honey harvests (> 40 kg of honey per hive). Using a Boolean classification approach with a criterion of maximum January NDVI > 0.5, and honey harvest data from 2010 – 2018 in southwest Australia across 16 wide-ranging apiary sites were classified correctly with 78% accuracy.

Incorporating climatic variables, particularly January monthly rainfall and median maximum daily temperature, with NDVI improved the classification of good quality harvest years to 90%. However, moderate and poor harvest quality years had a larger overlap of the weather and NDVI data for compared with good quality harvest years.

The results of this study will allow beekeepers to more accurately predict honey harvest quality prior to the start of honey production and adjust their preparation and site selection accordingly.

The approach used here in developing the predictive model can be readily applied to honey harvests from other floral sources, allowing beekeepers from other regions or countries to develop models suited to their own particular floral assemblages and climates.

6 Published Paper: Machine learning regression model for predicting honey harvests

Campbell T, Dods K, Dixon K, Fearns P, Handcock R (2020) Machine Learning Regression Model for Predicting Honey Harvests. Agriculture, 10 (4), pp. 118. https://doi.org/10.3390/agriculture10040118

agriculture

MDPI

Article

Machine Learning Regression Model for Predicting Honey Harvests

Tristan Campbell [1,*], Kingsley W. Dixon [2], Kenneth Dods [3], Peter Fearns [2] and Rebecca Handcock [4]

1 Computing and Mathematical Sciences, School of Electrical Engineering, Curtin University, Perth 6102, Australia
2 School of Molecular and Life Sciences, Curtin University, Perth 6102, Australia; kingsley.dixon@curtin.edu.au (K.W.D.); peter.fearns@curtin.edu.au (P.F.)
3 Chem Centre, Perth 6102, Australia; KDods@chemcentre.wa.gov.au
4 Curtin Institute for Computation, Curtin University, Perth 6102, Australia; rebecca.handcock@curtin.edu.au
* Correspondence: tristan.campbell@postgrad.curtin.edu.au; Tel.: +61-448-569-707

Received: 9 February 2020; Accepted: 18 March 2020; Published: 9 April 2020

Abstract: Honey yield from apiary sites varies significantly between years. This affects the beekeeper's ability to manage hive health, as well as honey production. This also has implications for ecosystem services, such as forage availability for nectarivores or seed sets. This study investigates whether machine learning methods can develop predictive harvest models of a key nectar source for honeybees, *Corymbia calophylla* (marri) trees from South West Australia, using data from weather stations and remotely sensed datasets. Honey harvest data, weather and vegetation-related datasets from satellite sensors were input features for machine learning algorithms. Regression trees were able to predict the marri honey harvested per hive to a Mean Average Error (MAE) of 10.3 kg. Reducing input features based on their relative model importance achieved a MAE of 11.7 kg using the November temperature as the sole input feature, two months before marri trees typically start to produce nectar. Combining weather and satellite data and machine learning has delivered a model that quantitatively predicts harvest potential per hive. This can be used by beekeepers to adaptively manage their apiary. This approach may be readily applied to other regions or forage species, or used for the assessment of some ecosystem services.

Keywords: remote sensing; weather; *Corymbia calophylla*; honey; machine learning; prediction

1. Introduction

The beekeeping industry in Western Australia has grown rapidly in the past decade, from 660 registered beekeepers in 2010 to over 3000 in 2019 [1]. In addition, honey produced from Western Australia has some of the highest antimicrobial properties known for honey [2]. As these high antimicrobial honey varieties are produced from marri (*Corymbia calophylla*, Myrtaceae) and jarrah (*Eucalyptus marginata*, Myrtaceae) trees that occur across a large area (see Figure 1) of approximately 84,000 km^2 [3], beekeepers often travel long distances to inspect apiary sites and manage their beehives. Access to a tool to predict areas of higher and lower honey production would make apiary management more efficient and improve industry safety by reducing the amount of rural driving required.

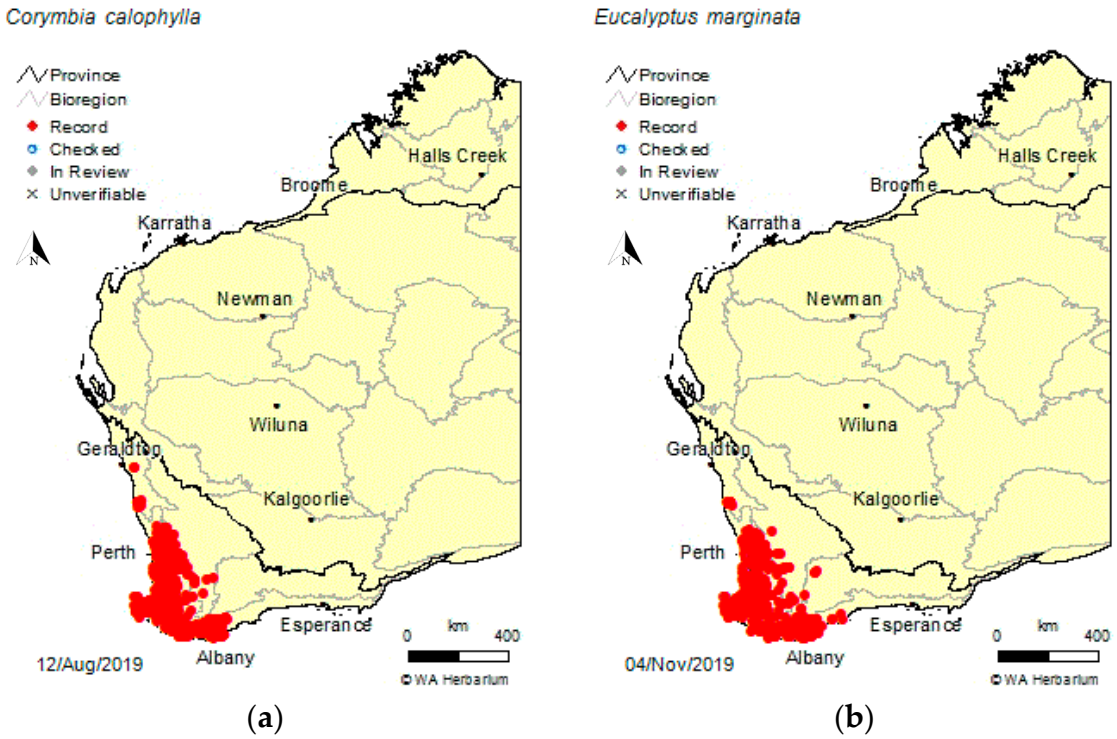


Figure 1. Geographic extent of *Corymbia calophylla* (**a**) and *Eucalyptus marginata* (**b**) [3].

Efforts to predict flowering patterns of Myrtaceous trees in Australia from weather data have often found complex relationships between weather and phenology. For example, a phenological study of four different species of *Eucalyptus* with coincident geographic ranges analysed 400 data points of flowering status and climate over 30 years [4] with the Generalised Additive Model for Location, Scale and Shape (GAMLSS) technique to assess the impact of minimum, mean and maximum temperature for the flowering period of those species, as well as flowering intensity in the preceding months and season. This study found that there is a non-linear relationship between temperature and flowering intensity, both of which varied between the four species. Two species flowered more intensely with increasing maximum temperatures, one species flowered more intensely with increasing minimum temperature and one species flowered less intensely with increasing maximum temperature. While temperature was consistently a key factor in flowering intensity, the effect was far from consistent.

While several studies on the relationship between satellite-derived vegetation indices and honey production in south-east Australia have shown promising outcomes [5–7], these have been qualitative in nature, with no demonstration of quantitative relationships, nor predictive model development.

Hawkins and Thomson [8] developed a qualitative, relativistic model for landscape/regional scale nectar availability in subtropical Eastern Australia, covering an area of 314,400 Ha. This study identified some key factors that were related to nectar availability, notably the Gross Primary Productivity for the previous 12 months and the average annual solar radiation and rainfall for the previous 6 months. The study was broad, covering more than 50 different plant species from many different genera.

In this paper, we use machine learning regression methods to assess the capability of both weather data and satellite-derived vegetation and moisture related data to develop a honey harvest prediction model for marri honey. This includes an assessment of the spatial density of weather station data from the Australian Government's Bureau of Meteorology (BOM). The primary aim is to identify the key factors that influence marri honey production and identify potential limitations in the data sources for these key factors.

2. Materials and Methods

2.1. Honey Harvest Data

Honey harvest data used for this study were from the same dataset used by Campbell, Fearns [9]. The dataset is from two apiarists: one 'commercial' apiarist with ~700 hives and one 'hobby' apiarist with ~50 hives. This consisted of harvest data from 2011 to 2018, across 16 apiary sites (with not all apiary sites being used every year). Honey harvest data were reduced to the average weight of honey harvested per hive, per year, per apiary site (see Table 1). The data were also classified by this measure as to whether the harvest was a 'poor year' (<20 kg honey per hive), 'moderate year' (20–40 kg honey per hive) or 'good year' (>40 kg honey per hive).

Table 1. Summary of honey harvest data by site and year.

Site	2011	2012	2013	2014	2015	2016	2017	2018
101	*N/A*	*N/A*	*N/A*	*N/A*	39.3	20.3	12.5	35.0
102	*N/A*	*N/A*	*N/A*	*N/A*	36.3	0.0	4.0	30.0
103	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	6.0	10.0	28.1
201	*Not used*	*Not used*	49.6	*Not used*	52.2	*Not used*	*Not used*	*Not used*
202	*Not used*	*Not used*	65.6	*Not used*	45.5	*Not used*	*Not used*	*Not used*
204	16.1	*Not used*	71.0	*Not used*	*Not used*	*Not used*	26.8	40.2
205	0.0	*Not used*	*Not used*	*Not used*	48.2	*Not used*	13.4	38.8
206	*Not used*	0.0	18.8	*Not used*	29.5	*Not used*	*Not used*	16.1
207	*Not used*	0.0	18.8	*Not used*	29.5	*Not used*	*Not used*	16.1
208	*Not used*	8.0	16.1	16.1	12.1	*Not used*	*Not used*	16.1
209	*Not used*	*Not used*	18.8	*Not used*	20.1	*Not used*	*Not used*	16.1
210	*Not used*	*Not used*	56.3	*Not used*	67.0	*Not used*	*Not used*	*Not used*
211	*Not used*	*Not used*	49.6	*Not used*	52.2	*Not used*	*Not used*	*Not used*
212	8.9	*Not used*	42.9	*Not used*	32.1	*Not used*	11.6	26.8
213	*Not used*	17.1	*Not used*	*Not used*	*Not used*	*Not used*	44.2	*Not used*

Red = 'poor year' (<20 kg per hive), yellow = 'moderate year' (20–40 kg per hive) and green = 'good year' (>40 kg per hive).

While most apiary sites were within 65 km of the capital city of Perth (Western Australia, −31.95 latitude 115.86 longitude), sites extended from as far north as Dandaragan (140 km north of Perth, −30.66 latitude 115.70 longitude) to as far south as Boyup Brook (230 km south-east of Perth, −33.83 latitude 116.38 longitude). The locations of the apiary sites are shown in Figure 2. All sites experience a warm temperate climate zone [10] and fall within the highly biodiverse South West Australian Floristic Region [11]. While all sites are within the same broad climatic region, the geographic extent of the sites means there is a range of annual weather trends within the study area, with the sites north of Perth being hotter and drier and the site south of Perth being cooler and drier (further inland than Perth). The mean annual and summer weather statistics for the three main areas are summarised in Table 2.

Table 2. Summary of key summary weather data for the three regions where sites are located.

REGION	ANNUAL		SUMMER (DECEMBER–FEBRUARY)	
	Mean Max. Temperature	Rainfall	Mean Max. Temperature	Rainfall
Dandaragan (north of Perth)	25.9 °C	484.5 mm	33.8 °C	32.4 mm
Mundaring (Perth Hills)	22.6 °C	1069 mm	29.7 °C	49.3 mm
Boyup Brook (south of Perth)	22.4 °C	649.1 mm	29.0 °C	44.7 mm

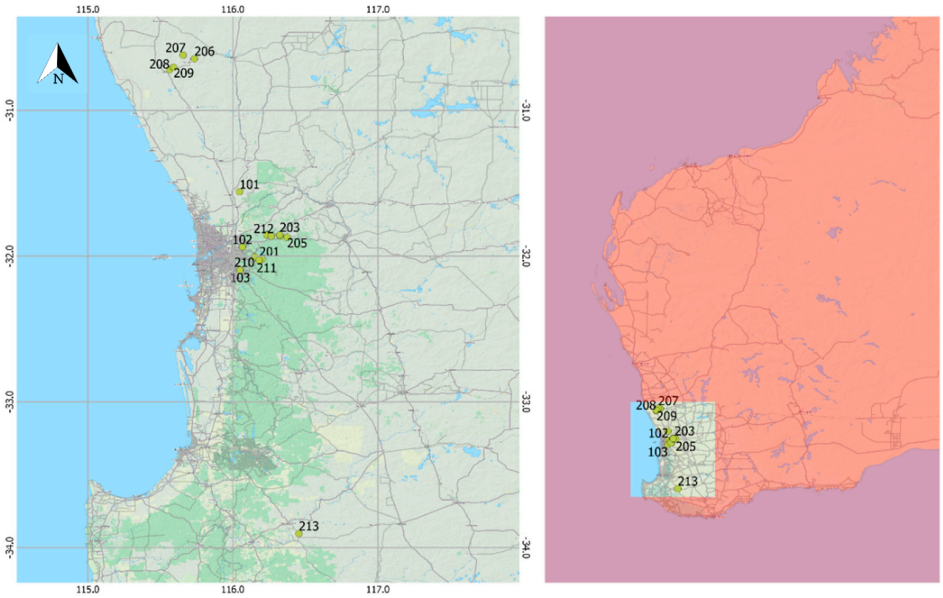


Figure 2. Apiary site locations indicated by yellow markers. Site labels correspond to those listed in Table 1. Coordinates are in World Geodetic System 1984 (WGS84).

There was also a large range of mature tree canopy cover across the sites. Using the vegetation structure products produced by Auscover [12] in combination with the height range of mature, flowering *Corymbia calophylla* (marri) trees of 10–40 m [13], the percentage of mature canopy cover for the apiary sites ranges from as low as 2.6% (for a farm site near Dandaragan) to as high as 61.9% (for a forest site in the Beelu National Park, near Mundaring) with a median mature canopy cover of 32.6%. Examples of these canopy cover extents are shown in Figure 3.

Figure 3. Examples of the minimum, median and maximum mature canopy cover across the apiary sites (images are all 1 km × 1 km extent).

2.2. Temperature, Rainfall and Solar Exposure Datasets

These standard meteorological datasets were retrieved from the Bureau of Meteorology (BOM)'s Australian Data Archive for Meteorology (ADAM), a database that holds weather observations dating back to the mid-1800s for some sites [14]. Daily weather observations stored within ADAM are readily accessible online via BOM's Climate Data Online portal [15]. Weather stations managed by BOM are installed and operated to consistent standards [16], providing a high degree of confidence when comparing data between years and sites.

While the individual weather station measurements have a high degree of confidence, weather stations in the study area are widely spaced, with up to 38.9 km between apiary sites and the nearest

weather station (Table 3). With the majority of the surveyed extent of the marri trees covering ~84,000 km^2 [3], only ~11,000 km^2 (or 13% of this area) has a temperature station within 10 km. This is shown spatially in Figure 4. The relatively sparse weather station network means that the extrapolation of temperature data to apiary sites may introduce some errors. Additionally, while a 10 km buffer from the rainfall recording stations covers 50% of the surveyed extent of marri trees, the spatial extent of rainfall events, particularly summer thunderstorms, can be much less.

Table 3. Distances between apiary sites and weather stations Bureau of Meteorology (BOM) weather station locations retrieved from [17].

Apiary Site	Nearest Temperature Station	Distance (km)	Nearest Rainfall Station	Distance (km)
101	PEARCE RAAF	12.3	MARBLING	3.7
102	BICKLEY	10.5	MAIDA VALE	5.2
103	BICKLEY	9.7	ROLEYSTONE	3.6
201	BICKLEY	6.2	BICKLEY	6.2
202	BICKLEY	23.9	CHIDLOW	5.5
203	BICKLEY	24.5	WOOROLOO	4.7
204	BICKLEY	19.3	LAKE LESCHENAULTIA	1.8
205	BICKLEY	27.0	WOOROLOO	6.9
206	BADGINGARRA RESEARCH STN	38.9	CHELSEA	4.9
207	BADGINGARRA RESEARCH STN	33.3	TAMBREY	2.4
208	LANCELIN (DEFENCE)	34.6	DANDARAGAN WEST	3.2
209	LANCELIN (DEFENCE)	37.9	DANDARAGAN WEST	1.2
210	BICKLEY	1.2	BICKLEY	1.2
211	BICKLEY	4.8	BICKLEY	4.8
212	BICKLEY	20.1	CHIDLOW	0.1
213	BRIDGETOWN COMPARISON	29.9	BOYUP BROOK	10.0

Shaded cells indicate either temperature stations further than 10 km from the apiary site or rainfall stations more than 2 km from the apiary site.

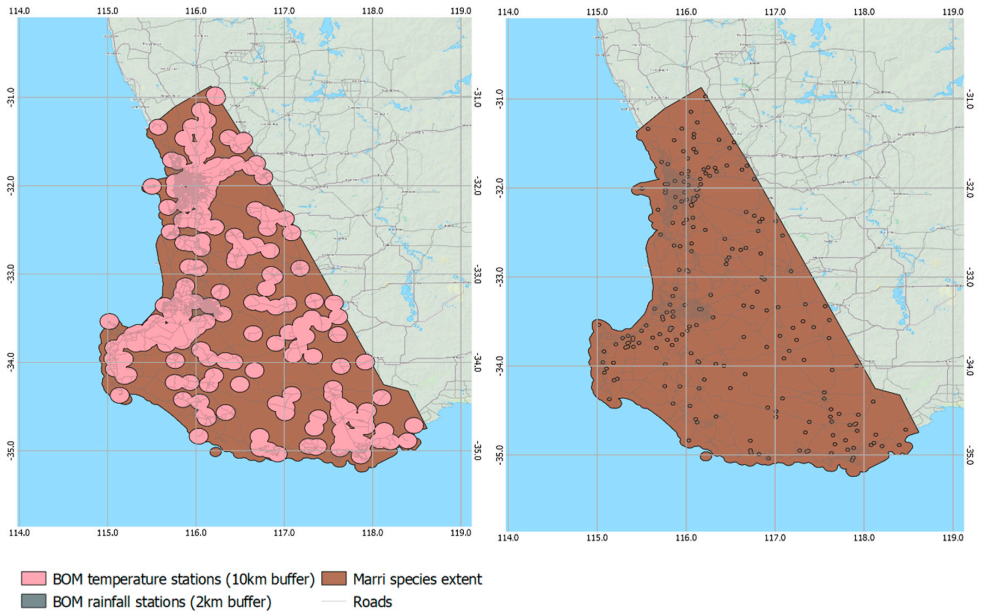
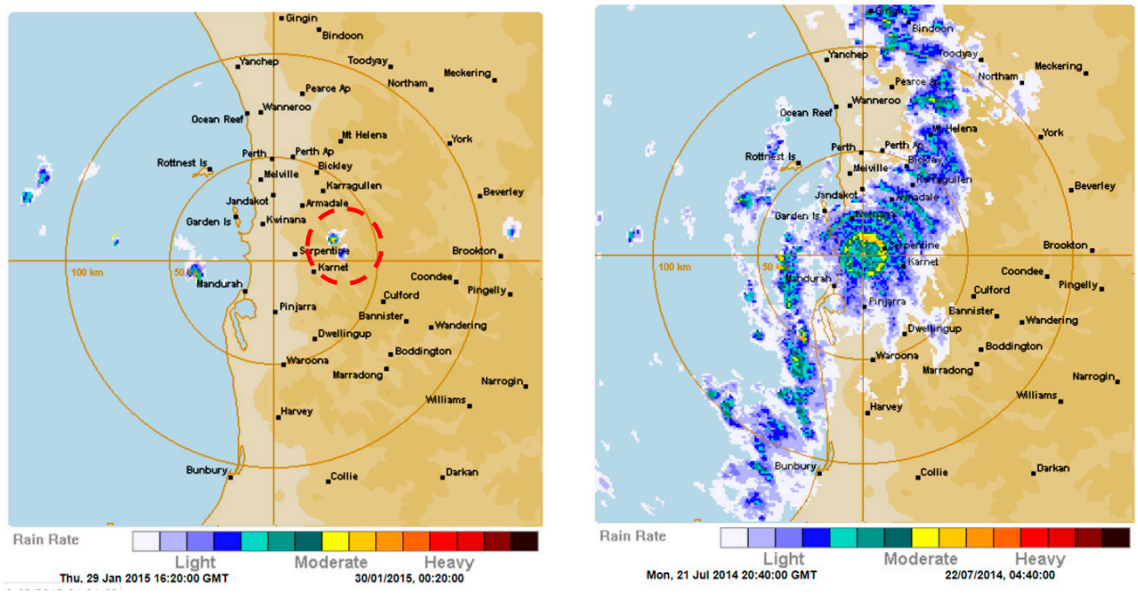


Figure 4. Coverage of Bureau of Meteorology (BOM) rainfall and temperature weather stations over the geographic extent of marri trees. Coordinates are in WGS84.

An example of the rainfall from such a storm event is shown in Figure 5. The main storm feature is approximately 18 km across its long axis but the more intense central area, where rainfall may be as high as 200 mm/h in this case, is just over 2 km across. Reducing the buffer around the BOM rainfall stations to 2 km means that only 3.1% of the marri spatial extent is sufficiently close to a rainfall station to reliably detect these isolated, yet intense, summer rain events. With summer rainfall being one of the key indicators of marri honey harvest weight [9], this lack of coverage may limit the ability of the BOM rainfall data to reliably classify or predict honey harvest weight.



(a) summer storm **(b)** winter cold front

Figure 5. Examples of rainfall events from the Perth region viewed in the Bureau of Meteorology (BOM) rain radar for Perth [18] for (**a**) a localised and intense summer storm event in January (highlighted in red), and (**b**) the storm front of a winter thunderstorm in July.

The distances between the apiary sites where the honey harvest data is from and the nearest weather stations are provided in Table 3, with temperature stations more than 10 km away and rainfall stations greater than 2 km away highlighted. A total of 75% of the apiary sites are more than 10 km from the nearest temperature station and 75% of the apiary sites are more than 2 km from the nearest rainfall station. Only one apiary site was within both 10 km of a temperature station and 2 km of a rainfall station.

2.3. Satellite-Derived Data

All other input datasets for modelling were derived from the Moderate Resolution Imaging Spectrometer (MODIS) sensor which is carried on the Terra and Aqua satellites [19], with data retrieved from the Application for Extracting and Exploring Analysis Ready Samples (AppEEARS) online portal [20]. A summary of the scientific basis for the creation of these datasets follows below. All datasets were retrieved from the AppEEARS portal at a pixel size of 1000 × 1000 m, from the pixel in which the apiary site is located. Retrieving data from coincident pixels meant that the subsequent integration of the data into a single database was a straightforward process. This pixel size was selected based on the typical forage range for honeybees with moderate nectar availability being 1–2 km [21]. This spatial resolution scale is also in line with studies elsewhere in the world investigating links between honeybee foraging and satellite data, such as eastern Australia [6] and Europe [22]. These studies concluded that, due to the foraging distances, the larger pixels of MODIS datasets were a more

accurate representation of the foraging conditions for a given apiary site than data from sensors with a higher spatial resolution.

All satellite derived data sets were filtered using the quality control channel contained in the AppEEARS download files. The particulars of the quality control parameters for each dataset can be found in the references provided in the following paragraphs. The quality control filter was set to only allow the highest quality data through into the machine learning algorithm input dataset.

The Gross Primary Production (GPP) MODIS-derived product was developed by Running and Nemani [23], based on the theory that productivity of crops with sufficient water and nutrient availability is linearly related to the amount of absorbed solar radiation and the efficiency of its use [24]. The MOD17 data product [25] used for this study incorporates the vegetation conversion efficiency factor (ε), as well as the effect of water stress and cold conditions on this conversion factor. Ground-based validation of the MOD17 data by Turner and Ritts [26] found that MOD17 GPP is responsive to general trends associated with local climate and land use, but tended to overestimate GPP in low productivity areas and underestimate GPP in high production areas.

The MOD16A2 Net Evapotranspiration data product [27] is based on the retrieval of key parameters of the Penman–Monteith evapotranspiration formula from MODIS data [28]. A reliability study of this data product in dry, heterogenous forests against ground measurements yielded a strong correlation of $r^2 = 0.82$ [29], indicating the high overall reliability of the product.

The Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) were both retrieved from the MOD13Q1 data product [30]. NDVI has been found to be strongly related to leaf chlorophyll content, whereas EVI is more indicative of canopy structural variations, including leaf area index (LAI) [31].

The Marri Flowering Index (MFI), developed by Campbell and Fearns [32], was designed for direct detection of marri flowers from MODIS data, based on an analysis of spectroradiometer surveys of this particular species. While the MFI has proven to be an effective index for classifying the marri honey harvest weight for some apiary sites with a high proportion of canopy cover [32], it appears to be less reliable across multiple apiary sites and years [9].

The Normalised Difference Water Index (NWDI) [33] is sensitive to changes in the liquid water content of vegetation canopies. This is a proven metric for vegetation water stress [34], which may impact bud development and nectar production.

2.4. Classification and Regression Tree Analysis

Classification and regression trees [35] are machine learning methods commonly used to build predictive models from input data [36]. Both methods partition the input dataset into smaller subsets and perform a simple prediction for each subset. By repeating this process over the entire dataset and combining the simple predictions together, a 'tree' type classification or regression structure is made. Classification trees are used for predictive models based on discreet classes or values, while regression trees are used for continuous variables or ranges.

To develop a predictive model, we tested two kinds of regression and classification approaches. Boosted Regression Trees (BRTs) are versatile, tree-based regression methods that can handle input features of different data types, while handling complex nonlinear relationships as well as interaction effects between the input features [37]. Random Forest trees have become popular in recent years for remote sensing and ecological applications due to their ability to handle both high data dimensionality and multicollinearity, which are common in multi-spectral remote sensing data, and the fact that they are both fast and insensitive to overfitting [38]. The Scikit-learn code for Python was used for the analysis [39], with the *GradientBoostingRegressor*, *GradientBoostingClassifier*, *RandomForestRegressor* and *RandomForestClassifier* functions initially used to assess which approach resulted in the most accurate predictive model.

2.5. Results

Input data were retrieved from several different sources for periods both preceding and during the main honey flow period for marri trees (February). The sources, periods and abbreviations used hereafter for these data are provided in Table 4. In total, a multidimensional database of 115 input features was created from the data summarised in Table 4.

Table 4. Summary of features used in the regression tree analysis.

Feature	Months Used	Data Source
MaT: Median maximum monthly temperature (°C)	Median maximum monthly temperature during flowering (MaTD) to the median maximum monthly temperature for the 11 months preceding flowering (MaT11)	Australian Data Archive for Meteorology (BOM) [15]
MiT: Median minimum monthly temperature (°C)	Median minimum monthly temperature during flowering (MiTD) to the median minimum monthly temperature for the 11 months preceding flowering (MiT11)	
T40: Number of days above 40 °C	Total number of days above 40 °C during flowering (T40D) to total number of days above 40 °C for the 6 months preceding flowering (T406)	
T25: Number of days below 25 °C	Total number of days below 25 °C during flowering (T25D) to the total number of days below 25 °C for the 6 months preceding flowering (T256)	
R: Monthly rainfall (mm)	Total rainfall during flowering (RD) to the total rainfall during the 11 months preceding flowering (R11)	
SRad: Nett mean monthly Solar Exposure (MJ/m^2)	Mean solar exposure during flowering (SRad) to the mean solar exposure for the 11 months preceding flowering (SRad11)	
GPP: Gross Primary Productivity ($kgCm^2$)	Gross Primary Productivity from 1 month prior to flowering (GPP1) to the Gross Primary Productivity from the 11 months prior to flowering (GPP11)	Application for Extracting and Exploring Analysis Ready Samples (AρρEEARS) [20]-MOD17A2H product
PSN: Net Photosynthesis	Net Photosynthesis from 1 month prior to flowering (PSN1) to the Net Photosynthesis from 11 months prior to flowering (PSN11)	AρρEEARS—MOD17A2H product
ET: Evapotranspiration ($kg/m^2/day$)	Evapotranspiration from 1 month prior to flowering (ET1) to the Evapotranspiration from the 11 months prior to flowering (ET11)	AρρEEARS—MOD16A2 product
LE: Average Latent Heat Flux ($J/m^2/day$)	Average Latent Heat Flux from 1 month prior to flowering (LE1) to the Average Latent Heat Flux from the 11 months prior to flowering (LE11)	AρρEEARS—MOD16A2 product
PET: Total Potential Evapotranspiration ($kg/m^2/day$)	Total Potential Evapotranspiration from 1 month prior to flowering (PET1) to the Total Potential Evapotranspiration from the 11 months prior to flowering (PET11)	AρρEEARS—MOD16A2 product
PLE: Average Potential Latent Heat Flux ($J/m^2/day$)	Average Potential Latent Heat Flux from 1 month prior to flowering (PLE1) to the Average Potential Latent Heat Flux from the 11 months prior to flowering (PLE11)	AρρEEARS—MOD16A2 product
NDVIA: Average Normalized Difference Vegetation Index	Average NDVI from 1 month prior to flowering (NDVIAv1) to the average NDVI for the 11 months prior to flowering (NDVIAv11)	AρρEEARS—MOD13A1 product
NDVIM: Maximum Normalized Difference Vegetation Index	Maximum NDVI from 1 month prior to flowering (NDVIMx1) to the maximum NDVI for the 11 months prior to flowering (NDVIMx11)	AρρEEARS—MOD13A1 product
EVIA: Average Enhanced Vegetation Index	Average EVI from 1 month prior to flowering (EVIAv1) to the average EVI for the 11 months prior to flowering (EVIAv11)	AρρEEARS—MOD13A1 product
EVIM: Maximum Enhanced Vegetation Index	Maximum EVI from 1 month prior to flowering (EVIMx1) to the maximum EVI for the 11 months prior to flowering (EVIMx11)	AρρEEARS—MOD13A1 product
MFI: Marri Flowering Index	Maximum Marri Flowering Index (MFI) value for February	AppEEARS—derived from MODOCGA product
NDWI: Normalized Difference Water Index	Average NDWI from 1 month prior to flowering (NDWIAv1) to the average NDWI for the 11 months prior to flowering (NDWIAv11)	AppEEARS—derived from MOD09A1 product

To assess the predictive accuracy of each algorithm, both the regression and classifier functions for the gradient boosted and Random Forest tree algorithms were run for all 115 input features, with an increasing number of trees in the model. No complexity limit was given due to the relatively small size of the dataset and the resulting short run time. The output from this testing (see Table 5) shows that the Random Forest functions worked better for the classification approach, based on the 'poor year', 'moderate year' and 'good year' ratings, and the Gradient Boosted functions worked better for predicting the honey harvest weights via the regression model.

Table 5. Summary of predictive errors for different sized Random Forests.

Number of Algorithm Trees	RANDOM FOREST TREES			GRADIENT BOOSTED TREES		
	Honey Weight Regression	Honey Weight Classification	Honey Class Classification	Honey Weight Regression	Honey Weight Classification	Honey Class Classification
5	11.68 kg	25%	50%	13.49 kg	50%	42%
10	13.63 kg	25%	50%	11.38 kg	25%	42%
20	12.46 kg	25%	42%	10.42 kg	18%	42%
50	11.48 kg	25%	42%	10.55 kg	18%	42%
100	11.31 kg	25%	33%	10.35 kg	18%	42%
200	11.63 kg	25%	33%	10.33 kg	18%	42%
500	11.56 kg	25%	33%	10.33 kg	18%	42%
1000	11.29 kg	25%	33%	10.33 kg	18%	42%
5000	11.43 kg	25%	33%	10.33 kg	18%	42%

Note that the 'Honey weight classification' is performed by doing the 'Honey weight regression', then classifying the predicted honey weight into the appropriate class of 'poor year', 'moderate year', or 'good year'. The 'Honey class classification' output is provided by the relevant 'Classifier' function. The lowest predictive errors are highlighted in green.

The lowest predictive errors came from the use of Gradient Boosted Regression (GBR), with a mean average error of +/− 10.3 kg for the weight, with this weight being in the correct class 82% of the time.

The partial dependence plots for the 10 features with the highest feature importance from the full feature input model are shown in Figure 6. These plots show that the model predictions have the highest dependence on the mean maximum temperature for October–January (MaT4), with mean maximum temperatures below 27 °C having a strong positive relationship and mean maximum temperatures above 27 °C having a negative relationship. There are also strong relationships with Gross Primary Production from the 11 months preceding the flowering period (GPP11), with low GPP associated with lower honey production, and a higher evapotranspiration for January (ET1) having a positive relationship with higher honey production.

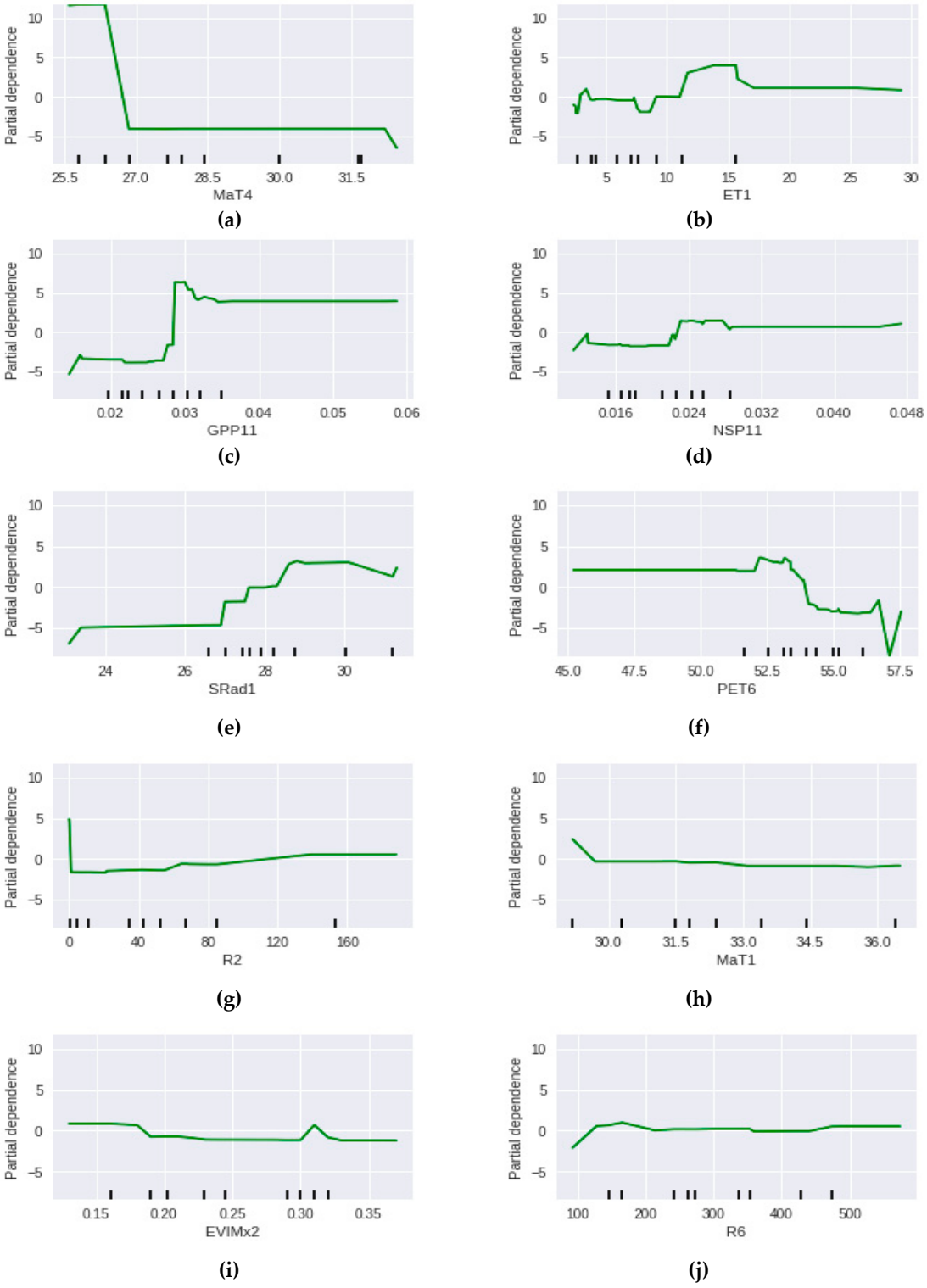


Figure 6. Partial dependence plots for 10 most important features (Gradient Boosted Regression). Refer to Table 4 for the feature acronym definitions. (**a**) Average maximum temperature for October to January (**b**) evapotranspiration for January (**c**) Gross Primary Productivity from March to January (**d**) Net Photosynthesis For March to January (**e**) Solar Radiation for January (**f**) Total Potential Evapotranspiration For August to January (**g**) rainfall from December to January (**h**) average maximum temperature in January (**i**) maximum Enhanced Vegetation Index from December to January (**j**) rainfall from August to January.

In order to determine whether the dimensionality of the input features could be reduced without compromising the predictive accuracy of our models, the GBR function was run multiple times with fewer input features for each run. The selection of input features for each run was on the basis of the highest feature importance from the preceding model. Table 6 shows the VIs and predictive errors for this process. The MaT4 feature is clearly the most important input, with a VI of 37.7% even when all 115 features are used in the regression modelling. This is followed by ET1 and GPP, both at below 10% when all feature inputs are used. The columns highlighted in green show the models with the lowest error, by class (five input features) and by weight (two input features). These reduced input feature models both have slightly higher honey weight prediction errors than the full feature model (increased error of 0.52 kg versus a honey harvest range of 0.0 kg to 71.0 kg in the training and testing datasets). If this simplified approach were to be applied to larger datasets for the prediction of honey harvest weight, a significantly smaller number of feature inputs can be used, with a minimal reduction in predictive accuracy, reducing the size of the multidimensional dataset required.

Table 6. Gradient Boosted Regression errors and feature importance (import.) for differing number of input features.

# Features	All	10	8	6	5	4	3	2	1
Honey weight	10.33 kg	11.75 kg	9.56 kg	12.87 kg	10.85 kg	11.61 kg	10.91%	10.42 kg	11.72 kg
Honey class. (1–3)	17%	17%	17%	25%	8%	17%	17%	17%	25%
Honey class. (1–2)	0%	0%	0%	0%	0%	0%	0%	0%	0%
MaT4 import.	37.7%	30.7%	31.7%	45.7%	48.9%	56.8%	57.0%	63.2%	100%
ET1 import.	7.9%	18.6%	15.4%	25.0%	26.3%	29.3%	31.0%	36.8%	
GPP11 import.	5.9%	11.3%	14.4%	10.0%	11.1%	8.4%	12.0%		
NSP11 import.	5.7%	9.8%	12.5%	7.5%	7.3%	5.5%			
SRad1 import.	4.0%	8.0%	8.8%	6.0%	6.5%				
PET6 import.	3.7%	6.2%	6.6%	5.8%					
R2 import.	3.7%	5.4%	6.1%						
MaT1 import.	3.2%	4.6%	4.6%						
EVIMx2 import.	2.8%	3.9%							
R6 import.	2.7%	1.6%							

The models with the lowest errors are highlighted in green. Gray cells are where the input features were not used in the classification algorithm.

The 'Honey class. (1–2)' row in Table 6 is calculated by classifying the predicted honey weight as either 'good year' or 'below good year', based on the high degree of clustering found with key weather and satellite inputs for 'good years' by Campbell and Fearns [9]. This tight clustering for 'good years', and the 0% error in the predictive regression classification, means that for this dataset the 'good year' prediction is both an accurate and a robust model. While this cannot assist apiarists with preparation for 'poor years', when their bees can sometimes starve due to the lack of nectar, the apiarists can prepare for a 'good year' in advance and therefore increase the honey production, compared with being unprepared for a 'good year' and having insufficient hives or other equipment to facilitate efficient use of the abundant resource.

Although the predictive regression model developed solely from the MaT4 and ET1 features has a relatively low error and is shown to be quite robust at predicting harvests that are 'good years', both input features require data from the month immediately preceding the honey flow. If the predictive model developed here was used operationally by beekeepers to adaptively manage their beehives, this timing of the key input data would limit the time available for apiarists to prepare for a 'good year'. To determine whether a predictive model could be generated with a longer lead time into the honey flow, mean maximum temperatures for the individual months from October to January were also tested in the GBR, as well as combinations of these months. The VIs from this process are summarised in Table 7. While the original MaT4 and ET1 features retained a high importance, MaT3 (mean maximum temperature for November) was also assessed as a key feature input, based on a visual assessment of

the clustering of each feature versus honey harvest weight (see Figure 7). While the model does not produce the lowest errors, it is nonetheless a reliable predictor (particularly for 'good years' versus non-good years). The lower accuracy of the prediction is offset by the lead time to the honey flow; with honey flow generally starting in late January to early February [32], having a strong indicator of an upcoming good harvest by the end of November gives apiarists approximately two months to prepare for the predicted conditions.

Table 7. Gradient Boosted Regression errors and the feature importance of mean temperatures 1–4 months before flowering and evapotranspiration 1 month before flowering.

# Features	10	8	6	4	3	MaT3 + ET1	MaT3
Honey weight	10.09 kg	10.34 kg	9.64 kg	11.16 kg	10.44 kg	8.92 kg	11.72 kg
Honey class. (1–3)	17%	25%	17%	25%	17%	17%	25%
Honey class. (1–2)	0%	0%	0%	0%	0%	0%	0%
ET1 import.	28.3%	28.3%	28.4%	35.1%	36.8%	38.1%	
MaT 1–4 import.	24.5%	35.1%	33.8%	17.1%	38.3%		
MaT 2–3 import.	24.1%	13.9%	16.9%	37.5%	24.9%		
MaT 2 import.	5.7%	8.2%	8.9%	10.4%			
MaT 3–4 import.	3.2%	3.4%	3.2%				
MaT 1 import.	8.9%	8.8%	8.9%				
MaT 2–4 import.	1.2%	1.8%					
MaT 1–2 import.	3.6%	0.6%					
MaT 3 import.	0.4%					61.9%	100%
MaT 4 import.	0.2%						

The models with the lowest errors are highlighted in green. Gray cells are where the input features were not used in the classification algorithm.

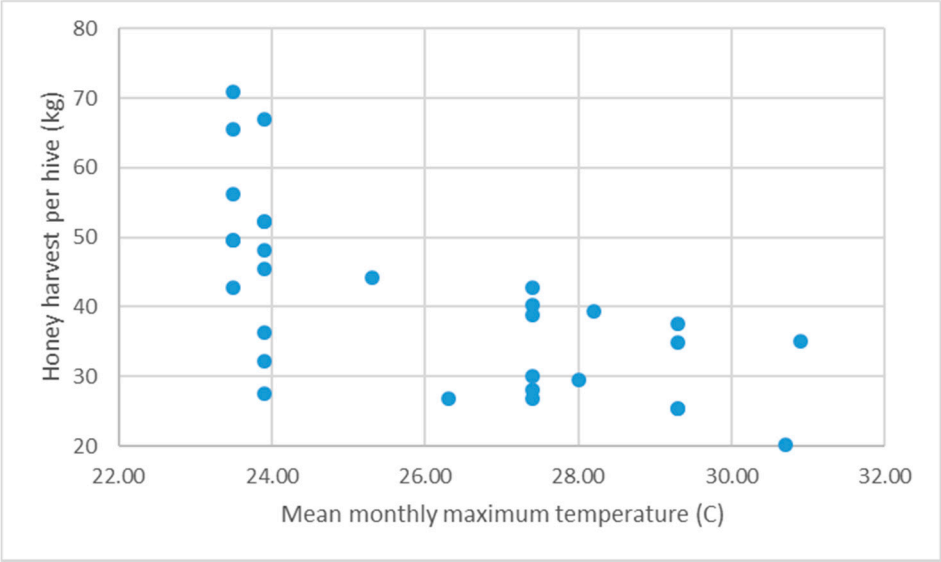


Figure 7. Mean monthly maximum temperature for November (MaT3) versus honey harvest weight.

3. Discussion

From the regression tree analysis performed in this study, there are a range of different factors that influence the honey harvest of marri trees and the timeframe of some of these factors (e.g., Gross Primary Productivity) can be as much as a year before the honey flow starts. Despite the range of available input variables that we explored, we found that the inputs to the predictive model can be reduced significantly with a minimal reduction in the accuracy of the predictive model. The mean maximum temperature for the few months preceding honey flow (November to January) and the evapotranspiration in the month before honey flow (January) were found to describe the majority of

the variability in honey harvest weight almost as accurately as the full multidimensional dataset of 115 input features.

The potential limitations in the accuracy of the weather station data available from BOM, due to the sparsity of stations across the extent of marri trees versus the spatial variability of the observations, may mean that some relationships between marri honey production and weather may not be fully captured in the existing database. The sparse nature of rainfall stations compared to the localised nature of summer rainfall events means that the influence of rainfall in particular may have been poorly characterised in our input dataset. Unfortunately, due to this sparsity of rainfall stations, excluding apiary sites from the database that have temperature stations more than 10 km away and/or rainfall stations more than 2 km away would result in only one apiary site in the database with honey harvests from two of the study years. This is insufficient data for the development of a predictive model, particularly as both of these years are 'good years' (>40 kg of marri honey per hive).

Rainfall immediately preceding and during the flowering period is one of the key factors in harvest quality (see Figure 8), with heavy summer rainfall events after flowering commences actually knocking stamen, and sometimes flowers, from the trees [40]. This presents an issue with the development of a predictive model as, even with good conditions in the lead up to the peak flowering period, a localised storm may downgrade a prospective harvest from a 'good year' to a 'moderate year' or even a 'poor year' in a matter of hours.

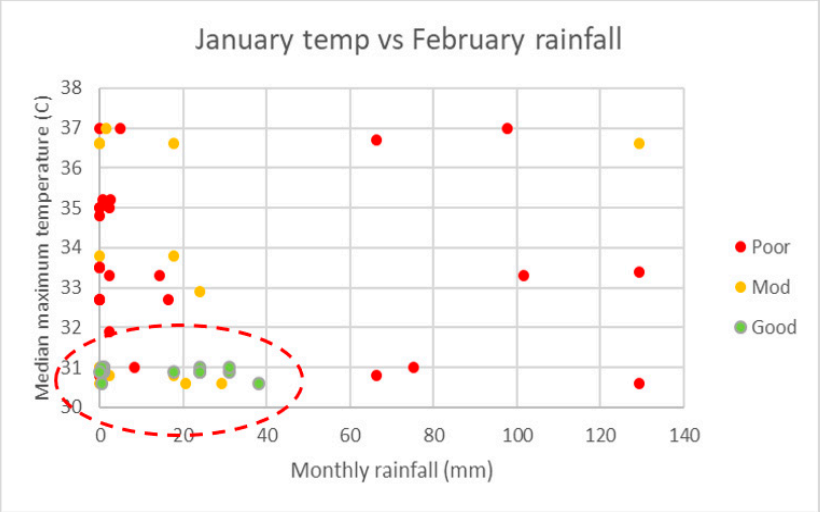
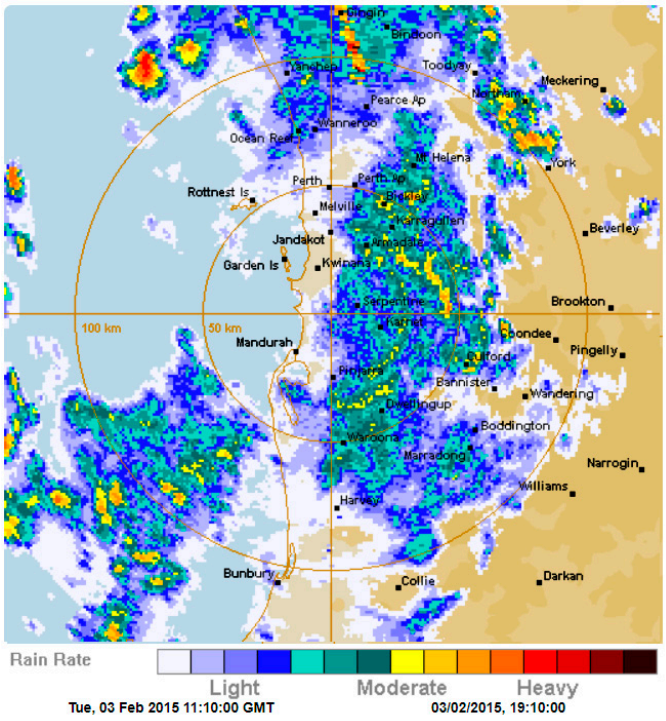


Figure 8. Relationship between January temperature February rainfall and honey harvest (excerpt from Campbell and Fearns [9]).

Such a localised storm may have been the cause of the quality of some of the 'moderate year' harvests in the Perth Hills in 2015 being incorrectly classified as 'good years' by the predictive model; the November mean maximum temperature was within the range for a good harvest that season, but three of the eight apiary sites only yielded a harvest that was a 'moderate year'. Rainfall for all of the apiary sites was below 40 mm for February (in the range 18.6–30.0 mm). The rainfall radar image from the main rainfall event at the start of the month, shown in Figure 9, contains localised patches of more intense rainfall (up to 80 mm/hr) that are under 2 km across. The presence of one of these localised higher intensity zones over an apiary site may well have increased the rainfall received by the site to over 40 mm for the period, the upper limit for good harvests from the currently available data [9].



Level	Colour		Approx. Rainfall Intensity (mm/hr)
0	clear	Not visible	Under 0.2
1		Off-white	0.5
2		Sky-blue	1.5
3		Light Blue	2.5
4		Blue	4
5		Light Cyan	6
6		Cyan	10
7		Dark Cyan	15
8		Yellow	20
9		Yellow-orange	35
10		Orange	50
11		Orange-red	80
12		Red	120
13		Dark Red	200
14		Maroon	300
15		Dark Brown	over 360

Figure 9. Rainfall radar map for min rainfall event over the Perth Hills in early February 2015 [18,41].

With marri honey being the largest annual honey harvest in Western Australia [42], the highly specific conditions required for a good harvest may be less likely to occur with the projected climate models for the extent to which marri trees occur. The Regional Climate Model developed for southwest Western Australia by Andrys and Kala [42] assessed projections of both mean and extreme weather events over the period from 2030 to 2059 under a high greenhouse gas emissions scenario. Compared with historic data from 1970 to 1999, mean maximum temperatures for spring and summer are projected to increase by between 0.5–2.0 °C. With more than 80% of the 'good years' occurring in seasons where the preceding November mean maximum temperature is less than 24.0 °C. As the November mean maximum temperature for the Perth Hills region is already at 25.0 °C, the probability of this criteria for a good harvest being met will likely decrease. In addition, while the projections for summer rainfall vary considerably between models, the models all predict that the intensity of summer rainfall events will increase, resulting in an increase in the probability of localised rainfall events of sufficient intensity to reduce the harvest quality.

4. Conclusions

The development of a multidimensional database with 115 factors that may influence the honey production of marri trees was subjected to a regression tree analysis. The GBR models achieved the highest predictive accuracy when all features from the multi-dimensional dataset were used, with an average error of +/− 10.33 kg for the weight, with the weight being in the correct class 82% of the time. A similar level of predictive accuracy was achieved with a revised GBR model using the input features with the five highest values of feature importance. This regression model was able to predict the honey yield per hive with a mean average error (MAE) of 10.85 kg and classify the harvest into the correct quality category with 92% accuracy.

Analysis of the predictive accuracy of GBRs with different input features highlighted that 'good years' could be predicted robustly compared with 'moderate years' and 'poor years' (100% accuracy with only one or two input features in the models). With mean maximum temperature in the months leading up to the marri honey flow consistently rating among the highest importance values, testing

the model with various combinations found that using the November mean maximum temperature on as the only input feature into the GBR algorithm could predict the honey harvest weight per hive to an MAE of 11.72 kg, classify it into the correct class with 75% accuracy and predict a 'good year', versus other types of year, to 100% accuracy. With the honey flow typically starting in February, giving apiarists a reliable predictor of a good season two months ahead of time will allow them to prepare their equipment, including establishing new hives, before the flow starts to improve the production in years with good honey harvests.

In an operational setting, the accuracy of the prediction models may be restricted by the availability of weather data, with less than 13% of the extent of marri trees having a temperature station within 10 km. Intense localised summer rainfall events also have an important role to play in making the model more widely applicable. Only 3.1% of the marri areas have a rainfall station within 2 km. Access to more spatially accurate weather data may be able to improve the regression model's predictive accuracy.

With cooler weather and lower summer rainfall required for good marri honey harvests, the climate projections for southwest Western Australia indicate that good harvests are likely to become rarer in the future, with mean maximum temperatures and the intensity of summer rainfall both projected to increase.

While this study has been focused on honey production from a single species endemic to South West Australia, the weather stations and satellite data used to develop the model, for example, the Royal Netherlands Meteorological Institute (KNMI) Climate Explorer tool [43], have collected over 10 TB of weather data from multiple weather agencies around the world (including the BOM data used for this study). The MODIS input feature data are freely available global datasets produced by NASA. If these data are used in conjunction with honey harvest data from other species and/or regions, predictive models could foreseeably be developed using the methodology employed for in study for other honey harvests.

Author Contributions: Conceptualization, T.C. and P.F.; methodology, T.C.; investigation, T.C.; writing-original draft preparation, T.C.; writing-review and editing, K.W.D., K.D., P.F. and R.H.; funding acquisition, K.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the Beekeeping Industry Council of Western Australia (BICWA), the Western Australian government's Department of Primary Industries and Regional Development (DPIRD) and ChemCentre as part of the Grower Groups Research and Development Grant, Round 2-GGRD2 2016-1700179-INDUSTRY STANDARDS OPTIMISING STORAGE AND SUPPLY VOLUME OF WA MONO-FLORAL HONEY.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Thomson, J. *Western Australia a Sweet Spot for Beekeeping*; Department of Primary Industries and Regional Development: Perth, Australia, 2019.
2. Irish, J.; Blair, S.; Carter, D. The Antibacterial Activity of Honey Derived from Australian Flora. *PLoS ONE* **2011**, *6*, e18229. [CrossRef]
3. Herbarium, W.A. *Florabase—the Western Australian Flora*; Department of Environment and Conservation: Perth, Australia, 1998.
4. Hudson, I.L.; Kim, S.; Keatley, M. Climatic influences on the flowering phenology of four Eucalypts: A GAMLSS approach. In Proceedings of the 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Cairns, Australia, 13–17 July 2009.
5. Arundel, J.; Winter, S.; Gui, G.; Keatley, M. A web-based application for beekeepers to visualise patterns of growth in floral resources using MODIS data. *Environ. Model. Softw.* **2016**, *83*, 116–125. [CrossRef]
6. Webber, E. *Eucalypt Leaf-Flush Detection from Remotely Sensed (MODIS) Data*; Department of Infrastructure Engineering-Geomatics, University of Melbourne: Melbourne, Australia, 2011.
7. Winter, S.; Leach, J.; Keatley, M.; Arundel, J. *BeeBox Application User Manual*; Burns, C., Ed.; Rural Industries Research and Development Corporation: Canberra, Australia, 2013.

8. Hawkins, B.; Thomson, J.; Mac Nally, R. Regional patterns of nectar availability in subtropical eastern Australia. *Landsc. Ecol.* **2018**, *33*, 999–1012. [CrossRef]
9. Campbell, T.; Fearns, P.; Dods, K.; Dixon, K. Prediction and detection of honey harvests from remote sensing and weather data. *Int. J. Eng. Sci. Res. Technol.* **2019**, *8*, 7–88.
10. Peel, M.C.; Finlayson, B.L.; Mcmahon, T.A. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci. Discuss.* **2007**, *4*, 439–473. [CrossRef]
11. Beard, J. *A New Phytogeographic Map of Western Australia*; Western Australian Herbarium Research Notes; Western Australian Herbarium: Kensington, Australia, 1980; Volume 3, pp. 37–58.
12. Scarth, P. *Vegetation Height and Structure—Derived from ALOS-1 PALSAR, Landsat and ICESat/GLAS, Australia Coverage*; Phinn, S., Ed.; Joint Remote Sensing Research Program, University of Queensland: Brisbane, Australia, 2009.
13. Brooker, M.I.H.; Kleinig, D.A. Field guide to eucalypts. In *South-western and Southern Australia*; Bloomings Books: Melbourne, Australia, 2001; Volume 2.
14. Meteorology, B.O. (Ed.) Australian Data Archive for Meteorology. In *Conference on Managing Australian Climate Variability*; NSW: Albury, Australia, 2000.
15. Bureau of Meteorology. Climate Data Online. Available online: http://www.bom.gov.au/climate/data (accessed on 21 April 2019).
16. Canterford, R. *Guidelines for the Siting and Exposure of Meterological Instruments and Observing Facilities*; Bureau of Meteorology, Department of the Environment, Sports and Territories: Melbourne, Australia, 1997.
17. Bureau of Meteorology. Weather Station Directory. Available online: http://www.bom.gov.au/climate/data/stations/ (accessed on 8 February 2019).
18. The Weather Chaser. Perth Radar—128km Rain Rate. Available online: http://www.theweatherchaser.com/radar-loop/IDR703-perth-serpentine (accessed on 21 March 2019).
19. Barnes, W.L.; Pagano, T.S.; Salomonson, V.V. Prelaunch characteristics of the moderate resolution imaging spectroradiometer (MODIS) on EOS-AMI. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1088–1100. [CrossRef]
20. AppEEARS Team. Application for Extracting and Exploring Analysis Ready Samples (AppEEARS). NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center: Sioux Falls, SD, USA. Available online: https://lpdaacsvc.cr.usgs.gov/appeears/ (accessed on 12 March 2019).
21. Hagler, J.R.; Mueller, S.; Teuber, L.R.; Machtley, S.A.; Van Deynze, A. Foraging range of honey bees, Apis mellifera, in alfalfa seed production fields. *J. Insect Sci.* **2011**, *11*, 144. [CrossRef] [PubMed]
22. Lynn, B.C. Relation of Honey Production in Apis Mellifera Colonies to the Normalized Difference Vegetation Index and Other Indicators. Ph.D. Thesis, Department of Geography, University of North Carolina, Chapel Hill, NC, USA, 2013.
23. Running, S.W.; Nemani, R.R.; Heinsch, F.A.; Zhao, M.; Reeves, M.; Hashimoto, H. A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production. *BioScience* **2004**, *54*, 547–560. [CrossRef]
24. Monteith, J.L. Solar Radiation and Productivity in Tropical Ecosystems. *J. Appl. Ecol.* **1972**, *9*, 747–766. [CrossRef]
25. Running, S.W.; Mu, Q.; Zhao, M. *MOD17A2H MODIS/Terra Gross Primary Productivity 8-Day L4 Global 500m SIN Grid V006*; NASA EOSDIS Land Processes DAAC, 2015. Available online: https://lpdaac.usgs.gov/products/mod17a2hv006/ (accessed on 24 March 2019).
26. Turner, D.P.; Ritts, W.D.; Cohen, W.B.; Gower, S.T.; Running, S.W.; Zhao, M.; Costa, M.H.; Kirschbaum, A.A.; Ham, J.M.; Saleska, S.R.; et al. Evaluation of MODIS NPP and GPP products across multiple biomes. *Remote Sens. Environ.* **2006**, *102*, 282–292. [CrossRef]
27. Running, S.W.; Mu, Q.; Zhao, M. *MOD16A2 MODIS/Terra Net Evapotranspiration 8-Day L4 Global 500m SIN Grid V006*; NASA EOSDIS Land Processes DAAC, 2017. Available online: https://lpdaac.usgs.gov/products/mod16a2v006/ (accessed on 24 March 2019).
28. Mu, Q.; Heinsch, F.A.; Zhao, M.; Running, S.W. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sens. Environ.* **2007**, *111*, 519–536. [CrossRef]
29. Miranda, R.D.Q.; Galvíncio, J.D.; Moura, M.S.B.D.; Jones, C.A.; Srinivasan, R. Reliability of MODIS Evapotranspiration Products for Heterogeneous Dry Forest: A Study Case of Caatinga. *Adv. Meteorol.* **2017**, *2017*. [CrossRef]

30. Didan, K. *MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006*; NASA EOSDIS Land Processes DAAC, 2015. Available online: https://lpdaac.usgs.gov/products/mod13q1v006/ (accessed on 24 March 2019).
31. Schnur, M.T.; Xie, H.; Wang, X. Estimating root zone soil moisture at distant sites using MODIS NDVI and EVI in a semi-arid region of southwestern USA. *Ecol. Inform.* **2010**, *5*, 400–409. [CrossRef]
32. Campbell, T.; Fearns, P. Honey crop estimation from space: Detection of large flowering events in Western Australian forests, in ISPRS TC I Mid-term Symposium "Innovative Sensing—From Sensors to Methods and Applications". In Proceedings of the 2018 International Society for Photogrammetry and Remote Sensing, Karlsruhe, Germany, 4–5 October 2018; pp. 79–86.
33. Gao, B.-C. NDWI A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water From Space. *Remote Sens. Environ.* **1996**, *58*, 257–266. [CrossRef]
34. Clay, D.E.; Kim, K.I.; Chang, J.; Clay, S.A.; Dalsted, K. Characterizing Water and Nitrogen Stress in Corn Using Remote Sensing. *Charact. Water Nitrogen Stress Corn Using Remote Sens.* **2006**, *98*, 579–587. [CrossRef]
35. Breiman, L. *Classification and Regression Trees*; Wadsworth International Group: Belmont, CA, USA, 1984.
36. Loh, W.-Y. Classification and Regression Trees. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2011**, *1*, 14–23. [CrossRef]
37. Elith, J.; Leathwick, J.R.; Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [CrossRef] [PubMed]
38. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
39. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Vanderplas, J. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
40. Leyland, D. *Review of Historic Marri Harvest Records*; Personnal communication: Chidlow, Australia, 2015.
41. Painter, S. *Jarrah Honey Crisis as Yield Wiped out, in The West Australian*; Seven West Media: Perth, Australia, 2010.
42. Andrys, J.; Kala, J.; Lyons, T. Regional climate projections of mean and extreme climate for the southwest of Western Australia (1970–1999 compared to 2030–2059). *Obs. Theor. Comput. Res. Clim. Syst.* **2017**, *48*, 1723–1747. [CrossRef]
43. Oldenborgh, G. Climate Explorer: Starting Point. Available online: http://climexp.knmi.nl/start.cgi (accessed on 4 April 2019).

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

7 Review and Discussion

7.1 Significant Findings

As detailed in Section 1.3 (Research Objectives and Significance), the objective of this thesis is to develop tools to map and predict *Corymbia calophylla* honey harvests, utilising readily available temporospatial input data. The staged research program detailed in Section 1.4 (Research Design and Thesis Outline) identifies the research steps required to achieve this outcome, which includes development of supporting outcomes to assist apiarists with assessing the honey producing potential of their apiary sites on a seasonal basis.

The objective of the initial stage of research was to determine whether the relative abundance of flowering *Corymbia calophylla* trees could be assessed by readily available UAV's. The usefulness of this to apiarists would be the ability to achieve a better understanding of nectar availability of apiary sites at the start of and during the flowering period. Due to mature *Corymbia calophylla* trees being between 10 – 40 m height (Brooker and Kleinig 2001), and honeybees readily foraging over a kilometre from their hive (Hagler et al. 2011), this information is difficult to determine accurately from the ground. Creating an accurate tool for relative flower abundance between apiary sites would improve placement of beehives on optimal apiary sites for a season and monitor the progress of the flowering season.

In Chapter 3 (Publication 1: Campbell and Fearns (2018b)), analysis of high-resolution multispectral images acquired for *Corymbia calophylla* flowers showed that the image pixels containing flowers could be separated from non-flower pixels to better than 90% accuracy using the simple parallelepiped classification method. This initial analysis was performed on images with over 50 pixels per flower. To determine the minimum resolution for accurate classification of flower pixels, the images were progressively scaled to lower resolution and the classification algorithm re-applied. It was found that a minimum of 10 pixels per flower is required to maintain the 90% accuracy cover.

This minimum resolution information was combined with the resolution and FOV of a typical 12 MP RGB drone sensor to calculate a maximum flight height of 83 m above the tree canopy to achieve the 10 pixels per flower resolution. As *Corymbia calophylla* canopy height is typically between 10 – 40 m for mature trees, the UAV survey design needs to consider the canopy height to determine the optimum UAV height above the ground for optimum image resolution versus spatial coverage given the field of view.

With a robust approach developed to assess flower abundance while on site, the second stage of the research was to determine whether relative flower abundance could be determined remotely via satellite data. In Chapter 4 (Publication 2: Campbell and Fearns (2018a)), field spectroradiometer measurement data are presented that show a clear spectral separation in the visual and UV spectral ranges between *Corymbia calophylla* flowers and other material typical of these forests, including leaves, groundcover and bare earth. Statistical analysis of this separation using ANOVA and JM Distance methods highlighted the 518 nm to 557 nm wavelength range as the spectral range of peak separation.

To determine the ability of multispectral satellite sensors to detect relative differences in flower abundance, the spectra recorded by the field spectroratiometer were convolved with the spectral response functions of nine different satellite sensors to create a synthetic dataset of multi-spectral data for each sensor. From a JM Distance analysis of the synthetic spectral response, the MODIS sensor was the only sensor with the capability of reliably spectrally separating flower and non-flower pixels.

As satellite multispectral images would be the result of spectral mixing of multiple classes, the synthetic spectral responses for different classes (flowers, leaves and ground) were combined in different ratios to create a database of synthetic pixels of different spectral mixtures. By adjusting these ratios and recording the change in the nett spectral response, it was determined that MODIS Band 4 was the most sensitive to flower abundance (able to detect flowers once they cover 25% of a pixel).

Visual inspection of the field spectroradiometer data showed that while *Corymbia calophylla* flowers had a higher reflectance in the visual bands compared to other classes, they have a lower reflectance than other classes in the UV region. By dividing the visual band by the UV band, both spectral separation aspects are highlighted. A repeat analysis of synthetic mixed pixels for satellite sensors with spectral bands in these regions showed that the MODIS, VIIRS and Sentinel-3 sensors are highly sensitive to flower abundance using this metric, all being able to detect a change in flower abundance of less than 2% of the pixel with this revised scale.

$$Marri\ Flower\ Index\ (MFI)\ = \frac{Visual\ Band\ Reflectance\ (\sim 550\ nm)}{Near-UV\ Band\ Reflectance\ (< 400\ nm)}$$

Equation 1 - Calculation of the Marri Flowering Index (MFI)

This result was tested with real-world data by comparing the scaled MODIS data against 14 years of Corymbia calophylla honey harvest data from an apiary site near Perth. An ANOVA analysis of the results proved that there is a statistically significant separation of the MODIS data for harvests above 20 kg per hive and below 20 kg per hive. Based on this result, a binary classification of the scaled MODIS data into harvests greater than 20 kg and less than 20 kg of honey per hive was able to achieve a 78% classification accuracy. It is therefore inferred that satellite data is capable of detecting differences in flower abundance, as measured via honey weight collected by honeybees from these flowers.

While the first two publications resulted in methods to detect flowers during the flowering period, the other objective of the research was to predict honey harvests prior to the flowering period commencing. With several previous studies indicating that satellite data may provide honey yield predictions but incorporation of weather data would improve the result, Publication 3 (Section 5) assessed this approach using a decision tree approach.

Annual *Corymbia calophylla* harvest yields from 16 different apiary sites over 9 years were used for this assessment, with sites ranging from less than 5% to greater than 50% mature canopy cover. Weather data, consisting of temperature and rainfall for the two months preceding the harvest (January and February), were extracted from the nearest BOM weather station and monthly NDVI and EVI values for these months were extracted from the MODIS datasets. The scaled MODIS data for flower abundance, as described above, was also extracted for each site.

Of the satellite datasets, the January NDVI showed the greatest difference between honey harvests of different weights. An ANOVA analysis of this proved that the January NDVI was statistically separate for good versus moderate to poor harvest years (greater than 40 kg of honey per hive versus less than 40 kg of honey per hive). ANOVA analysis of the scaled MODIS flower abundance measurement indicated that this metric was able to separate moderate versus good years but not other classification pairs. It was therefore concluded that the MODIS flower abundance metric was only applicable to sites with extensive, mature forest canopy.

Analysis of the weather data showed that three key parameters were linked to different harvest classifications; cooler January temperatures generally resulted in good harvests, higher January rainfall in moderate harvests and higher February rainfall in poor harvests, although each of these parameters did overlap across the three honey harvest

classifications. Overall, there was stronger clustering of good harvests and more overlap between moderate and poor harvests.

Assessing pairs of NDVI and weather data against honey harvest classifications improved the clustering of the data and a Decision Tree was developed to predict the honey harvest classification. This classification method was able to classify good harvests to 90% accuracy and moderate and poor quality harvests to 66% and 69% accuracy respectively.

To further develop this prediction model, Publication 4 (Campbell et al. (2020), presented in Chapter 6) used machine learning to develop the optimum prediction model based on a wider range of weather and satellite data from up to 11 months prior to harvest (effectively from the end of the previous year's harvest). This included minimum and maximum average temperatures and rainfall for monthly and cumulative periods and a range of MODIS products commonly used for agricultural crop monitoring (such as Gross Primary Production and Evapotranspiration products).

The 115-dimensional matrix built from these datasets was used as input to a range of Random Forest and Gradient Boosted trees for regression and classification algorithms to determine which algorithm performed best. Gradient Boosted Regression achieved the most accurate prediction result with a MAE of +/- 10.33 kg against a harvest range of 0 – 80 kg. The Gradient Boosted Classification algorithm was able to classify the harvest into the correct quality class 82% of the time.

Analysis of the partial dependence of the input factors to the gradient boosted algorithms showed that the average maximum temperature from October to January (from weather station data) and the January Evapotranspiration (from MODIS data) were the two input factors with the highest importance. Rerunning the GBR algorithm with only these two inputs achieved a MAE of +/- 10.42 kg and the same classification accuracy of 82%.

A more detailed investigation of the effect of the average maximum temperature for individual months between October to January highlighted that the average maximum temperature for November described the majority of the harvest variations. Re-running the GBR algorithm using the November temperature and January ET inputs achieved a MAE of +/- 8.92 and a classification accuracy of 83% (with a 100% accuracy for classifying good versus poor and moderate harvests).

While this predictive model provides a sufficient degree of accuracy for apiarists, the model relies on data from the end of January (when *Corymbia calophylla* typically starts to flower)

and as a result does not provide sufficient time for apiarists to adapt their seasonal plans once the model is ready. Running the GBR with only the November average maximum temperature produces a model with a MAE of +/- 11.72 kg, with a classification accuracy of 75%. While this model has a lower accuracy than the model incorporating the January ET, it does give apiarists several months to plan ahead for the *Corymbia calophylla* harvest, with progressively more accurate models able to be generated as more weather and satellite data becomes available closer to the harvest period.

7.2 Limitations of the Research

While the research completed and published in Chapters 3 through 6 has successfully achieved the research objectives detailed in Section 1.3 (Research Objectives and Significance), there are limitations to the research and outcomes for each publication.

For the UAV survey specifications to detect flowering *Corymbia calophylla* presented in Chapter 3, the research indicated that a UAV survey with a resolution of 10 pixels per flower could classify flowers to 90% accuracy using the Parallelepiped method. However, as discussed in the publication the misclassification of background (i.e. non-flower pixels) as flower pixels is the main source of error. This is particularly an issue in areas where direct sunlight on light coloured material results in the RGB image bands being close to saturation (i.e. nearly white or white). As the research did not include testing of UAV imagery over *Corymbia calophylla* forests, the practical outcomes of this limitation needs to be tested under UAV survey conditions.

The accuracy of UAV surveys to measure flower abundance is also limited by the distribution of flowers on *Corymbia calophylla* trees. As shown in Figure 2 of Chapter 3, *Corymbia calophylla* trees can have flowers from the top of the crown, down the canopy and almost to the ground in some cases. UAV surveys conducted at greater than 60 m height with a camera in nadir orientation would be unable to detect flowers on the sides of the tree at viewing angles close to 90 degrees. Flowers may be detected on one side of trees near the edges of the UAV imagery at lower viewing angles. Therefore the proposed survey design may be suitable to measure relative flower abundances but not absolute abundances.

Chapter 4 investigated the hyperspectral separation of *Corymbia calophylla* flowers from other objects common to these forests and showed a clear separation in the visual and UV portions of the spectrum. The resulting analysis of synthetic multispectral data from nine different satellite-borne sensors showed a clear advantage of the MODIS, VIIRS and

Sentinel-3 systems for *Corymbia calophylla* flower detection. With only the MODIS data tested against real-world honey harvest records, there remains potential for another two sensors to detect flowering in *Corymbia calophylla* trees.

The real-world comparison of honey harvest data with MODIS-derived flower abundance in Chapter 4 proved the ability of this data to detect differences in relative flower abundances between years for an apiary site in an area of mature *Corymbia calophylla* with a high degree of canopy cover. The flower abundance measurement ability has not been tested for apiary sites with a lower proportion of mature *Corymbia calophylla* canopy, whether this is due to more variable forest types, less mature forest or landscape fragmentation from land clearing for either agricultural or residential purposes.

The initial honey yield prediction model developed in Chapter 5 successfully used a combination of weather and satellite data to predict the quality of the honey harvest immediately prior to the commencement of flowering for *Corymbia calophylla* trees. While the honey harvest data used to create the prediction model was acquired over a large geographic area (over 400 km between the northern and southern apiary sites), the locations are clustered in three areas and therefore not spread evenly over the range. The range also does not cover the geographic range over which *Corymbia calophylla* trees occur. The applicability of the model may be improved through incorporation of honey harvest data from a wider, and more equitably distributed, geographic range of apiary sites. As noted in the publication, beehive health was not included in the model input factors. In addition, a wider range of input data may be used, both over longer timeframes and assessing the impact of more extreme weather events within the current two-month time period of the input data.

With the predictive model presented in Chapter 5 built on averaged datasets close to honey harvest time using a simple Decision Tree approach, there is scope to improve this based on both the data inputs and the model development approach. These are largely addressed in Chapter 6. In this publication, a longer timeframe was taken for the input dataset (up to 11 months prior to the season), extreme weather events were incorporated into the model and a more advanced and robust machine learning algorithm was utilised. However, as detailed in the publication, some limitations are still present.

With temperature and summer rainfall identified as key input factors in Chapters 5 and 6, the accuracy of these datasets is crucial in model development. The BOM terrestrial weather stations used for model development in Chapter 6 are not always proximal to the

apiary sites. Steep temperature gradients during heatwaves and highly localised, intense summer rainfall events in the region means that the weather station network may not accurately capture the true weather conditions of an apiary site. As with Chapter 5 , the geographic spread of the sites did not cover the full extent of *Corymbia calophylla* forests, nor was beehive health incorporated into the model, though commercial apiarists aim to maintain constant hive health through good husbandry.

7.3 Future Work and Recommendations

In addition to the potential to reduce the limitations discussed above in Section 7.2, there are several areas where further research may improve the existing model, or extend the applicability of the results. These are summarised as follows:

1. **Field testing of UAV and satellite flower abundance detection**

   The flower abundance detection approaches detailed in Chapters 3 and 4 have undergone limited field testing to date. Their efficacy can be verified by a combined UAV field survey program over a *Corymbia calophylla* flowering season, with regular UAV surveys conducted over several different apiary sites of contrasting forest maturity, land use types, etc.

   The UAV classification accuracy may be improved through more detailed data processing, such as image texture metrics, and/or detailed survey design to create a 3D point cloud for the survey area (allowing bright pixels at ground level to be filtered out).

   The timing of these surveys should be scheduled to coincide with flyover times for the selected satellites, Sentinel-3 and VIIRS datasets in conjunction with MODIS. The capabilities of sensors with higher spatial resolution, but sub-optimum spectral bands can also be tested (such as Sentinel-2). Based on the required canopy coverage required, these sensors may perform well in fragmented landscapes as pixels of high content can be extracted for analysis, whereas the lower spatial resolution sensors will rely on the mixed response from different land use types.

   With the spectral library created from the research, potentially spectral end-member analysis could be used to extract more detailed information on flower coverage per pixel of satellite data.

   This work would both verify the accuracy of the datasets and establish whether the satellite data is sufficiently sensitive to measure changes in relative flower abundance for a given apiary site over time (allowing apiarists to estimate nectar availability/honey production remotely during the season).

2. **Increased geographic distribution of input honey harvest data**

The honey harvest data used to create the current predictive model included apiary sites over a sizeable distance, from Dandgaragan (140 km north of Perth) to Boyup Brook (230 km south-east of Perth). These sites were clustered together in three main localities with a combined areal coverage of approximately 9,500 km^2, which is 11% of the full distribution of the *Corymbia calophylla* trees. These apiary sites are from only three of the six Interim Biogeographic Regionalisation for Australia (IBRA) subregions where *Corymbia calophylla* trees occur (Environment 2012) and three of DPIRD's seven climatic regions in which the *Corymbia calophylla* trees occur (Kingwell 2003).

With a number of ecosystems and growing regions under- or not represented, honey harvest data is required from a significantly larger area to extend the model and ensure it is accurately predicting honey harvests across the distribution of the *Corymbia calophylla* trees. Obtaining honey harvest data from all regions also gives the potential for creating predictive models for individual regions (if sufficient data is received to train the model) as there is some evidence for different growth rates and phenological responses for *Corymbia calophylla* trees in different climatic regions (Ahrens et al. 2019).

3. **Improved input of weather data into model**

As discussed in Section 7.2, a limiting factor in the current model is the relative sparsity of weather stations across the extent of *Corymbia calophylla* trees. Of the 84,000 km^2 of area where *Corymbia calophylla* trees occur, only 11,000 km^2 (i.e. 13%) is within 10 km of a BOM weather station.

There are several alternative sources of weather data that may be used, such as daily and monthly gridded products of temperature, rainfall and other parameters created from BOM weather stations (Jones et al. 2009) and the more geographically spread weather station operated by DPIRD. In addition, there are several sources of satellite-derived temperature and rainfall data.

The accuracy of these various data sources can be assessed across the different climate and/or biogeographic zones in which *Corymbia calophylla* trees occur and the optimum data source determined for each weather parameter for each apiary site.

4. **Impact of climate change models on honey production**

With weather being a key factor of *Corymbia calophylla* honey production, the effect of weather conditions as defined by the predictive model can be used alongside climate change projection models to determine the likely impact of climate change on medium to long-term honey harvest volumes. This has been discussed qualitatively in Chapter 6

but this analysis can be significantly extended to produce quantitative models. Outcomes from this process can be used as inputs to strategic planning for migratory beehive management (Patel et al. 2020a)

5. **Incorporate bee health into the model**

   As identified in Section 7.2 and Chapter 5, bee health has not been incorporated into the predictive model. This may be measures of adult bee population and brood, as per Delaplane et al. (2013) or quality of nutrition available to the hive, both leading to and during the honey flow (Brodschneider and Crailsheim 2010).

   An alternative approach is to predict hive activity based on weather forecasts, with temperature and solar radiation being key input factors for forage and scout bee activity levels (Clarke and Robert 2018).

6. **Adapt model to other honey producing species and regions**

   As the model development process for *Corymbia calophylla* utilises readily available weather and satellite data, predictive models can be readily created for other honey harvests. For Western Australia, this may be for *Eucalyptus marginata* for example, another high-grade antimicrobial honey (Irish et al. 2011), and the Banksia woodland north of Perth, which flowers in winter (Lee and Manning 1997) and is crucial for maintaining bee health and building hives prior to pollination of crops in spring (Manning 1995).

   However, the model development process need not be limited to Western Australia. In theory, the process may be applied to any honey producing region or species where sufficient honey harvest data are available to develop the model.

7.4 Concluding remarks

As discussed in Section 1.3 (Research Objectives and Significance), the primary research objective of this thesis was to develop a predictive model of marri honey harvest weight, utilising readily available temporospatial input data. This has been achieved by the research detailed in Chapters 3 to 6, with outcomes developed as detailed below:

- Publication 1 (Chapter 3) has established survey design parameters for detection of typical sized canopy tree flowers in the SWAFR with readily available UAV systems without requiring full CASA UAV certification. This allows apiarists to achieve a better measure of the extent of flowering at their apiary sites when inspecting them than can be achieved from the ground.

- Publication 2 (Chapter 4) proved the ability of readily available satellite data to map active honey producing areas via direct detection of flowering *Corymbia calophylla* trees in areas of high canopy cover, providing a method to potentially assess current nectar producing status remotely.
- Chapter 5 identified initial key factors from both weather and satellite data that precede the honey producing period for *Corymbia calophylla*, particularly in good harvest years. These factors were able to determine whether an apiary site would produce a good harvest to 90% accuracy immediately prior to flowering commencing.
- Publication 4 (Chapter 6) extended the research in Chapter 5 through the use of machine learning to create a model to predict the weight of honey harvested to 87% accuracy immediately prior to flowering and 85% accuracy two months prior to flowering.

The staged development of the research through the three peer reviewed publications explicitly meets the research objective set out in Section 1.3, with the predictive model using readily available spatiotemporal data. The process used to create the model detailed in Chapter 6 can readily be applied to honey produced from other species, or regions, providing the basis of a widely applicable method to predict honey harvest yields from locations with sufficient availability of historical harvest records.

8 Bibliography

Every reasonable effort has been made to acknowledge the owners of copyright material. I would bepleased to hear from any copyright owner who has been omitted or incorrectly acknowledged

Abburu S, Babu Golla S (2015) Satellite Image Classification Methods and Techniques: A Review International Journal of Computer Applications 119:20 - 25 doi:10.5120/21088-3779

Adam E, Mutanga O (2009) Spectral discrimination of papyrus vegetation (Cyperus papyrus L.) in swamp wetlands using field spectrometry ISPRS Journal of Photogrammetry and Remote Sensing 64:612-620 doi:10.1016/j.isprsjprs.2009.04.004

Ahrens CW et al. (2019) Adaptive variation for growth and resistance to a novel pathogen along climatic gradients in a foundation tree Evolutionary Applications 12:1178-1190 doi:10.1111/eva.12796

Aizen MA et al. (2019) Global agricultural productivity is threatened by increasing pollinator dependence without a parallel increase in crop diversification Global Change Biology 0 doi:10.1111/gcb.14736

Al-Ahmadi A, Subedi A, Wang G, Choudhary R, Fakhoury A (2018) Detection of charcoal rot (Macrophomina phaseolina) toxin effects in soybean (Glycine max) seedlings using hyperspectral spectroscopy Computers and Electronics in Agriculture 150:188 - 195 doi:10.1016/j.compag.2018.04.013

Al-Gaadi KA, Hassaballa AA, Tola E, Kayad AG, Madugundu R, Alblewi B, Assiri F (2016) Prediction of potato crop yield using precision agriculture techniques PloS one 11

Anderson JR (1976) A land use and land cover classification system for use with remote sensor data vol 964. US Government Printing Office,

Andrys J, Kala J, Lyons T (2017) Regional climate projections of mean and extreme climate for the southwest of Western Australia (1970–1999 compared to 2030–2059) Observational, Theoretical and Computational Research on the Climate System 48:1723-1747 doi:10.1007/s00382-016-3169-5

Aono Y, Kazui K (2008) Phenological data series of cherry tree flowering in Kyoto, Japan, and its application to reconstruction of springtime temperatures since the 9th century International Journal of Climatology 28:905-914 doi:10.1002/joc.1594

Arditti J, Elliott J, Kitching IJ, Wasserthal LT (2012) 'Good Heavens what insect can suck it'–Charles Darwin, Angraecum sesquipedale and Xanthopan morganii praedicta Botanical Journal of the Linnean Society 169:403-432 doi:10.1111/j.1095-8339.2012.01250.x

Arundel J, Winter S, Gui G, Keatley M (2016) A web-based application for beekeepers to visualise patterns of growth in floral resources using MODIS data Environmental Modelling & Software 83:116 - 125 doi:10.1016/j.envsoft.2016.05.010

Asner GP, Townsend AR, Braswell BH (2000) Satellite observation of El Nino effects on Amazon forest phenology and productivity Geophysical research letters 27:981-984

Atkinson PM, Tatnall ARL (1997) Introduction Neural networks in remote sensing International Journal of Remote Sensing 18:699-709 doi:10.1080/014311697218700

Bao S, Cao C, Chen W, Tian H (2017) Spectral features and separability of alpine wetland grass species Spectroscopy Letters 50:245 - 256 doi:10.1080/00387010.2016.1240088

Barbour L, White B, Boruff B, Narongsirikul M, Day C, Colgan-Zito E Fire impact on honey bee product production: an indicator of flora health. In: Prescribed Burning Conference 2019, Perth, Australia, 2019.

Barnes WL, Pagano TS, Salomonson VV (1998) Prelaunch characteristics of the moderate resolution imaging spectroradiometer (MODIS) on EOS-AMI IEEE Transactions on Geoscience and Remote Sensing 36:1088-1100 doi:10.1109/36.700993

Barrell J, Grant J (2015) High-resolution, low-altitude aerial photography in physical geography: A case study characterizing eelgrass (Zostera marina L.) and blue mussel (Mytilus edulis L.) landscape mosaic structure Progress in Physical Geography 39:440 - 459 doi:10.1177/0309133315578943

Beard J (1980) A new phytogeographic map of Western Australia vol 3. Perth, Australia

Beaubien J (1979) Forest type mapping from Landsat digital data Photogrammetric Engineering and Remote Sensing 45:1135-1144

Beck PSA, Atzberger C, Høgda KA, Johansen B, Skidmore AK (2006) Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI Remote Sensing of Environment 100:321-334 doi:https://doi.org/10.1016/j.rse.2005.10.021

Beeton R (2009) Advice to the Minister for the Environment, Heritage and the Arts from the Threatened Species Scientific Committee (the Committee) on Amendment to the list of Threatened Species

under the Environment Protection and Biodiversity Conservation Act 1999 (EPBC Act). Canberra, Australia

Belgiu M, Drăguţ L (2016) Random forest in remote sensing: A review of applications and future directions ISPRS Journal of Photogrammetry and Remote Sensing 114:24-31 doi:10.1016/j.isprsjprs.2016.01.011

Bischoff M, Lord J, Robertson A, Dyer A (2013) Hymenopteran pollinators as agents of selection on flower colour in the New Zealand mountains: salient chromatic signals enhance flower discrimination New Zealand Journal of Botany 51:181-193 doi:10.1080/0028825X.2013.806933

Blomstedt W (2014) Mapping The Phenology of European Honey Bee Nectar Flows. University of Edinburgh

Bolstad P, Lillesand T (1991) Rapid maximum likelihood classification Photogrammetric Engineering and Remote Sensing 57:67-74

Bravo LA, Ulloa N, Zuñiga GE, Casanova A, Corcuera LJ, Alberdi M (2001) Cold resistance in Antarctic angiosperms Physiologia Plantarum 111:55-65 doi:10.1034/j.1399-3054.2001.1110108.x

Breiman L (1984) Classification and regression trees / Leo Breiman ... [et al.]. Belmont, Calif : Wadsworth International Group, Belmont, Calif

Brodschneider R, Crailsheim K (2010) Nutrition and health in honey bees Apidologie 41:278-294 doi:10.1051/apido/2010012

Brooker MIH, Kleinig DA (2001) Field guide to eucalypts. vol Volme 2, South-western and southern Australia. Bloomings Books, Melbourne, Australia

Butt B, Turner MD, Singh A, Brottem L (2011) Use of MODIS NDVI to evaluate changing latitudinal gradients of rangeland phenology in Sudano-Sahelian West Africa Remote Sensing of Environment 115:3367-3376 doi:https://doi.org/10.1016/j.rse.2011.08.001

Camarda B (2007) How honey saved Bunny Australian Performance Horse April-May:40 - 42

Campbell T (2019) Ground-based spectroradiometer measurements of vegetation and groundcover of Corymbia calopylla forests in Western Australia. Curtin University. doi:10.25917/5d47b72fc294f

Campbell T, Dixon K, Dods K, Fearns P, Handcock R (2020) Machine Learning Regression Model for Predicting Honey Harvests Agriculture 10:118

Campbell T, Fearns P (2018a) Honey crop estimation from space: Detection of large flowering events in Western Australian forests. Paper presented at the ISPRS TC I Mid-term Symposium “Innovative Sensing – From Sensors to Methods and Applications", Karlsruhe, Germany,

Campbell T, Fearns P (2018b) Simple remote sensing detection of Corymbia calophylla Flowers using common 3 –band imaging sensors Remote Sensing Applications: Society and Environment 11:51-63 doi:10.1016/j.rsase.2018.04.009

Campbell T, Fearns P, Dods K, Dixon K (2019) Prediction and detection of honey harvests from remote sensing and weather data International Journal of Engineering Sciences & Research Technology 8:73 - 88

Carreiras JMB, Pereira JMC, Shimabukuro YE, Stroppiana D (2003) Evaluation of compositing algorithms over the Brazilian Amazon using SPOT-4 VEGETATION data International Journal of Remote Sensing 24:3427-3440 doi:10.1080/0143116021000021251

Carvalho S, Schlerf M, van der Putten W, Skidmore A (2013) Hyperspectral reflectance of leaves and flowers of an outbreak species discriminates season and successional stage of vegetation International Journal of Applied Earth Observation and Geoinformation 24:32-41 doi:10.1016/j.jag.2013.01.005

Chittka L, Shmida A, Troje N, Menzel R (1994) Ultraviolet as a component of flower reflections, and the colour perception of Hymenoptera Vision Research 34:1489-1508

Chittka L, Waser N (1997) Why red flowers are not invisible to bees Israel Journal of Plant Sciences 45:169-183 doi:10.1080/07929978.1997.10676682

Cho LH, Yoon J, An G (2017) The control of flowering time by environmental factors Plant Journal 90:708-719 doi:10.1111/tpj.13461

Christenhusz M, Byng J (2016) The number of known plant species in the world and its annual increase Phytotaxa 261:201-217 doi:10.11646/phytotaxa.261.3.1

Clarke D, Robert D (2018) Predictive modelling of honey bee foraging activity using local weather conditions Apidologie 49:386-396 doi:10.1007/s13592-018-0565-3

Clay DE, Kim KI, Chang J, Clay SA, Dalsted K (2006) Characterizing Water and Nitrogen Stress in Corn Using Remote Sensing Characterizing Water and Nitrogen Stress in Corn Using Remote Sensing 98:579-587

Clerici N, Weissteiner CJ, Gerard F (2012) Exploring the use of MODIS NDVI-based phenology indicators for classifying forest general habitat categories Remote Sensing 4:1781-1803

Cochrane MA (2000) Using vegetation reflectance variability for species level classification of hyperspectral data International Journal of Remote Sensing Applications 21:2075-2087 doi:10.1080/01431160050021303

Congalton RG (2010) Remote Sensing: An Overview GIScience & Remote Sensing 47:443-459 doi:10.2747/1548-1603.47.4.443

Cooper CE, Withers PC, Mawson PR, Bradshaw SD, Prince J, Roberston H (2002) Metabolic ecology of cockatoos in the south-west of Western Australia Australian Journal of Zoology 50:67-76 doi:10.1071/ZO00067

Cooper CE et al. (2003) Characteristics of Marri (Corymbia calophylla) fruits in relation to the foraging behaviour of the Forest Red-tailed Black Cockatoo (Calyptorhynchus banksii naso) Journal of the Royal Society of Western Australia 86:139-142

Cronquist A, Berry P, Zimmermann M, Dilcher DL, Stevens P, Stevenson D (2019) Angiosperm. Encyclopedia Britannica, Inc. https://www.britannica.com/plant/angiosperm. Accessed 26 April 2020

Dallon D (2003) Comparison of the Analytical Spectral Devices FieldSpec Pro JR and the Apogee/StellarNet Model SPEC-PAR/NIR Spectroradiometers. Utah State University, Utah

Darwin C (1859) On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. London : John Murray, 1859,

Davison EM, Tay FCS (1989) Phenology of Eucalyptus marginata on Sites Infested With Phytophthora cinnamomi Australian Journal of Botany 37:193-206 doi:10.1071/BT9890193

De Castro AI, Ehsani R, Ploetz R, Crane JH, Abdulridha J (2014) Optimum spectral and geometric parameters for early detection of laurel wilt disease in avocado Remote Sensing of Environment 171:33-44 doi:10.1016/j.rse.2015.09.011

De Groot AP (1952) Amino acid requirements for growth of the honeybee (Apis mellifica L.) Experientia 8:192-194 doi:10.1007/BF02173740

Delaplane KS, Van Der Steen J, Guzman-Novoa E (2013) Standard methods for estimating strength parameters of Apis mellifera colonies vol 52. Taylor & Francis. doi:10.3896/IBRA.1.52.1.03

Didan K (2015) MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 doi:10.5067/MODIS/MOD13Q1.006

Donohue RJ, Lawes RA, Mata G, Gobbett D, Ouzman J (2018) Towards a national, remote-sensing-based model for predicting field-scale crop yield Field Crops Research 227:79-90 doi:https://doi.org/10.1016/j.fcr.2018.08.005

Du J, He Z, Yang J, Chen L, Zhu X (2014) Detecting the effects of climate change on canopy phenology in coniferous forests in semi-arid mountain regions of China International Journal of Remote Sensing 35:6490-6507 doi:10.1080/01431161.2014.955146

Du Q (2018) 2.09 - Pixel Unmixing. In: Liang S (ed) Comprehensive Remote Sensing. Elsevier, Oxford, pp 186-198. doi:https://doi.org/10.1016/B978-0-12-409548-9.10337-9

Duchemin B et al. (2006) Monitoring wheat phenology and irrigation in Central Morocco: On the use of relationships between evapotranspiration, crops coefficients, leaf area index and remotely-sensed vegetation indices Agricultural Water Management 79:1-27 doi:https://doi.org/10.1016/j.agwat.2005.02.013

Duda T, Canty M (2002) Unsupervised classification of satellite imagery: Choosing a good algorithm International Journal of Remote Sensing 23:2193-2212 doi:10.1080/01431160110078467

Dyer A, Boyd-Gerny S, McLoughlin S, Rosa M, Simonov V, Wong B (2012) Parallel evolution of angiosperm colour signals: common evolutionary pressures linked to hynemopteran vision Proceedings of TheRoyal Society 279:3606-3615 doi:10.1098/rspb.2012.0827

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees Journal of Animal Ecology 77:802-813 doi:10.1111/j.1365-2656.2008.01390.x

Enderle D, Weih R (2005) Integrating Supervised and Unsupervised Classification Methods to Develop a More Accurate Land Cover Classification Journal of the Arkansas Academy of Science 59:65 - 73

Environment Dot (2012) Interim Biogeographic Regionalisation for Australia (Regions - States and Territories) v. 7 (IBRA). Canberra, Australia. doi:http://intspat01.ris.environment.gov.au/fed/catalog/search/resource/details.page?uuid=%7BFB89EEC9-5ABE-4CCD-B50E-7D485A3BAA4C%7D

Estes JE, Jensen JR, Tinney LR (1977) The Use of Historical Photography For Mapping Archaeological Sites Journal of Field Archaeology 4:441-447 doi:10.1179/009346977791490104

Fan L, Gao Y, Brück H, Bernhofer C (2009) Investigating the relationship between NDVI and LAI in semi-arid grassland in Inner Mongolia using in-situ measurements Theoretical and Applied Climatology 95:151-156 doi:10.1007/s00704-007-0369-2

Feilhauer H, Asner G, Martin R, Schmidtlein S (2010) Brightness-normalized Partial Least Squares Regression for hyperspectral data Journal of Quantitative Spectroscopy & Radiative Transfer 111:1947-1957 doi:10.1016/j.jqsrt.2010.03.007

Fenster CB, Armbruster WS, Wilson P, Dudash MR, Thomson JD (2004) Pollination syndromes and floral specialization Pollination syndromes and floral specialization 35:375-403

Fitter AH, Fitter RSR, Harris ITB, Williamson MH (1995) Relationships Between First Flowering Date and Temperature in the Flora of a Locality in Central England Functional Ecology 9:55-60 doi:10.2307/2390090

Fleming R (2018) Identification of chewed Marri nuts by cockatoos and parrots. Perh, Australia

Fuentes S, Tongson EJ, De Bei R, Gonzalez Viejo C, Ristic R, Tyerman S, Wilkinson K (2019) Non-invasive tools to detect smoke contamination in grapevine canopies, berries and wine: A remote sensing and machine learning modeling approach Sensors 19:3335

Gallai N, Salles J-M, Settele J, Vaissière BE (2009) Economic valuation of the vulnerability of world agriculture confronted with pollinator decline Ecological Economics 68:810-821 doi:10.1016/j.ecolecon.2008.06.014

Gandhi GM, Parthiban S, Thummalu N, Christy A (2015) Ndvi: Vegetation Change Detection Using Remote Sensing and Gis – A Case Study of Vellore District Procedia Computer Science 57:1199-1210 doi:10.1016/j.procs.2015.07.415

Gao B-C (1996) NDWI A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water From Space Remote Sensing of Environment 58:257 - 266

Gao J et al. (2019) Modeling alpine grassland forage phosphorus based on hyperspectral remote sensing and a multi-factor machine learning algorithm in the east of Tibetan Plateau, China ISPRS Journal of Photogrammetry and Remote Sensing 147:104-117 doi:https://doi.org/10.1016/j.isprsjprs.2018.11.015

Gill RA (1989) An economic evaluation of alternative management practices and enterprise structures in the Australian beekeeping industry. Dept. of Agricultural Economics and Business Management, University of New England, Armidale, N.S.W.

Gitelson AA, Merzlyak MN (1997) Remote estimation of chlorophyll content in higher plant leaves International Journal of Remote Sensing 18:2691-2697 doi:10.1080/014311697217558

Gole C (2006) The Southwest Australia Ecoregion - Jewel of the Australian Continent. Western Australia

Google Earth (2016).

Goward SN, Markham B, Dye DG, Dulaney W, Yang J (1991) Normalized difference vegetation index measurements from the advanced very high resolution radiometer Remote Sensing of Environment 35:257-277 doi:https://doi.org/10.1016/0034-4257(91)90017-Z

Grainger J (1939) STUDIES UPON THE TIME OF FLOWERING OF PLANTS: ANATOMICAL, FLORISTIC AND PHENOLOGICAL ASPECTS OP THE PROBLEM Annals of Applied Biology 26:684-704 doi:10.1111/j.1744-7348.1939.tb06994.x

Gross J, Heumann B (2014) Can flowers provide better spectral discrimination between herbaceous wetland species than leaves? Remote Sensing Letters 5:892-901 doi:10.1080/2150704X.2014.973077

Guerschman JP, Oyarzabal M, Malthus TJ, McVicar TR, Byrne G, Randall LA, Stewart JB (2012) Evaluation of the MODIS-based vegetation fractional cover product. New South Wales, Australia

Gutman G, Byrnes R, Masek J, Covington S (2008) Towards monitoring land-cover and land-use changes at a global scale: the global land survey 2005 Photogramm Eng Remote Sens 74

Gutman G, Huang C, Chander G, Noojipady P, Masek JG (2013) Assessment of the NASA–USGS Global Land Survey (GLS) datasets Remote Sensing of Environment 134:249-265 doi:https://doi.org/10.1016/j.rse.2013.02.026

Hadjimitsis DG et al. (2010) Atmospheric correction for satellite remotely sensed data intended for agricultural applications: impact on vegetation indices Nat Hazards Earth Syst Sci 10:89-95

Hagler J, Mueller S, Teuber L, Machtley S, Van Deynze A (2011) Foraging range of honey bees, Apis mellifera, in alfalfa seed production fields Journal of Insect Science 11

Han L et al. (2019) Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data Plant Methods 15:10 doi:10.1186/s13007-019-0394-z

Hastie H (2018) Sweet as: The rise and rise of West Australian honey. Perth, Western Australia

Hastono T, Santoso AJ, Pranowo (2017) Honey yield prediction using Tsukamoto fuzzy inference system vol 2017-. doi:10.1109/EECSI.2017.8239150

Hawkins B, Thomson J, Mac Nally R (2018) Regional patterns of nectar availability in subtropical eastern Australia Landscape Ecology 33:999-1012 doi:10.1007/s10980-018-0647-7

Haydak MH (1970) Honey Bee Nutrition Annual Review of Entomology 15:143-156 doi:10.1146/annurev.en.15.010170.001043

Herbarium WA (1998) Florabase - the Western Australian Flora.

Herbarium WA (2015) Florabase. Department of Parks and Wildlife. https://florabase.dpaw.wa.gov.au/. Accessed 22 April 2015 2015

Herman BM, Caudill TR, Flittner DE, Thome KJ, Ben-David A (1995) Comparison of the Gauss-Seidel spherical polarized radiative transfer code with other radiative transfer codes Applied optics 34:4563-4572 doi:10.1364/ao.34.004563

Herold M, Roberts DA, Gardner ME, Dennison PE (2004) Spectrometry for urban area remote sensing—Development and analysis of a spectral library from 350 to 2400 nm Remote Sensing of Environment 91:304-319 doi:https://doi.org/10.1016/j.rse.2004.02.013

Hird JN, McDermid GJ (2009) Noise reduction of NDVI time series: An empirical comparison of selected techniques Remote Sensing of Environment 113:248-258 doi:https://doi.org/10.1016/j.rse.2008.09.003

Hochman Z, Horan H Graincast: near real time wheat yield forecasts for Australian growers and service providers. In: Proceedings of the 2019 Agronomy Australia Conference, Wagga Wagga, Australia, 25 - 29 August 2019 2019.

Hogda KA, Karlsen SR, Solheim I Climatic change impact on growing season in Fennoscandia studied by a time series of NOAA AVHRR NDVI data. In: IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217), 2001. IEEE, pp 1338-1340

Holzworth DP et al. (2014) APSIM – Evolution towards a new generation of agricultural systems simulation Environmental Modelling & Software 62:327-350 doi:https://doi.org/10.1016/j.envsoft.2014.07.009

Honkavaara E, Khoramshahi E (2018) Radiometric Correction of Close-Range Spectral Image Blocks Captured Using an Unmanned Aerial Vehicle with a Radiometric Block Adjustment Remote Sensing 10:256 doi:10.3390/rs10020256

Hopper SD, Brown AP (2007) A revision of Australia' s hammer orchids (Drakaea: Orchidaceae), with some field data on species-specific sexually deceived wasp pollinators Australian Systematic Botany 20:252 doi:10.1071/SB06033

Hu S, Dilcher DL, Jarzen DM, Winship Taylor D (2008) Early steps of angiosperm pollinator coevolution Proceedings of the National Academy of Sciences of the United States of America 105:240-245 doi:10.1073/pnas.0707989105

Hudson IL, Kim S, Keatley M (2009) Climatic influences on the flowering phenology of four Eucalypts: a GAMLSS approach. Paper presented at the The 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Cairns, Australia, 13 - 17 July 2009

Huete A, Justice C, Van Leeuwen W (1999) MODIS vegetation index (MOD13) Algorithm theoretical basis document 3

Huete AR (1988) A soil-adjusted vegetation index (SAVI) Remote Sensing of Environment 25:295-309 doi:https://doi.org/10.1016/0034-4257(88)90106-X

Huntsdale J (2019) Honey producers hand-feed bees during drought to save hives, with sting likely for consumers. ABC News, Illawarra

Iqbal F, Lucieer A, Barry K (2018) Simplified radiometric calibration for UAS-mounted multispectral sensor European Journal of Remote Sensing 51:301-313 doi:10.1080/22797254.2018.1432293

Irish J (2006) Honey has an antifungal effect against Candida species Medical Mycology 44:289-291 doi:10.1080/13693780500417037

Irish J, Blair S, Carter D (2011) The Antibacterial Activity of Honey Derived from Australian Flora PLoS ONE 6 doi:10.1371/ journal.pone.0018229

Jamali S, Jönsson P, Eklundh L, Ardö J, Seaquist J (2015) Detecting changes in vegetation trends using time series segmentation Remote Sensing of Environment 156:182-195 doi:10.1016/j.rse.2014.09.010

Jiao W, Zhang L, Chang Q, Fu D, Cen Y, Tong Q (2016) Evaluating an enhanced vegetation condition index (VCI) based on VIUPD for drought monitoring in the continental United States Remote Sensing 8:224

Johnstone RE, Kirkby T (2008) Distribution, status, social organisation, movements and conservation of Baudin's Cockatoo (Calyptorhynchus baudinii) in South-west Western Australia Records of the Western Australian Museum 25:107 - 118 doi:10.18195/issn.0312-3162.25(1).2008.107-118

Jones D, Wang W, Fawcett R (2009) High-quality spatial climate data-sets for Australia Australian Meteorological and Oceanographic Journal 58:233-248 doi:10.22499/2.5804.003

Jönsson P, Eklundh L (2002) Seasonality extraction by function fitting to time-series of satellite sensor data, IEEE T. Geosci. Remote, 40, 1824–1832.

Jönsson P, Eklundh L (2004) TIMESAT—a program for analyzing time-series of satellite sensor data Computers & Geosciences 30:833-845 doi:https://doi.org/10.1016/j.cageo.2004.05.006

Karasinski J (2018) The economic valuation of Australian managed and wild honey bee pollinators. Curtin University, Perth, W.A.

Kerdiles H, Rembold F, Leo O, Boogaard H, Hoek S CST, a freeware for predicting crop yield from remote sensing or crop model indicators: Illustration with RSA and Ethiopia. In: 2017 6th International Conference on Agro-Geoinformatics, 7-10 Aug. 2017 2017. pp 1-6. doi:10.1109/Agro-Geoinformatics.2017.8047071

Key C, Benson N (2006) Landscape Assessment: Ground measure of severity, the Composite Burn Index; and Remote sensing of severity, the Normalized Burn Ratio. In. pp LA 1-51

Khoram S, Koch F, Wiele Cv, Nelson S (2012) Remote Sensing. Springer. doi:10.1007/978-1-4614-3103-9

Kimes DS, Holben BN, Tucker CJ, Newcomb WW (1984) Optimal directional view angles for remote-sensing missions International Journal of Remote Sensing 5:887-908 doi:10.1080/01431168408948876

Kingwell RS (2003) Quality Assurance Certification and Implementation: Growers' Costs and Perceived Benefits

Kisevic M et al. (2011) Spectral reflectance profile of Caulerpa racemosa var. cylindracea and Caulerpa taxifola in the Adriatic Sea Acta Adriatica 52:21-28

Klein A-M, Vaissière BE, Cane JH, Steffan-Dewenter I, Cunningham SA, Kremen C, Tscharntke T (2007) Importance of pollinators in changing landscapes for world crops Proceedings Biological sciences 274:303-313 doi:10.1098/rspb.2006.3721

Kogan FN (1995) Application of vegetation index and brightness temperature for drought detection Advances in Space Research 15:91-100 doi:https://doi.org/10.1016/0273-1177(95)00079-T

Kratz M et al. Essential amino acid composition of emerging bees fed different food sources. In: Bee Industry Council of Western Australia 2019 Industry Conference, Perth, Western Australia, 2019.

Kremen C et al. (2007) Pollination and other ecosystem services produced by mobile organisms: a conceptual framework for the effects of land-use change vol 10. Oxford, UK. doi:10.1111/j.1461-0248.2007.01018.x

Kross A, McNairn H, Lapen D, Sunohara M, Champagne C (2015) Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops International Journal of Applied Earth Observation and Geoinformation 34:235-248 doi:https://doi.org/10.1016/j.jag.2014.08.002

Kyalo Kiema JB (2013) Chapter 6 - Remote Sensing Application Supporting IWRM in Kenya. In: Paron P, Olago DO, Omuto CT (eds) Developments in Earth Surface Processes, vol 16. Elsevier, pp 51-66. doi:https://doi.org/10.1016/B978-0-444-59559-1.00006-2

Lancashire PD, Bleiholder H, Boom TVD, Langelüddeke P, Stauss R, Weber E, Witzenberger A (1991) A uniform decimal code for growth stages of crops and weeds Annals of Applied Biology 119:561-601 doi:10.1111/j.1744-7348.1991.tb04895.x

Landmann T, Piiroinen R, Makori D, Abdel-Rahman E, Makau S, Pellikka P, Raina S (2015) Application of hyperspectral remote sensing for flower mapping in African savannas Remote Sensing of Environment 166:50 - 60 doi:10.1016/j.rse.2015.06.006

Lary DJ, Alavi AH, Gandomi AH, Walker AL (2016) Machine learning in geosciences and remote sensing Geoscience Frontiers 7:3-10 doi:https://doi.org/10.1016/j.gsf.2015.07.003

Last F (2001) Keeping Records The Horticulturalist 10:9 - 11

Lautenbach S, Seppelt R, Liebscher J, Dormann CF (2012) Spatial and Temporal Trends of Global Pollination Benefit (Trends of Global Pollination Benefit) PLoS ONE 7:e35954 doi:10.1371/journal.pone.0035954

Lawes R et al. Monitoring Grain Production across the Australian Continent In: 2019 Agronomy Australia Conference, Wagga Wagga, Australia, 25 - 29 August 2019 2019.

Lee A, Manning R (1997) Beekeeping In Western Australia vol 9-1997, 4344 edn. Department of Agriculture and Food, Western Australia, Perth, Australia

Leyland D (2015) Review of historic marri harvest records. Personnal communication, Chidlow, Western Australia

Liang S (2012) Chapter 1 - A Systematic View of Remote Sensing. In: Liang S, Li X, Wang J (eds) Advanced Remote Sensing. Academic Press, Boston, pp 1-31. doi:https://doi.org/10.1016/B978-0-12-385954-9.00001-0

Lieutier F et al. (2017) Chapter Three - From Plant Exploitation to Mutualism. In: Sauvion N, Thiéry D, Calatayud P-A (eds) Advances in Botanical Research, vol 81. Academic Press, pp 55-109. doi:https://doi.org/10.1016/bs.abr.2016.10.001

Liu Z-J, Wang X (2016) A perfect flower from the Jurassic of China Historical Biology 28:707-719 doi:10.1080/08912963.2015.1020423

Loh W-Y (2011) Classification and regression trees Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1:14-23 doi:10.1002/widm.8

Long W, Srihann S Land cover classification of SSC image: unsupervised and supervised classification using ERDAS Imagine. In: IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium, 20-24 Sept. 2004 2004. pp 2707-2712 vol.2704. doi:10.1109/IGARSS.2004.1369859

Lu Q, Tang M (2011) Detection of Hidden Bruise on Kiwi fruit Using Hyperspectral Imaging and Parallelepiped Classification. Paper presented at the International Conference on Environmental Science and Engineering, Indonesia,

Lughadha E et al. (2016) Counting counts: revised estimates of numbers of accepted species of flowering plants, seed plants, vascular plants and land plants with a review of other recent estimates Phytotaxa 272:82 - 88 doi:10.11646/phytotaxa.272.1.5

Lynn BC (2013) Relation of honey production in Apius Mellifera colonies to the NDVI and other indicators. University of North Carolina

Magiera A, Feilhauer H, Otte A, Waldhardt R, Simmering D (2013) Relating canopy reflectance to the vegetation composition of mountainous grasslands in the Greater Caucasus Agriculture, Ecosystems and Environment 177:101-112 doi:10.1016/j.agee.2013.05.017

Malila WA, Lambeck PF, Crist EP, Jackson RD, Pinter Jr PJ (1980) Landsat features for agricultural applications

Manning R (1995) Honeybee pollination, technical data for potential honey-bee pollinated crops and orchards in Western Australia vol 5-1995, 4298 edn. Department of Agriculture and Food, Western Australia, Perth, Australia

Manning R (2008) Pollen from Western Australia at a glance vol 9-2008. Department of Agriculture and Food, Western Australia, Perth, Australia

Manning R (2011) Research into Western Australian honeys vol 9-2011. South Perth, Western Australia

Margono B et al. (2012) Mapping and monitoring deforestation and forest degradation in Sumatra (Indonesia) using Landsat time series data sets from 1990 to 2010 Environmental Research Letters 7:1-16 doi:10.1088/1748-9326/7/3/034010

Maxwell AE, Warner TA, Fang F (2018) Implementation of machine-learning classification in remote sensing: an applied review International Journal of Remote Sensing 39:2784-2817 doi:10.1080/01431161.2018.1433343

Menzel A et al. (2006) European phenological response to climate change matches the warming pattern Global Change Biology 12:1969-1976 doi:10.1111/j.1365-2486.2006.01193.x

Meteorology Bo (2000) Australian Data Archive for Meteorology. Paper presented at the Conference on Managing Australian Climate Variability, Albury, NSW, Australia,

Miranda RdQ, Galvíncio JD, Moura MSBd, Jones CA, Srinivasan R (2017) Reliability of MODIS Evapotranspiration Products for Heterogeneous Dry Forest: A Study Case of Caatinga Advances in Meteorology 2017 doi:10.1155/2017/9314801

Miyoshi GT et al. (2020) A Novel Deep Learning Method to Identify Single Tree Species in UAV-Based Hyperspectral Images Remote Sensing 12:1294

Monteith JL (1972) Solar Radiation and Productivity in Tropical Ecosystems Journal of Applied Ecology 9:747-766 doi:10.2307/2401901

Morris JL et al. (2018) The timescale of early land plant evolution Proceedings of the National Academy of Sciences 115:E2274-E2283 doi:10.1073/pnas.1719588115

Moulin S, Kergoat L, Viovy N, Dedieu G (1997) Global-scale assessment of vegetation phenology using NOAA/AVHRR satellite measurements Journal of Climate 10:1154-1170

Mouradov A, Cremer F, Coupland G (2002) Control of Flowering Time The Plant Cell 14:S111 doi:10.1105/tpc.001362

MSS Standard Interface Document (1978). General Electric Company, Beltsville, USA

Mu Q, Heinsch FA, Zhao M, Running SW (2007) Development of a global evapotranspiration algorithm based on MODIS and global meteorology data Remote Sensing of Environment 111:519-536 doi:10.1016/j.rse.2007.04.015

Munyati C (2004) Use of Principal Component Analysis (PCA) of Remote Sensing Images in Wetland Change Detection on the Kafue Flats, Zambia Geocarto International 19:11-22 doi:10.1080/10106040408542313

Myers N, Mittermeier C, Mittermeier R, Fonseca G, Kent J (2000) Biodiversity hotspots for conservation priorities Nature 403:853-858

Painter S (2010) Jarrah honey crisis as yield wiped out. Perth

Pal M, Mather PM (2003) An assessment of the effectiveness of decision tree methods for land cover classification Remote sensing of environment 86:554-565

Patel V, Biggs EM, Pauli N, Boruff B (2020a) Using a social-ecological system approach to enhance understanding of structural interconnectivities within the beekeeping industry for sustainable decision making Ecology and Society 25:24 doi:10.5751/ES-11639-250224

Patel V, Pauli N, Biggs E, Barbour L, Boruff B (2020b) Why bees are critical for achieving sustainable development Ambio doi:10.1007/s13280-020-01333-9

Pedregosa F et al. (2011) Scikit-learn: Machine Learning in Python J Mach Learn Res 12:2825-2830

Peel MC, Finlayson BL, Mcmahon TA (2007) Updated world map of the Köppen-Geiger climate classification Hydrology and Earth System Sciences Discussions 4:439-473

Perumal K, Bhaskaran R (2010) Supervised Classification Performance of Multispectral Images Journal of Computing 2:124 - 129

Petanidou T, Kallimanis AS, Sgardelis SP, Mazaris AD, Pantis JD, Waser NM (2014) Variable flowering phenology and pollinator use in a community suggest future phenological mismatch Acta Oecologica 59:104-111 doi:10.1016/j.actao.2014.06.001

Plaza A, Martinez P, Perez R, Plaza J (2004) A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data IEEE Transactions on Geoscience and Remote Sensing 42:650-663 doi:10.1109/TGRS.2003.820314

Portigal F, Holasek R, Mooradian G, Owensby P, Dicksion M, Fene M, Driggett D (1997a) Vegetation classification using red-edge first derivative and green peak statistical moment indices with the Advanced Airborne Hyperspectral Imaging System (AAHIS). Paper presented at the Third International Airborne Remote Sensing Conference and Exhibition, Copenhagen, Denmark,

Portigal F et al. (1997b) Vegetation classification using red-edge first derivative and green peak statistical moment indices with the Advanced Airborne Hyperspectral Imaging System (AAHIS). Paper presented at the Third International Airborne Remote Sensing Conference and Exhibition, Copenhagen, Denmark,

Potts SG et al. (2016) Assessment Report on Pollinators, Pollination and Food Production.

Prasad AK, Chai L, Singh RP, Kafatos M (2006) Crop yield estimation model for Iowa using remote sensing and surface parameters International Journal of Applied Earth Observation and Geoinformation 8:26-33 doi:https://doi.org/10.1016/j.jag.2005.06.002

Price J (1994) How unique are spectral signatures? Remote Sensing of Environment 49:181-185

Proffitt T (2019) Effect of poor marri flowering on viticulture, Discussion followed by email edn., Margaret River, Australia

Pu R (2009) Broadleaf species recognition with in situ hyperspectral data International Journal of Remote Sensing 30:2759-2779 doi:10.1080/01431160802555820

Quintano C, Fernández-Manso A, Shimabukuro YE, Pereira G (2012) Spectral unmixing International Journal of Remote Sensing 33:5307-5340 doi:10.1080/01431161.2012.661095

Rees A (1966) The physiology of ornamental bulbous plants The Botanical Review 32:1-23 doi:10.1007/BF02858583

Remotely piloted aircraft systems – operation of excluded RPA (other than model aircraft) (2016), v1.2 edn.,

Richards JA (2013) Remote sensing digital image analysis: An introduction vol 9783642300622. doi:10.1007/978-3-642-30062-2

Richards JA, Jia X (1999) The Interpretation of Digital Image Data. In: Richards JA, Jia X (eds) Remote Sensing Digital Image Analysis: An Introduction. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 75-88. doi:10.1007/978-3-662-03978-6_3

Rocha H, Dias J Honey Yield Forecast Using Radial Basis Functions. In: Nicosia G, Pardalos P, Giuffrida G, Umeton R (eds) Machine Learning, Optimization, and Big Data, Cham, 2018// 2018. Springer International Publishing, pp 483-495

Roshan N, Rippers T, Locher C, Hammer K (2017) Antibacterial activity and chemical characteristics of several Western Australian honeys compared to manuka honey and pasture honey Archives of Microbiology 199:347-355 doi:10.1007/s00203-016-1308-3

Rouse J, Haas R, Schell J, Deering D (1974) Monitoring vegetation systems in the Great Plains with ERTS NASA special publication 351:309

Running SW, Mu Q, Zhao M (2015) MOD17A2H MODIS/Terra Gross Primary Productivity 8-Day L4 Global 500m SIN Grid V006. doi:10.5067/MODIS/MOD17A2H.006

Running SW, Mu Q, Zhao M (2017) MOD16A2 MODIS/Terra Net Evapotranspiration 8-Day L4 Global 500m SIN Grid V006. doi:10.5067/MODIS/MOD16A2.006

Running SW, Nemani RR, Heinsch FA, Zhao M, Reeves M, Hashimoto H (2004) A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production BioScience 54:547-560 doi:10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO

Sagili R, Pankiw T (2007) Effects of protein-constrained brood food on honey bee (Apis mellifera L.) pollen foraging and colony growth Behavioral Ecology and Sociobiology 61:1471-1478 doi:10.1007/s00265-007-0379-1

Scarth P (2009) Vegetation height and structure - derived from ALOS-1 PALSAR, Landsat and ICESat/GLAS, Australia coverage. Brisbane, Australia. doi:http://dx.doi.org/10.4227/05/5703458340442

Schmidt KS, Skidmore AK (2003) Spectral discrimination of vegetation types in a coastal wetland Remote Sensing of Environment 85:92-108 doi:10.1016/S0034-4257(02)00196-7

Schnur MT, Xie H, Wang X (2010) Estimating root zone soil moisture at distant sites using MODIS NDVI and EVI in a semi-arid region of southwestern USA Ecological Informatics 5:400-409 doi:10.1016/j.ecoinf.2010.05.001

Schowengerdt RA (2006) Remote Sensing : Models and Methods for Image Processing. Elsevier Science & Technology, Burlington, UNITED STATES

Seiler RA, Kogan F, Wei G (2000) Monitoring weather impact and crop yield from NOAA AVHRR data in Argentina Advances in Space Research 26:1177-1185 doi:https://doi.org/10.1016/S0273-1177(99)01144-8

Sellers PJ, Tucker CJ, Collatz GJ, Los SO, Justice CO, Dazlich DA, Randall DA (1994) A global 1° by 1° NDVI data set for climate studies. Part 2: The generation of global fields of terrestrial biophysical parameters from the NDVI International Journal of Remote Sensing 15:3519-3545 doi:10.1080/01431169408954343

Shrestha M, Dyer A, Boyd-Gerny S, Wong B, Burd M (2013) Shades of red: Bird-pollinated flowers target the specific colour discrimination abilities of avian vision New Phytologist 198:301-310 doi: 10.1111/nph.12135

Singh RP, Roy S, Kogan F (2003) Vegetation and temperature condition indices from NOAA AVHRR data for drought monitoring over India International Journal of Remote Sensing 24:4393-4402 doi:10.1080/0143116031000084323

Somerville D (2005) Fat bees skinny bees : a manual on honey bee nutrition for beekeepers : a report for the Rural Industries Research and Development Corporation / by Doug Somerville. Manual on honey bee nutrition for beekeepers. Barton, A.C.T. : Rural Industries Research and Development Corporation, Barton, A.C.T.

Souza C et al. (2013) Ten-Year Landsat Classification of Deforestation and Forest Degradation in the Brazilian Amazon Remote Sensing 5:5493-5513 doi:10.3390/rs5115493

Sparks T, Lines J (2008) Chapters in the life of Robert Marsham Esq.

FRS of Stratton Strawless. St. Margaret's Church, Norfolk, United Kingdom

Strahler A (1981) Stratification of natural vegetation for forest and rangeland inventory using Landsat digital imagery and collateral data International Journal of Remote Sensing 2:15-41

Strahler AH (1980) The use of prior probabilities in maximum likelihood classification of remotely sensed data Remote sensing of Environment 10:135-163

Sulik J, Long D (2015) Spectral indices for yellow canola flowers International Journal of Remote Sensing 36:2751-2765 doi:10.1080/01431161.2015.1047994

Sulik J, Long D (2016) Spectral considerations for modeling yield of canola Remote Sensing of Environment 184:161 - 174 doi:10.1016/j.rse.2016.06.016

Swain PH, Davis SM (1981) Remote sensing: The quantitative approach IEEE Transactions on Pattern Analysis & Machine Intelligence:713-714

Testa S, Soudani K, Boschetti L, Borgogno Mondino E (2018) MODIS-derived EVI, NDVI and WDRVI time series to estimate phenological metrics in French deciduous forests International Journal of Applied Earth Observation and Geoinformation 64:132-144 doi:https://doi.org/10.1016/j.jag.2017.08.006

Thenkabail P, Enclona E, Ashton M, Van Der Meer B (2004) Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications Remote Sensing of Environment 91:354-376

Thomson J (2019) Western Australia a sweet spot for beekeeping. Department of Primary Industries and Regional Development,

Thorp KR, Wang G, West AL, Moran MS, Bronson KF, White JW, Mon J (2012) Estimating crop biophysical properties from remote sensing data by inverting linked radiative transfer and ecophysiological models Remote Sensing of Environment 124:224-233 doi:https://doi.org/10.1016/j.rse.2012.05.013

Tooke F, Battey NH (2010) Temperate flowering phenology Journal of Experimental Botany 61:2853-2862 doi:10.1093/jxb/erq165

Turner DP et al. (2006) Evaluation of MODIS NPP and GPP products across multiple biomes Remote Sensing of Environment 102:282-292 doi:10.1016/j.rse.2006.02.017

Udall S (1966) Earth's Resources to be Studied From Space. Office of the Secretary,

Ullah S, Schlerf M, Skidmore A, Hecker C (2012) Identifying plant species using mid-wave infrared (2.5–6 µm) and thermal infrared (8–14 µm) emissivity spectra Remote Sensing of Environment 118:95 - 102 doi:10.1016/j.rse.2011.11.008

Upreti D et al. (2019) A comparison of hybrid machine learning algorithms for the retrieval of wheat biophysical variables from sentinel-2 Remote Sensing 11:481

Vaiphasa C, Ongsomwang S, Vaiphasa T, Skidmore AK (2005) Tropical mangrove species discrimination using hyperspectral data: A laboratory study Estaurine Coastal and Shelf Science 65:371-379 doi:10.1016/j.ecss.2005.06.014

Vaissière BE Honey bee stocking rate, pollinator visitation, and pollination effectiveness in upland cotton grown for hybrid seed production. In: ISHS Acta Horticulturae 288: VI International Symposium on Pollination, Tilburg, Netherlands, 1991. International Society for Horticultural Science (ISHS), Leuven, Belgium, pp 359-363. doi:10.17660/ActaHortic.1991.288.58

Value of Agricultural Commodities Produced, Australia, 2014-15 (2016). Australian Bureau of Statistics, Canberra, Australia

Van Dijk A, Callis SL, Sakamoto CM, Decker WL (1987) Smoothing vegetation index profiles: An alternative method for reducing radiometric disturbance in NOAA/AVHRR data Photogrammetric Engineering and Remote Sensing 53:1059-1067

van Dijk J, Gomboso J, Levantis C (2016) Australian honey bee industry: 2014–15 survey results vol 16.18. Department of Agriculture and Water Resources, Canberra, Australia

Vlad V, Stefan V, Ion V, Ion N, DM M (2008) Researches regarding the prognosis of the flowering stage and honey potential yield of sunflower crops Bulletin of University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca Agriculture 65

Von Holle B, Wei Y, Nickerson D (2010) Climatic Variability Leads to Later Seasonal Flowering of Floridian Plants (Delayed Subtropical Flowering) PLoS ONE 5:e11500 doi:10.1371/journal.pone.0011500

Voorhies EC (1933) Economic aspects of the bee industry. Agricultural Experiment Station,

Wang AX, Tran C, Desai N, Lobell D, Ermon S Deep transfer learning for crop yield prediction with remote sensing data. In: Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, 2018. pp 1-5

Wang Q, Adiku S, Tenhunen J, Granier A (2005) On the relationship of NDVI with leaf area index in a deciduous forest site Remote Sensing of Environment 94:244-255 doi:https://doi.org/10.1016/j.rse.2004.10.006

Wannebo A, Rosenzweig C (2003) Remote sensing of US cornbelt areas sensitive to the El Ni@o-Southern Oscillation International Journal of Remote Sensing 24:2055-2067 doi:10.1080/01431160210154786

Webber E (2011) Eucalypt leaf-flush detection from remotely sensed (MODIS) data. University of Melbourne

Westman W, Price C (1987) Remote detection of air pollution stress to vegetation -

Laboratory-level studies. Paper presented at the International Geoscience and Remote Sensing Symposium, Michigan, USA,

Wiegand C, Gautreaux M (1977) Landsat agricultural land-use survey Photogrammetric Engineering and Remote Sensing 43:207-216

Williams DL, Coiner JC Utilization of LANDSAT imagery for mapping vegetation on the millionth scale. In: Proceedings of the NASA Earth Resources Survey Symposium, 1975. NASA Houston, TX, pp 53-65

Winter S, Leach J, Keatley M, Arundel J (2013) BeeBox Application User Manual.

Yan E, Wang G, Lin H, Xia C, Sun H (2015) Phenology-based classification of vegetation cover types in Northeast China using MODIS NDVI and EVI time series International Journal of Remote Sensing 36:489-512 doi:10.1080/01431161.2014.999167

Yongji W, Qingwen Q, Ying L (2018) Unsupervised Segmentation Evaluation Using Area-Weighted Variance and Jeffries-Matusita Distance for Remote Sensing Images Remote Sensing 10:1193 doi:10.3390/rs10081193

You J, Li X, Low M, Lobell D, Ermon S Deep gaussian process for crop yield prediction based on remote sensing data. In: Thirty-First AAAI Conference on Artificial Intelligence, 2017.

Zhang X et al. (2019) A deep learning-based approach for automated yellow rust disease detection from high-resolution hyperspectral uav images Remote Sensing 11:1554

Zielinski TJ, Harvey E, Sweeney R, Hanson DM (2005) Quantum States of Atoms and Molecules Journal of Chemical Education 82:1880 doi:10.1021/ed082p1880.2

Zomer R, Trabucco A, Ustin S (2009) Building spectral libraries for wetlands land cover classification and hyperspectral remote sensing Journal of Environmental Management 90:2170-2177 doi:10.1016/j.jenvman.2007.06.028

Appendix A - Statements of Contribution from Co-Authors

The following pages contain written statements detailing the candidate's contribution to each of the publications that are included in this thesis, with each co-author endorsing the stated contribution with their signature.

To Whom It May Concern,

I, Tristan Campbell, contributed the following work to the paper/publication entitled "Simple remote sensing detection of Corymbia calophylla Flowers using common 3 –band imaging sensors" (DOI https://doi.org/10.1016/j.rsase.2018.04.009):

- Conceived and designed the study
- Supervised the data acquisition
- Conducted the data analysis
- Wrote the first draft of the publication
- Completed the edits to the paper following internal and peer review.

[redacted]

I, as a Co-Author, endorse that this level of contribution by the candidate indicated above is appropriate.

Dr Peter Fearns [redacted]

To Whom It May Concern,

I, Tristan Campbell, contributed the following work to the paper/publication entitled "Honey crop estimation from space: Detection of large flowering events in Western Australian forests" (DOI https://doi.org/10.5194/isprs-archives-XLII-1-79-2018):

- Conceived and designed the study
- Conducted the data acquisition (under supervision)
- Conducted the data analysis
- Wrote the first draft of the publication
- Completed the edits to the paper following internal and peer review.

[redacted]

I, as a Co-Author, endorse that this level of contribution by the candidate indicated above is appropriate.

Dr Peter Fearns [redacted]

To Whom It May Concern,

I, Tristan Campbell, contributed the following work to the paper/publication entitled "Machine learning regression model for predicting honey harvests" (DOI https://doi.org/10.3390/agriculture10040118):

- Conceived and designed the study
- Sourced the data referenced in the publication
- Conducted the data analysis
- Wrote the first draft of the publication
- Completed the edits to the paper following internal and peer review.

[redacted]

I, as a Co-Author, endorse that this level of contribution by the candidate indicated above is appropriate.

Dr Kingsley Dix[redacted]

Dr Kenneth Dods [redacted]

Dr Peter Fearns [redacted]

Dr Rebecca Handcock [redacted]

Appendix B – Publication Copyright Statements

Copyright Statement

I warrant that I have obtained, where necessary, permission from the copyright owners to use any third-party copyright material reproduced in the thesis (e.g. questionnaires, artwork, unpublished letters), or to use any of my own published work (e.g. journal articles) in which the copyright is held by another party (e.g. publisher, co-author).

Individual copyright permission statements for each publication are provided in the following sections.

First publication: Simple remote sensing detection of *Corymbia calophylla* Flowers using common 3 –band imaging sensors

Campbell T, Fearns P. Simple remote sensing detection of Corymbia calophylla Flowers using common 3 –band imaging sensors. Remote Sensing Applications: Society and Environment. 2018;11:51-63.https://doi.org/10.1016/j.rsase.2018.04.009.

This article was published by Elsevier. Their copyright statement (available at https://www.elsevier.com/about/policies/copyright#Author-rights) includes the following definitions:

Authors can use their articles, in full or in part, for a wide range of scholarly, non-commercial purposes as outlined below:

- *Use by an author in the author's classroom teaching (including distribution of copies, paper or electronic)*
- *Distribution of copies (including through e-mail) to known research colleagues for their personal use (but not for Commercial Use)*
- *Inclusion in a thesis or dissertation (provided that this is not to be published commercially)*
- *Use in a subsequent compilation of the author's works*
- *Extending the Article to book-length form*
- *Preparation of other derivative works (but not for Commercial Use)*
- *Otherwise using or re-using portions or excerpts in other works*

These rights apply for all Elsevier authors who publish their article as either a subscription article or an open access article. In all cases we require that all Elsevier authors always

include a full acknowledgement and, if appropriate, a link to the final published version hosted on Science Direct.

Second publication: Detection of large flowering events in Western Australian forests

Campbell T, Fearns P. Honey crop estimation from space: Detection of large flowering events in Western Australian forests. ISPRS TC I Mid-term Symposium "Innovative Sensing – From Sensors to Methods and Applications"; Karlsruhe, Germany: International Society for Photogrammetry and Remote Sensing; 2018. p. 79 - 86.

This article was published by the International Society for Photogrammetry and Remote Sensing and distributed under the Creative Commons Attribution 4.0 License (as stated on the publications DOI: https://doi.org/10.5194/isprs-archives-XLII-1-79-2018). This license (available at https://creativecommons.org/licenses/by/4.0/) states that:

Subject to the terms and conditions of this Public License, the Licensor hereby grants You a worldwide, royalty-free, non-sublicensable, non-exclusive, irrevocable license to exercise the Licensed Rights in the Licensed Material to:

- *reproduce and Share the Licensed Material, in whole or in part; and*
- *produce, reproduce, and Share Adapted Material.*

If You Share the Licensed Material (including in modified form), You must retain the following if it is supplied by the Licensor with the Licensed Material:

- *identification of the creator(s) of the Licensed Material and any others designated to receive attribution, in any reasonable manner requested by the Licensor (including by pseudonym if designated);*
- *a copyright notice;*
- *a notice that refers to this Public License;*
- *a notice that refers to the disclaimer of warranties;*
- *a URI or hyperlink to the Licensed Material to the extent reasonably practicable.*

As the publication is referenced in this thesis as per international standards (including the DOI), the attribution requirements have been met.

Third publication: Machine learning regression model for predicting honey harvests

This paper was published by MPDI under the Creative Commons Attribution 4.0 License (http://www.ijesrt.com/index.html). As per the second publication (see above), as this

publication is referenced in this thesis as per international standards (including the DOI), the attribution requirements of the Creative Commons Attribution 4.0 License have been met.